# Ecological approach to investigations of non-communicable health challenges: cancers, diabetes mellitus and obesity at the world population level


**A thesis submitted by**

**Wenpeng You**

Bachelor of Biology Education

Master of Science

**For the degree**

## Doctor of Philosophy

Adelaide Medical School

The University of Adelaide

February 2018

# Contents

## Executive abstract

Globally, cancers, diabetes mellitus and obesity have emerged as the major health and development challenges which are responsible for millions of deaths annually. Population-based prevention strategies have been advocated and adopted as a public health approach. However, unfortunately, no country has achieved their expected results in the past 30 years.

An important way to control cancers, diabetes mellitus and obesity is to focus on reducing the risk factors associated with these non-communicable diseases. However, previous case or cohort studies into the risk factors associated with the three epidemics have controversial findings which may be the results of circumstantial study designs. It may be necessary to use broadly based ecological study to obtain new insights into the associations between risk factors and epidemic at population level.

The international health organizations, such as the WHO and the International Diabetes Federation (IDF) monitor and publish country specific health data in relation to the cancer, diabetes and obesity. These data have been helping governments, policy-makers, funders and researchers track and investigate the priorities of health research and development based on public health needs and ensure that funds and resources are used to meet the priorities.

In terms of tracking and investigating the risk factors of the epidemics, ecological studies have several advantages in study designs over case or cohort researches: 1) More risk factors can be included in the data analysis. 2) Cumulative/ prolonged effects of risk factors on epidemics can be considered in the studies through backdating the risk data. 3) The data on risk variables used in ecological studies are objective because they are collected independent from epidemiological data. In patient-based surveys or anonymised clinical records people with any disease tend to exaggerate negative life events in comparison to people with average or good health. For instance, obese people may misinform how much sugar they have consumed trying to appear more cautious in their dietary choices than they really are.

With the advantages of ecological studies, this thesis seeks to show that reduced natural selection, nutrition/diet and birth behaviour may be independent predictors of the modern noncommunicable epidemics. To achieve this, we collected and analysed data from 191 countries across over 30 years in ten investigations:

Natural selection is considered a force of evolution that adapts populations to their environments. However, humans manipulated their environments and supplemented natural properties of their bodies by medical procedures and technologies, so that natural selection no longer is a force of adaptation. Its operation as a force differentiating reproductive success of individuals has been seriously relaxed. This allows practically any person to pass their genes to the next generation, thus leading to accumulation of deleterious mutations whose effects are controlled by artificial means.

In Investigations 1-3, it is proposed that modern humans may not be naturally well adapted to the current environment because their survival capacity and "fitness" have been maintained by application of high levels of medical services, nutrition and public health advocacy. The studies were conducted through analysing correlations between relaxed natural selection indexed by the Biological State Index ($I_s$) with incidence rates of cancers and Type 1 diabetes mellitus, and prevalence rates of sex-specific obesity.

Meat has been advocated as one of the major contributors to obesity prevalence because it contains high energy component of fat. It is a fact that selective breeding, butchery and cooking which aim for leanness (more protein) have minimized the fat intake in our daily diet. However, meat is still reported as a contributor to body weight increase significantly because of its protein content.

Investigation 4 hypothesized that meat protein in modern diet may have been providing energy surplus to our daily life which contributes to obesity. The hypothesis was examined through analysing the correlations between obesity prevalence and total meat and meat protein consumption respectively.

Both meat and sugar (sucrose) in our daily diet contain the slower digested component and cause insulin resistance. However, it is widely accepted that sugar has been a major contributor to obesity. The role of meat in this regard has not been widely recognised. Investigation 5 compared the use of sugar and meat to predict obesity prevalence worldwide showing that meat availability predicts increase of obesity to the same extent as sugar availability.

Red meat and processed meat have been proposed as the major predictors of prostate cancer, but those studies are circumstantial, and the findings are controversial. Total meat (flesh) has not been associated with prostate cancer.

Investigation 6 postulated that total meat (flesh) may be an independent predictor of prostate cancer. This postulation was examined using country specific data, from a global perspective, that population with more total meat consumption, may have higher

incidence rate of prostate cancer, with empirical, macro-level data collected from the major international organizations.

Gluten has been considered as the trigger of a number of diseases. Worldwide, incidence of gluten-related diseases is increasing. Wheat, the storage proteins, is the main source of gluten, but the adverse effects of wheat on obesity have not been tested.

Investigation 7 analysed and compared the associations between obesity prevalence and wheat, rice and maize, and identified that wheat is the hidden risk factor of obesity. Contrarily, consumption of maize and rice showed the protective role in obesity prevalence. Therefore, the adverse effects of wheat on increasing body weight may have been covered by maize and rice when cereals consumption is advocated as the healthy diet component.

Previous studies into the relationship between low parity and risk of cancers revealed that the decreasing number of children born into a family was associated with the risk of cancers of the mother and a few other cancers of family members. However, these studies did not identify that parity may be the most influential predictor of breast cancer and ovarian cancer. Neither did these studies show that greater parity has the protecting effects on developing site cancers of family members.

Investigation 8 hypothesized that greater family size may protect the whole family from developing cancers. The hypothesis was examined through analysing relationships between total fertility rate, indexing family size and incidence rates of male and female cancers.

Investigations 9 and 10 analysed and compared the contributing effects of multiple risk factors of female breast cancer and ovarian cancer and identified that low parity (indexed by birth rate) may be the most influential risk factor of female breast cancer and ovarian cancers respectively.

The information gathered from the ten studies reveals that 1) Reduced natural selection may be the significant predictor of cancer, Type 1 diabetes and obesity; 2) Meat consumption may be the risk predictor of obesity and prostate cancer; 3) Wheat may be a hidden contributor to obesity prevalence worldwide. 4) The number of children born into a family may be the strong predictor of female breast cancer and ovarian cancer and it may be associated with the cancer risk of all family members.

In general terms, the investigations presented in this thesis show that “ecological analyses” of worldwide data confirm known relationships between some risk factors and incidence/prevalence of non-communicable diseases and can reveal new, hitherto unknown relationships, that are interpretable in the context of human biology.

## Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Wenpeng You and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis (as listed below) resides with the copyright holder(s) of those works. I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, the Australasian Digital Theses Program (ADTP) and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

_____ February 2018

Wenpeng You Date

## Acknowledgements

I would firstly like to thank my supervisors, Professor Maciej Henneberg and Dr Arthur Saniotis for all their time and effort spent aiding me in my research and writing.

Acknowledgement must also go to the following organizations and their staff who helped collect and define the research data:

- The World Health Organization (WHO) - Gretchen Stevens and Regina Guthold
- The International Agency for Research on Cancer (IARC) - Jacques Ferlay, Martyn Plummer, Katiuska Veselinovic, Catherine de Martel
- The World Cancer Research Fund UK -Beverley Smith from
- The World Bank- Ana Margarida Fernandes
- Food and Agriculture Organization (FAO)- Josef Schmidhuber, Luigi Castaldi, Salar Tayyib, Suzanne Vancliffetoresi, Fabio Grita;
- The International Diabetes Federation (IDF) - Katherine Ogurtsova, Marie-Astrid Thielens
- The Institute for Health Metrics and Evaluation (IHME) - Clint Kruchoski, Amy VanderZanden, Brent Bell
- The Imperial College London- Majid Ezzati
- The International Sugar Organization (ISO)- Sergey Gudoshnikov, Maureen Josiah
- The United Nations (UN)- Francois Pelletier, Armando Serrano Lombillo

I would like to thank Professor Ian Symonds and Professor Frank Ruehli for assisting me in consolidating the hypotheses of studies, and thanks goes to Dr Renata Henneberg for her professional supervision of manuscript writing and editing.

I would also like to acknowledge staff at the University of Zürich for their assistance and support while I was visiting Zürich (Switzerland) to collect data. Specifically, these people include Professor Frank Ruehli, Dr Kaspar Staub, Dr Francesco Galassi, Marianne Ott, Gülfirde Akgül and Dr Nikola Koepke.

Thanks go to Dr Damien Belobrajdic from the Commonwealth Scientific and Industrial Research Organisation (CIRSO) who assisted me in drafting and revising the first manuscript during my PhD candidature.

Michael Brockhouse aided in technical aspects of data analysis and should be acknowledged.

My research has been supported by the Australian Postgraduate Award, Adelaide Medical School Travel Award, The University of Adelaide and the Mäxi Foundation Zürich.

I would like to thank members of the Biological Anthropology and Comparative Anatomy Research Unit for their support during my entire PhD candidature. They are Dr Arthur Saniotis, Mr Dante Roccisano, Dr Malcolm Brinn, Dr Arjun Burlakoti and Dr Teghan Lucas.

I also thank Dr Michael Xiaoliang Tong from the Public Health School of the University of Adelaide for his assistance in writing the thesis.

The big thanks must go to my parents-in-law for their full support. I also thank my parents for their constant reminding that I must look after myself as I had passed the right age to work and study that hard.

I appreciate my wife, Lingling Hao's support when she had to look after a young kid and had the second baby boy.

I will not forget that my two kids, Hao You and Thomas You, put blanket on me when I fell asleep on the carpet due to long hours of study and work.

## Articles included in this thesis

1) Wenpeng You, Maciej Henneberg. Cancer incidence increasing globally: The role of relaxed natural selection. Evolutionary Applications, 2017, DOI: 10.1111/eva.12523

2) Wenpeng You, Maciej Henneberg. *Type 1 diabetes prevalence increasing globally and regionally: the role of natural selection and life expectancy at birth*. BMJ Open Diabetes Research and Care, 2016;4:e000161. doi:10.1136/bmjdrc-2015-000161

3) Wenpeng You, Maciej Henneberg. Reduced natural selection contributes to global obesity increase more in males than in females due to more environmental modifications in female body mass. PLOS ONE (awaiting final decision after two revisions)

4) Wenpeng You, Maciej Henneberg. *Meat consumption providing a surplus energy in modern diet contributes to obesity prevalence: an ecological analysis*. BMC Nutrition, 20162:22. DOI: 10.1186/s40795-016-0063-9

5) Wenpeng You, Maciej Henneberg. *Meat in modern diet, just as bad as sugar, correlates to worldwide obesity: an ecological analysis*, Journal of Nutrition & Food Sciences, 6 (4), 1-10.

6) Wenpeng You, Maciej Henneberg. "Meat consumption and prostate cancer incidence-global and regional associations (Abstract)" *BJU INTERNATIONAL*. Vol. 118. 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY-BLACKWELL, 2016. The manuscript, titled Prostate cancer incidence is correlated to total meat (flesh) intake– A cross-national analysis of 172 countries, is under review by BJU International.

7) Wenpeng You, Maciej Henneberg. *Cereal Crops Are not Created Equal: Wheat Consumption Associated with Obesity Prevalence Globally and Regionally*; AIMS Public Health, 2016, 3(2): 313-328. doi: 10.3934/publichealth.2016.2.313

8) Wenpeng You, Frank J. Rühli, Renata Henneberg, Maciej Henneberg. Greater family size associated with less cancer risk: an ecological analysis of 178 countries. Under review by BMC Cancer after two revisions

9) You, W.P., Symonds, I., Rühli, F.J. and Henneberg, M. (2018). Decreasing Birth Rate Determining Worldwide Incidence and Regional Variation of Female Breast Cancer. Advances in Breast Cancer Research, 7, 1-14. https://doi.org/10.4236/abcr.2018.71001

10) Wenpeng You, Ian Symonds, Maciej Henneberg. Low parity may be a significant determinant of ovarian cancer worldwide. Under review by BMC Journal of Ovarian Research.

## Conference presentations made by the candidate in relation to this thesis

1) ***Evolved poor melanin productivity explains high risk of melanoma in Europeans independent of ultraviolet rays' exposure.*** The 11th annual Florey Postgraduate Research Conference at the National Wine Centre 20 September 2017.

2) ***Birth rate determining worldwide incidence and regional variation of female breast cancer.*** ASMR (Australian Society of Medical Research) SA Annual Scientific Meeting, Adelaide Convention and Exhibition Centre, Adelaide, 7 June 2017.

3) ***The Role of Reduced Natural Selection in Boosting Male and Female Obesity Prevalence Worldwide.*** 10th annual Florey Postgraduate Research Conference at the National Wine Centre 19 September 2017.

4) ***Meat Protein, the Later Digested Nutrients, Contributing to Global Obesity.*** The Wednesday Wrap for clinicians and researchers. The Royal Adelaide Hospital, Adelaide, 17 August 2016.

5) ***Meat consumption and prostate cancer incidence - global and regional associations.*** The 17th Asia-Pacific Prostate Cancer Conference (APCC). Melbourne Convention and Exhibition Centre, Melbourne, 31 August to 3 September 2016.

6) ***Similar Correlation of Meat and Sugar to Global Obesity Prevalence***. The 18th International Conference on Nutrition and Food Sciences (World Academy of Science, Engineering and Technology, Dorint Airport-Hotel Zürich, 21-22 July 2016.

7) ***Reduced Natural Selection May Boost Cancer Incidence through Accumulating Mutations.*** ASMR (Australian Society of Medical Research) SA Annual Scientific Meeting, Adelaide Convention and Exhibition Centre, Adelaide, 8 June 2016.

8) ***Per capita wheat consumption and prevalence of obesity - global and regional associations.*** the 54th National Scientific Conference of the Australian Society for Medical Research (ASMR), Stamford Plaza, Adelaide, 15-18 November 2015.

9) ***Reduced Opportunity for Natural Selection Is Associated with Worldwide Increase of Type 1 Diabetes.*** The 2015 SA Population Health Conference (Public Health Association of Australia), Education Development Centre, Adelaide, 31 Oct 2015.

10) ***Global and Regional Correlations between Wheat Consumption and Obesity Prevalence- Cross-Sectional Analysis of Data at Population Level.*** the 9th International Conference on Diet and Activity Methods (ICDAM) Congress, Brisbane Convention and Exhibition Centre, Brisbane, 1 -3 September 2015.

11) ***Food Group Intake and Obesity among Adults.*** The 28th Annual ASHB Conference (Australasian Society for Human Biology (ASHB), Glenelg Comfort Inn, Adelaide, 10-12 December 2014.

# Chapter 1: Relaxed natural selection & global public health challenges

Article 1/10: Cancer incidence increasing globally: The role of relaxed natural selection (Published at Evolutionary Applications 2017)

Wenpeng You[1], Maciej Henneberg[1,2]

[1] Biological Anthropology and Comparative Anatomy Unit, Adelaide Medical School, the University of Adelaide, Adelaide, SA, Australia,

[2] Institute of Evolutionary Medicine, University of Zürich, Zürich, Switzerland

**Published:** Wenpeng You, Maciej Henneberg. Cancer incidence increasing globally: The role of relaxed natural selection. Evolutionary Applications, DOI: 10.1111/eva.12523

✉ **Correspondence:** Wenpeng You wenpeng.you@adelaide.edu.au

## Contextual Statement

Natural selection is the key mechanism of evolution, which changes heritable traits characteristics of a population over generations.

In the past 100-150 years, medical care advancement has reduced natural selection more than previously thought. It has been like a double-edged sword acting on human noncommunicable diseases. It saves the lives of those people with the non-communicable diseases which have strong heredity. Meanwhile, it offers the opportunities for those people to pass on their deleterious genes/mutations to their next generation. With 4-5 successive generations subject to the reduced natural selection, the phenotype of the accumulated deleterious genes/mutations may be noticeable.

We hypothesized and tested that reduced natural selection (indexed by opportunity for reproduction at population level) may be an independent predictor of incidence of cancers that have a heritable background.

# Statement of Authorship

## Statement of Authorship

| | |
|---|---|
| Title of Paper | Cancer incidence increasing globally: The role of relaxed natural selection |
| Publication Status | ☑ Published ☐ Accepted for Publication<br>☐ Submitted for Publication ☐ Publication Style |
| Publication Details | Wenpeng You, Maciej Henneberg. Cancer incidence increasing globally: The role of relaxed natural selection. Evolutionary Applications, DOI: 10.1111/eva.12523 |

### Principal Author

| | | | |
|---|---|---|---|
| Name of Principal Author (Candidate) | Wenpeng You | | |
| Contribution to the Paper | Conceived the hypothesis, collected data, performed analysis on all samples, interpreted data, wrote manuscript and acted as corresponding author. | | |
| Overall percentage (%) | 60 | | |
| Signature | | Date | 22/12/2017 |

### Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

| | | | |
|---|---|---|---|
| Name of Co-Author | Maciej Henneberg | | |
| Contribution to the Paper | Supervised development of work, formalized and interpreted data analysis, and evaluated and edited manuscript | | |
| Signature | | Date | 22/12/2017 |

| | | | |
|---|---|---|---|
| Name of Co-Author | NA | | |
| Contribution to the Paper | | | |
| Signature | | Date | |

Please cut and paste additional co-author panels here as required.

# Abstract

Cancer incidence increase has multiple aetiologies. Mutant alleles accumulation in populations may be one of them due to strong heritability of many cancers. The opportunity for the operation of natural selection has decreased in the past ~150 years because of reduction of mortality and fertility. Mutation-selection balance may have been disturbed in this process and genes providing background for some cancers may have been accumulating in human gene pools. Worldwide, based on the WHO statistics for 173 countries the index of the opportunity for selection is strongly inversely correlated with cancer incidence in peoples aged 0-49 and in people of all ages. This relationship remains significant when GDP, life expectancy of older people ($e_{50}$), obesity, physical inactivity, smoking and urbanization are kept statistically constant for fifteen (15) out of twenty-seven (27) individual cancers incidence rates. Twelve (12) cancers which are not correlated to relaxed natural selection after considering the six potential confounders are largely attributable to external causes like viruses and toxins. Ratios of the average cancer incidence rates of the 10 countries with highest opportunities for selection to the average cancer incidence rates of the 10 countries with lowest opportunities for selection are 2.3 (all cancers at all ages), 2.4 (all cancers in 0-49 years age group), 5.7 (average ratios of strongly genetically based cancers) and 2.1 (average ratios of cancers with less genetic background).



**Keywords:** Biological State Index, Mutations, Life expectancy, Cancer heritability


## Introduction

Worldwide, cancer incidence rate has increased to make it the second leading cause of death after cardiovascular disease. Environmental factors, such as tobacco smoking, urbanization and its associated pollution and changing diet patterns together with increased wealth associated with better medical services and extended post-reproductive life span have been considered responsible for this phenomenon. Prevention and treatment measures focusing on environmental factors have been implemented, but little progress in reducing incidence of cancers has been made [1].

Malignant neoplasms are results of somatic mutations of certain genes [2, 3]. Studies investigating transmission of cancer susceptibility in family lines suggested genetic background for incidence of many types of malignancies [4]. It is possible, then that this background contributes to increasing incidence of cancers at the population level.

Mutations are more common than previously thought [5-7]. For instance, it has been estimated that an average neonate has some 74 *de novo* point mutations [5, 8]. Multiple mutations may accumulate in genomes over time spanning just a few generations [9]. When selection against a certain mutation does not operate, the frequency of mutated alleles doubles every generation [10]. The mutation load is directly proportional to the mutation rate and inversely proportional to the rate of selection [10, 11]. Thus, when selection rates approach zero mutation load approaches infinity. These rates are expressed per generation. Human generations do overlap due to the length of the reproductive life span which in females is approximately 30 years. Assuming, for simplicity's sake, zero selection, it can be shown that mutation load at a given locus can triple or quadruple during one century (2-3 generations). In the recent past, selection operating in human populations has been significantly relaxed [8, 11, 12] by medical and fpublic health actions. This results in accumulation of mutations, especially mildly deleterious mutations. Interactions between alleles of various loci may magnify mutation rates including rates of somatic mutations that result in neoplastic cell growth because of the way DNA replicates and is repaired which is similar in germline and in somatic cells [8]. Combination of effects of mutations with relaxed selection produces a real possibility of deterioration of biological integrity of human organisms, observable in the time of a few generations in most advanced societies.

Human morphological characteristics that have a heritable, polygenic background have been evolving during the Holocene very fast: e.g. rate of cranial capacity change was -10.8 darwins while the cranial index (the ratio of braincase width to its length) changed at a rate of +65.2 darwins [13] and stature at +606.2 darwins [14]. These are polygenic characters with incomplete heritability, we cite them here as an illustration of how development of technological and social adaptations lowering natural selection rates in the last few millennia can influence the course of change in human biological characteristics.

Natural selection is a process that differentiates reproduction of individual genomes into new generations depending on how genetic endowment of parents influences the number of offspring that will replace them in the future [15]. Following Fisher's (1958) definition of the reproductive value, "Biological State Index ($I_{bs}$)" has been proposed to measure an opportunity for an average member of a population to pass genes to the

next generation. $I_{bs}$ calculation combines data on mortality and fertility [9, 16-19]. The formula [16, 17] for $I_{bs}$ calculation is:

$$I_{bs} = 1 - \sum_{x=0}^{x=\omega} d_x s_x$$

Where:

$d_x$ = the frequency of deaths at age x

$s_x$ =the probability of not possessing the complete number of births at age x

$\omega$: the age at death of the oldest member of the group

$I_{bs}$ expresses a probability for an average individual born into a population to pass on genes to the next generation. Index value of 1.0 means that there is no opportunity for natural selection through differential mortality because all individuals survive until the end of their reproductive period.

This index is a more precise calculation than what Crow (1958) called the $P_d$ [11]– proportion of individuals dying before reaching age of reproduction that is used to calculate the index of total selection due to mortality. For this index a "…source of error is that no allowance was made for women who died during the childbearing period after having one or more children." [11]. In the $I_{bs}$ such allowance is made by using $s_x$ and $d_x$ values for ages 15-49 years. By analogy to the Crow's mortality index of $P_d/P_s$ (where $P_s$ is a proportion of individuals surviving to the reproductive age) an index of total opportunity for selection through differential mortality (including its portion during reproductive years) is constructed $I_s = (1-I_{bs})/I_{bs}$. Theoretically, following Fisher's formulation, the opportunity for selection must include the variance of fertility, more precisely, this portion of the variance of fertility that is heritable $V_f/x^2$ (where x is the average number of children per female surviving to the menopause and $V_f$ variance of this number). In humans, however, heritable variance in actual fertility is very low even in couples who do not control family size. According to our study [20] of 7503 births from 1525 Polish and American historical couples in 12 groups free of conscious birth control, the genetic variance of fertility is less than 0.01 of its squared mean. Furthermore, considering that in developed countries conscious birth control has been practiced for

over 100 years and became widespread in at least the last two generations, further diminishing any heritable fertility differentials, the contribution of genetic variance in fertility to the opportunity for natural selection in humans is practically non-existent. Therefore, the use of $I_s$ measuring opportunity for selection through differential mortality provides sufficient approximation of the maximum selective pressures in modern human populations.

The primary role of natural selection is that of the "janitor of the gene pool" purging deleterious mutations. In the past ~100 years, there has been a great reduction in mortality and in fertility that has been limiting the overall opportunity for natural selection [9, 17, 21]. It follows that genes potentially providing background for some cancers have been accumulating in various populations. Cancer incidence may be greater in those populations who have experienced less opportunity for natural selection.

We hypothesise that in a global perspective, extent of relaxation of natural selection in various national populations may be positively correlated to greater cancer incidence.

## Materials and Methods

The country specific variables were collected for this ecological study.

1. Dependent variables: The GLOBOCAN 2012 estimates of incidence rates (C50) [22]

We extracted the cumulative incidence rates of all cancers excl. non-melanoma skin cancer (C00-97, but C44) among people of all ages and people aged 0-49 years respectively for 184 countries. We also captured separate estimates of incidence rates of 27 site cancers from the same source of data for people of all ages. The site cancers are: Lip and oral cavity (C00-08), Nasopharynx (C11), Other pharynx (C09-10,C12-14), Oesophagus (C15), Stomach (C16), Colorectum (C18-21), Liver (C22), Gallbladder (C23-24), Pancreas (C25), Larynx (C32), Lung (C33-34), Melanoma of skin (C43), Kaposi sarcoma (C46), Breast (C50), Cervix uteri (C53), Corpus uteri (C54), Ovary (C56), Prostate (C61), Testis (C62), Kidney (C64-66), Bladder (C67), Brain (C70-72), Thyroid (C73), Hodgkin lymphoma (C81), Non-Hodgkin lymphoma (C82-85,C96), Multiple myeloma (C88+C90) and Leukaemia (C91-95). The cancer incidence rate indicates the number per 100,000 persons who were diagnosed with cancer in 2012. The rate was age-standardized using the world standard population to increase the comparability.

Women age 50+ years enter menopause, which brings their fertility to zero. Female reproductive behaviour has been associated with various female cancers [23-25]. The $I_{bs}$ reflects mortality up to the age 50 years, considered the end of the reproductive life span, because $s_{50+}$ values equal zero (thus any $d_{50+}$ values are multiplied by zero). This

means that natural selection we measure cannot “reach” beyond age 50 years. For these reasons, we included specifically the incidence rate of all cancers in the age range 0-49 years (pre-reproductive and reproductive life span) since these cancers can directly produce mortality and fertility differentials influencing reproductive success of individuals.

2. Independent variable: The index ($I_s = (1-I_{bs})/I_{bs}$) of natural selection opportunity at population level

The $I_{bs}$ was calculated [16, 17] with the data of the world fertility published by United Nations in 2008 [26] and the data of life tables published by World Health Organization (WHO) in 2009 [27].

James Crow [11], based on the Fisher’s (1958) concept of the reproductive value [28], proposed to measure the total opportunity for natural selection (I) as the ratio of variance in offspring size of a couple (V) to the squared average offspring size of a couple ($x^2$) that will replace parents in the next generation. In the application to human populations this approach encounters two problems. The first is the birth control, which is very substantial in many modern societies. The second one is the overlapping of generations due to long reproductive period of females and males. The first problem can be tackled by separating contributions of fertility and mortality to the opportunity for selection and using only the portion of selection resulting from mortality. According to Crow [11], the index of opportunity for natural selection through differential mortality ($I_m$) is the ratio of individuals dying before reaching reproductive age ($P_d$) to the individuals surviving ($P_s$): $I_m = P_d/P_s$. Since not all individuals surviving to reproductive maturity will survive through the entire reproductive life span, a correction for deaths during the reproductive period is needed. This is introduced in the form of the Biological State Index ($I_{bs}$) that combines age specific mortality ($d_x$) with age specific opportunity for producing offspring in the future life ($s_x$) (Figure 1) [16, 17]. The Biological State Index accumulates mortality data in the way similar to “survival” biometric function of the life table and depends on the distribution of age specific relative fertility expressed as the fraction of the Total Fertility Rate remaining to be produced by a person of age x. Multiplication of the $I_{bs}$ value for a given population by the Total Fertility Rate of this population (number of children born by females surviving to the menopause) produces the Net Reproductive Rate, a generational measure of population growth. Details regarding $I_{bs}$ are explained in several previously published studies [13, 16-19]. Henneberg (1980) [20] proposed that considering low heritable variance of fertility and the widespread birth control that allow us to neglect opportunity for natural selection through differential fertility, the index of the total opportunity for selection in modern populations is: $I_s = (1-I_{bs})/I_{bs}$. The lower the value

of this index, the less opportunity for natural selection exists. None of the three indices discussed here ($I_m$, $I_{bs}$ or $I_s$) has any unit because they are ratios of offspring numbers or proportions and probabilities. Indices of the opportunity for selection measure the upper limit of the total selection pressure. Actual selection pressures can be lower because not all mortality differentials are heritable, but the magnitude of selection cannot exceed index values. Therefore, decreasing values of opportunity for selection indices certainly show reduction in possibility of selection to occur, while they do not measure the actual magnitude of selection that can be lower.

Gross Domestic Product per capita (GDP), life expectancy of older people ($e_{50}$), obesity (BMI ≥ 30 kg/m$^2$) prevalence rate [29], physical inactivity prevalence rate [30-32], smoking and urbanization [30, 33-35] have been associated with cancer initiation. They were considered as the confounders when we conducted the data analysis in this study.

3. The World Bank published data [36] on GDP

GDP is used as the index of socio-economic level and it is expressed in per capita purchasing power parity (PPP in current international USD) in 2010. Socio-economic levels measured with GDP have been related to cancer incidence rate [22, 33, 37, 38].

4. The United Nations Statistics Division estimates of the life expectancy [39]

Increasing life expectancy of older people, indexing ageing in this study, has been considered as a factor possibly promoting increasing cancer incidence [40, 41]. Therefore, the life expectancy ($e_{50}$, 1990-95) was extracted from abridged life tables (1950-2100) [39] published online by the United Nations.

5. The WHO Global Health Observatory (GHO) data on estimated obesity prevalence rate, physical inactivity, smoking rate and urbanization [29]

The obesity prevalence is expressed as the percent of population (2010) aged 18+ with Body Mass Index (BMI) ≥ 30 kg/m$^2$.

Physical inactivity is defined as the percent of a particular population attaining less than 150 minutes of moderate-intensity physical activity per week, or less than 75 minutes of vigorous-intensity physical activity per week, or equivalent in 2010.

Smoking is expressed as the percent of adults aged 15 years and over (age-standardized rate) who smoked any tobacco product daily in 2010.

Urbanization is expressed with the percent of total population living in urban areas in 2010. Urbanization, representing a major demographic shift, entails lifestyle changes,

including diet with more energy dense components, such as high fat and high alcohol consumption in daily diet, and less physical exercise.

## Data Selection

We used country specific cancer incidence rates, life tables and fertility rates (for $I_s$ calculation), GDP, life expectancy at 50 years of age ($e_{50}$), obesity prevalence rate, physical inactivity prevalence rate, smoking and urbanization for all countries where data were available. We aligned cancer incidence rates with $I_s$ by country and we obtained a set of data consisting of 173 countries. Quality of the country-specific cancer estimation depends upon the quality and the amount of the information available for each country [42]. For data robustness check, we clustered the countries with "high quality" data as defined by the International Agency for Research on Cancer (IARC) [42], obtaining a subset of data comprising 64 countries. This smaller data set was analysed separately from the other set of data consisting of all 173 countries. Country specific GDP, life expectancy ($e_{50}$), obesity prevalence rate, physical inactivity prevalence rate, smoking and urbanization were matched with the listing of 173 countries which have both cancer incidence rate and $I_s$. Numbers of countries included in the analysis of relationships with other variables may have differed somewhat because all information was not uniformly available for all countries.

All data included in this study were published by UN agencies. No ethical approval or written informed consent for participation was required.

## Data analysis

Various statistical analysis methods were applied in this study to explore the correlation between $I_s$ and cancer incidence rates. Each country was treated as an individual subject in the analysis. To examine the correlation between $I_s$ and cancer incidence rates, the analysis proceeded in five steps:

1. Pearson's r and nonparametric correlations (Spearman's "rho") were used to evaluate the strength and direction of the correlation between all the variables. Pearson's correlations and partial correlations were calculated using log-transformed (ln) variables to minimise non-homoscedascity of their distributions. Fisher's z-transformation of correlation coefficients was used to assess significance of individual correlation coefficients values and of differences between correlation coefficients values.

2. The independent relationships between $I_s$ and each cancer incidence rate for all ages were explored with partial correlation of Pearson's moment-product approach while we

controlled for the six variables, which are GDP [38], life expectancy ($e_{50}$) [43-45], obesity prevalence rate [46] [45], physical inactivity prevalence rate [46], smoking [45, 46] and urbanization [46]. Life expectancy ($e_{50}$) was not controlled for when the independent relationship between cancer incidence rate among the people aged 0-49 years and $I_s$ was studied because this potential confounder is not relevant to this group of people.

We controlled for GDP not only because it stands for cancer treatment service, but also because it is associated with cancer diagnoses level. Therefore, we considered GDP as a potential confounder and controlled for in our data analysis, which may reduce the influence of GDP associated cancer diagnose rate.

Urbanization, representing a major demographic shift, entails lifestyle changes, including diet with more energy dense components, such as high fat, high alcohol consumption, less vegetables and fruits in daily diet, and less physical exercise [30-32].

Those individual (site) cancers whose incidence rates were significantly and negatively correlated to Is in partial correlation are classified as “cancers with strong genetic background”. Those individual cancers whose incidence rates were not significantly or negatively correlated to $I_s$ are called “Less genetic cancers”.

Cohen’s $f^2$ was used to calculate and to report the “effect size” in this study.

3. Standard multiple linear regression (enter) was performed to describe the relationships between the outcome variables (all cancers among all ages and 0-49 years age group) and the explanatory variables (GDP, life expectancy ($e_{50}$), obesity prevalence rate, physical inactivity prevalence rate, smoking [46] and urbanization [46]). Standard multiple linear regression (stepwise) was performed to identify the most significant predictors of all cancer incidence rates among all ages and 0-49 years respectively.

Life expectancy of older people ($e_{50}$) was not included as an independent predictor in the standard multiple linear regression analysis when we explored the relationships between all cancer incidence rate among the population aged 0-49 years and $I_s$ because this potential confounder is not relevant to this group of people.

4. In order to demonstrate the universal association between all cancer incidence rate (all ages) and $I_s$, we categorized the countries for correlation analyses based on: 1) the WHO regional classifications, Africa (AFR), Americas (AMR), Eastern Mediterranean (EMR), Europe (EU), South-East Asia (SEAR) and Western Pacific (WPR) [47]; 2) the World Bank income classifications: high income, upper middle income, low-middle income and low income; 3) countries with the strong contrast in terms of geographic distributions, per capita GDP levels and/or cultural backgrounds. We analysed the

correlation in the six country groupings: the Arab World [48], countries with English as the official language (government websites), the Organisation for Economic Co-operation and Development (OECD) [49], the Asia-Pacific Economic Cooperation (APEC) [50], Asia Cooperation Dialogue (ACD) [51] and the Southern African Development Community (SADC) [52]. In our analysis, we only included those countries for which we could access their data for the specific groupings. To a large extent, grouping countries for analysis may also allow us to align our findings against previous local or regional studies regarding heterogeneous cancer epidemiology due to various geographic locations and ethnicity.

Socioeconomic level in different regions has been considered as the major contributor to regional variations of cancer incidence rates [22, 37, 38]. Therefore, the correlation coefficients between groupings in different socioeconomic levels were compared with Fisher's r-to-z transformation.

5. IARC-WHO has reported that GDP is associated with cancer incidence rate [22, 37, 38, 53]. Naturally, this drove us to consider the incidence rate of all cancers (all ages) without the contributing effect of GDP. This allows us to explore the association between $I_s$ and incidence rate of all cancers (all ages) which excludes the contributing effect of GDP.

Scatter plots (simple regression analysis) were used to explore and visualize the correlations between all cancer incidence rate (all ages) and $I_s$. The strength and form of the relationship between incidence rate of all cancers (all ages) and $I_s$ was analysed using actual values of the two variables. The equation ($y = -41.97\ln(x) + 36.831$) of the best fitting trendline (logarithmic) displayed in the scatter plots analysis of relationship between GDP and all cancer (all ages) incidence rate was used to calculate and remove the contributing effect of GDP on all cancer (all ages) incidence rate. This allowed us to obtain a new dependent variable "Residual of all cancer (all ages) incidence standardised on GDP". The relationship between $I_s$ and "Residual of all cancers (all ages) incidence standardised on GDP" was explored with scatter plots (Figure 2).

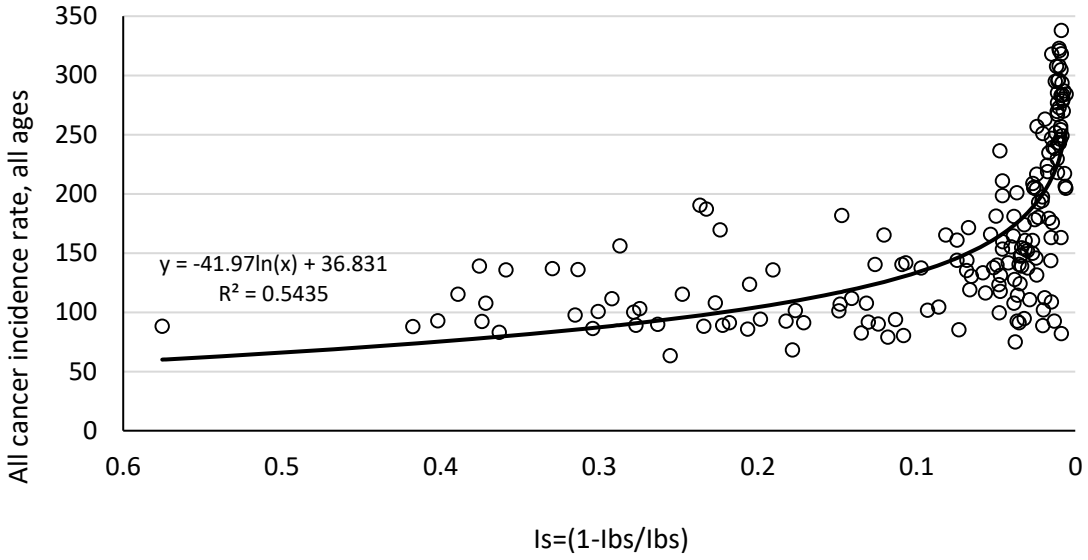
In order to assess the magnitude of possible changes in the incidence of cancers due to relaxation of natural selection we have calculated a "rate of incidence increase" by dividing the average incidence rates in the 10 countries with the lowest $I_s$ values by the average incidence rates in the 10 countries with the highest $I_s$ values. These rates allow us to estimate to what extent alteration of the mutation-selection balance over short periods could be responsible for the change in incidence. This is an approximate measure because other (non-genetic) factors may also influence incidence rates.

Pearson, non-parametric and partial correlations, and the multiple linear regression analysis were conducted using SPSS v. 22 (SPSS Inc., Chicago Il USA). Scatter plots and calculation of "Residual of all cancer (all ages) incidence standardised on GDP" were performed in Excel® (Microsoft 2016). The raw data are used for scatter plots and calculation of "Residual of all cancer (all ages) incidence standardised on GDP". The significance value is recorded for each correlation, and significance level is kept at the 0.05, but 0.01 and 0.001 levels can be found from the reported actual significance values. Standard multiple linear regression analysis criteria were set at probability of F to enter ≤ 0.05 and probability of F to remove ≥ 0.10.

## Results

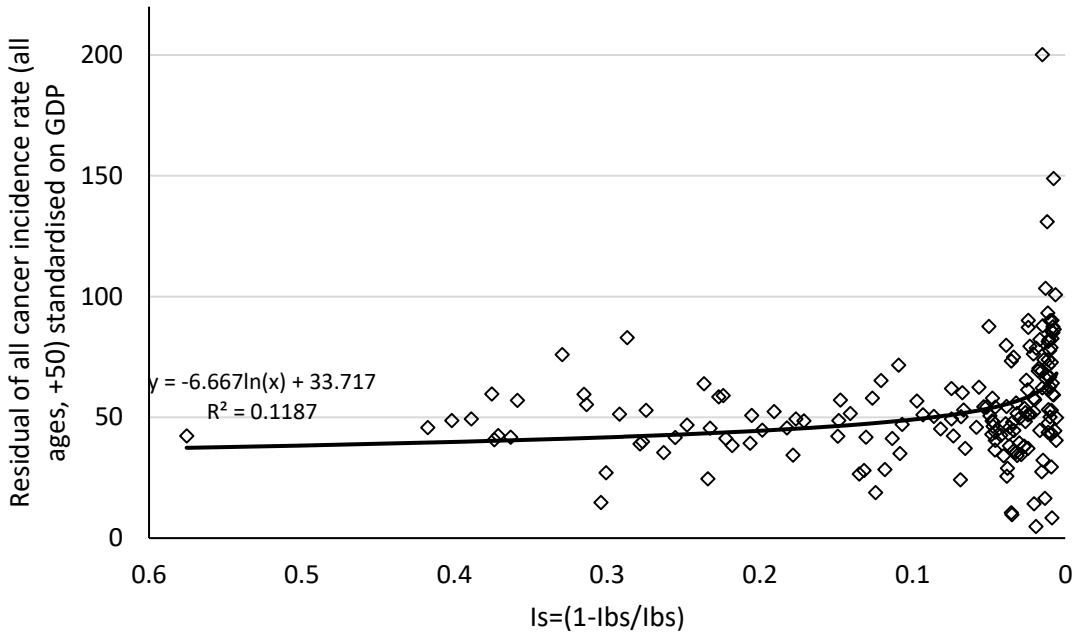
The relationship between $I_s$ and all cancer incidence was negative and strong ($R^2$=0.5435, Figure 1).

Figure 1. The relationship between $I_s$ and all cancer incidence rate (all ages)




When the contributing effect of GDP on all cancer incidence rates was removed, $I_s$ was still in negative and significant correlation to all cancer incidence ($R^2$=0.1187, Figure 2).

Figure 2. The relationship between $I_s$ and residual of all cancer incidence rate (+50, all ages) standardised on GDP




Globally, $I_s$ was significantly and negatively correlated to the incidence rates of all cancers at all ages (r=-0.738, p<0.001) and at 0-49 years (r=-0.719, p<0.001) in Spearman rho analysis (Table 1). This relationship trend remained (r=-0.319, p<0.001 and r=-0.380, p<0.001 respectively) when we controlled for potential confounding effects of GDP, life expectancy, obesity, physical inactivity, smoking and urbanization in partial correlation analysis (Table 1). When exploring partial correlations of $I_s$ to individual cancers significant negative correlation was found for 15 out 27 site cancers (Table 1). Similar results were observed in the correlation analysis with the data comprising 64 countries with “high quality” data (Table 1). Rates of incidence increase for all cancers at all ages (2.3) and in 0-49 years age group (2.4) are practically the same, while for individual cancers these rates of incidence increase vary from fractional (=decrease) for cancers not significantly or not negatively correlated with $I_s$ to over 10 for some cancers significantly negatively correlated with $I_s$. Overall, for cancers with the strong genetic background which were significantly negatively correlated with $I_s$ the average rate of

incidence increase is 5.7 while for the less genetic cancers the average rate of incidence increase is 2.1.

Relationships between $I_s$ and some site cancer correlations were illustrated in Figure 3. As can be seen, cancers that had predominantly external causes such as cervical cancer or oesophageal cancer showed no correlation to $I_s$, while those with possible genetic background do correlate with $I_s$. Partial correlation between $I_s$ and 15 cancers remained significant after removal of the confounding effects (Table 1) of the six potential confounders.

The multiple linear regression model (Table 2) showed that, globally, $I_s$ had the greatest beta coefficient than the potential confounders in the “Enter” analysis, whereas the stepwise regression model identified $I_s$ as the most significant predictor of all cancers incidence rates among all ages and 0-49. Similar results were revealed after the multiple linear regression model was calculated within the dataset which only included those 64 countries with “high quality” data.

Table 1 Spearman rho and partial correlations between $I_s$ and all cancers incidence rates of all ages and 0-49, and 27 site cancers respectively

| | All countries | | | | | | | Countries with "High quality" data | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Nonparametric | | | Partial | | | | Nonparametric | | | Partial | | | |
| Independent Variables (Cancer) | rho | p | N | r | p | df | Effect Size | rho | p | n | r | p | df | Effect Size |
| All cancers (C00-97, but C44), all ages | -0.738 | <0.001 | 173 | -0.319 | <0.001 | 98 | 0.113 | -0.650 | <0.001 | 64 | -0.348 | 0.024 | 40 | 0.132 |
| All cancers (C00-97, but C44), 0-49 ‡ | -0.719 | <0.001 | 173 | -0.380 | <0.001 | 99 | 0.168 | -0.607 | <0.001 | 64 | -0.446 | 0.003 | 41 | 0.248 |
| Bladder (C67) | -0.709 | <0.001 | 173 | -0.217 | 0.030 | 98 | 0.049 | -0.571 | <0.001 | 64 | -0.248 | 0.114 | 40 | 0.065 |
| Brain (C70-72) | -0.738 | <0.001 | 170 | -0.247 | 0.013 | 98 | 0.065 | -0.389 | <0.001 | 64 | -0.405 | 0.008 | 40 | 0.196 |
| Breast (C50) | -0.737 | <0.001 | 173 | -0.290 | 0.003 | 98 | 0.092 | -0.723 | <0.001 | 64 | -0.300 | 0.054 | 40 | 0.099 |
| Cervix uteri (C53) | 0.608 | <0.001 | 173 | 0.071 | 0.485 | 98 | 0.005 | 0.407 | <0.001 | 64 | -0.040 | 0.803 | 40 | 0.002 |
| Colorectum (C18-21) | -0.845 | <0.001 | 173 | -0.455 | <0.001 | 98 | 0.261 | -0.723 | <0.001 | 64 | -0.433 | 0.004 | 40 | 0.231 |
| Corpus uteri (C54) | -0.674 | <0.001 | 172 | -0.337 | <0.001 | 98 | 0.128 | -0.528 | <0.001 | 64 | -0.405 | 0.008 | 40 | 0.196 |
| Gallbladder (C23-24) | -0.509 | <0.001 | 158 | -0.226 | 0.024 | 98 | 0.054 | -0.096 | 0.452 | 63 | 0.106 | 0.502 | 40 | 0.011 |
| Hodgkin lymphoma (C81) | -0.666 | <0.001 | 166 | -0.270 | 0.007 | 98 | 0.078 | -0.491 | <0.001 | 64 | -0.347 | 0.024 | 40 | 0.137 |
| Kaposi sarcoma (C46) | 0.564 | <0.001 | 120 | 0.325 | 0.004 | 76 | 0.118 | 0.286 | 0.052 | 47 | 0.275 | 0.128 | 30 | 0.082 |
| Kidney (C64-66) | -0.850 | <0.001 | 167 | -0.485 | <0.001 | 98 | 0.308 | -0.562 | <0.001 | 64 | -0.425 | 0.005 | 40 | 0.221 |
| Larynx (C32) | -0.448 | <0.001 | 168 | -0.144 | 0.154 | 98 | 0.021 | 0.005 | 0.966 | 64 | -0.182 | 0.248 | 40 | 0.034 |
| Leukemia (C91-95) | -0.800 | <0.001 | 171 | -0.392 | <0.001 | 98 | 0.182 | -0.585 | <0.001 | 64 | -0.352 | 0.022 | 40 | 0.141 |
| Lip and oral cavity (C00-08) | -0.257 | <0.001 | 173 | -0.037 | 0.712 | 98 | 0.001 | -0.359 | 0.004 | 64 | -0.335 | 0.030 | 40 | 0.126 |
| Liver (C22) | 0.300 | <0.001 | 173 | -0.033 | 0.745 | 98 | 0.001 | 0.041 | 0.750 | 64 | 0.136 | 0.392 | 40 | 0.019 |
| Lung (C33-34) | -0.782 | <0.001 | 173 | -0.295 | 0.003 | 98 | 0.095 | -0.483 | <0.001 | 64 | -0.244 | 0.119 | 40 | 0.064 |
| Melanoma of skin (C43) | -0.482 | <0.001 | 168 | -0.155 | 0.124 | 98 | 0.025 | -0.613 | <0.001 | 63 | -0.283 | 0.069 | 40 | 0.087 |
| Multiple myeloma (C88, C90) | -0.663 | <0.001 | 157 | -0.236 | 0.018 | 98 | 0.059 | -0.547 | <0.001 | 64 | -0.077 | 0.626 | 40 | 0.006 |
| Nasopharynx (C11) | 0.221 | <0.001 | 154 | 0.114 | 0.257 | 98 | 0.013 | 0.334 | 0.008 | 63 | 0.144 | 0.364 | 40 | 0.021 |
| Non-Hodgkin lymphoma (C82-85, C96) | -0.524 | <0.001 | 173 | 0.031 | 0.756 | 98 | 0.001 | -0.565 | <0.001 | 64 | -0.114 | 0.472 | 40 | 0.013 |

| | All countries | | | | | | | Countries with “High quality” data | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Nonparametric | | | Partial | | | | Nonparametric | | | Partial | | | |
| Independent Variables (Cancer) | rho | p | N | r | p | df | Effect Size | rho | p | n | r | p | df | Effect Size |
| Esophagus (C15) | 0.009 | 0.907 | 172 | 0.132 | 0.189 | 98 | 0.018 | 0.008 | 0.951 | 64 | -0.004 | 0.978 | 40 | 0.000 |
| Other pharynx (C09-10, C12-14) | -0.347 | <0.001 | 168 | -0.091 | 0.367 | 98 | 0.008 | -0.371 | 0.003 | 63 | -0.263 | 0.093 | 40 | 0.074 |
| Ovary (C56) | -0.608 | <0.001 | 173 | -0.309 | 0.002 | 98 | 0.106 | -0.449 | <0.001 | 64 | -0.469 | 0.002 | 40 | 0.282 |
| Pancreas (C25) | -0.802 | <0.001 | 170 | -0.453 | <0.001 | 98 | 0.258 | -0.602 | <0.001 | 64 | -0.396 | 0.009 | 40 | 0.186 |
| Prostate (C61) | -0.498 | <0.001 | 173 | -0.114 | 0.260 | 98 | 0.013 | -0.577 | <0.001 | 64 | -0.301 | 0.053 | 40 | 0.100 |
| Stomach (C16) | -0.412 | <0.001 | 173 | -0.243 | 0.015 | 98 | 0.063 | 0.049 | 0.700 | 64 | -0.281 | 0.072 | 40 | 0.086 |
| Testis (C62) | -0.777 | <0.001 | 153 | -0.315 | <0.001 | 98 | 0.110 | -0.681 | <0.001 | 64 | -0.459 | 0.002 | 40 | 0.267 |
| Thyroid (C73) | -0.684 | <0.001 | 170 | -0.322 | <0.001 | 98 | 0.115 | -0.346 | 0.005 | 64 | -0.113 | 0.477 | 40 | 0.013 |
| GDP PPP 2010 | -0.853 | <0.001 | 168 | - | - | - | - | -0.760 | <0.001 | 63 | - | - | - | - |
| Life expect (e50), 1990-95 | -0.822 | <0.001 | 173 | - | - | - | - | -0.666 | <0.001 | 64 | - | - | - | - |
| Obesity | -0.572 | <0.001 | 173 | - | - | - | - | -0.003 | 0.984 | 64 | - | - | - | - |
| Physical inactivity | -0.315 | <0.001 | 132 | - | - | - | - | -0.103 | 0.453 | 55 | - | - | - | - |
| Smoking, Daily any tobacco product | -0.551 | <0.001 | 123 | - | - | - | - | -0.234 | 0.086 | 55 | - | - | - | - |
| Urbanization 2010 | -0.712 | <0.001 | 169 | - | - | - | - | -0.455 | <0.001 | 64 | - | - | - | - |

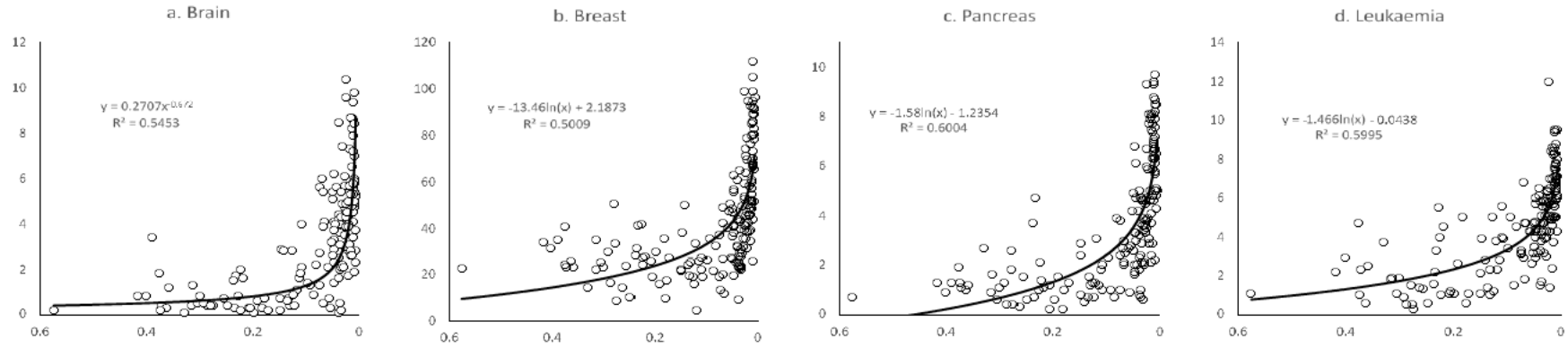
ǂ Life expectancy (e50) was not controlled as it is not relevant in population aged 0-49 years old.

Age Standardised Incidence Rate (ASR, All ages, World) per 100,000, GDP PPP is in per capita USD per year

Partial correlations were calculated when urbanization, GDP, life expectancy (e50) and smoking rate were kept statistically constant.

Figure 3. The relationship between relaxation of natural selection and incidence of selected cancers

***The strong correlation between relaxation of natural selection and incidence of selected cancers***




***The poor correlation between relaxation of natural selection and incidence of selected cancers***

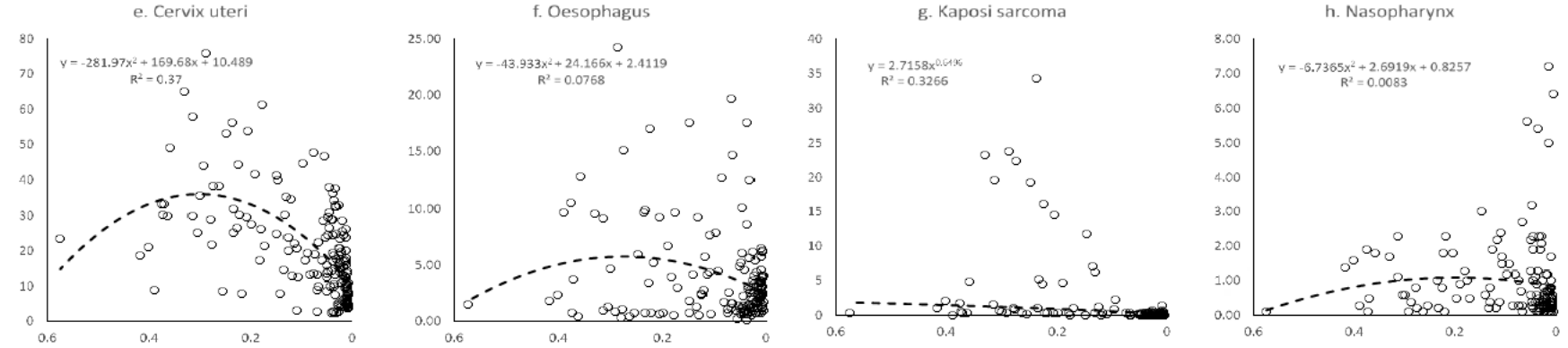


Table 2 Results of multiple linear regression analyses to identify predictors of cancer incidence

**Enter**

| | All countries (n=173) | | | | | Countries with "High quality" data (n=64) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | All ages | | 0-49 years old | | | All ages | | 0-49 years old | |
| Variable | Beta | Sig. | Beta | Sig. | Variable | Beta | Sig. | Beta | Sig. |
| $I_S$ | -0.373 | 0.014 | -0.523 | 0.000 | $I_S$ | -0.809 | 0.011 | -1.090 | <0.001 |
| GDP | 0.190 | 0.207 | 0.254 | 0.000 | GDP | -0.417 | 0.174 | -0.308 | 0.345 |
| Life expectancy | 0.126 | 0.241 | - | - | Life expectancy | 0.434 | 0.009 | - | - |
| Obesity | 0.100 | 0.211 | 0.040 | 0.608 | Obesity | 0.213 | 0.087 | -0.002 | 0.987 |
| Physical Inactivity | -0.071 | 0.339 | -0.057 | 0.436 | Physical Inactivity | -0.232 | 0.092 | -0.192 | 0.189 |
| Smoking | 0.220 | 0.007 | 0.204 | 0.011 | Smoking | -0.016 | 0.911 | 0.015 | 0.920 |
| Urbanization | -0.059 | 0.499 | -0.164 | 0.056 | Urbanization | -0.131 | 0.547 | -0.224 | 0.336 |

**Stepwise**

| | All countries (n=173) | | | | Countries with "High quality" data (n=64) | | | |
|---|---|---|---|---|---|---|---|---|
| | All ages | | 0-49 years old | | All ages | | 0-49 years old | |
| Rank | Variable | Adjusted $R^2$ | Variable | Adjusted $R^2$ | Rank | Variable | Adjusted $R^2$ | Variable | Adjusted $R^2$ |
| 1 | Is | 0.589 | $I_S$ | 0.597 | 1 | $I_S$ | 0.400 | $I_S$ | 0.340 |
| 2 | Smoking | 0.616 | Smoking | 0.625 | 2 | Physical Inactivity | 0.442 | Urbanization | 0.442 |
| 3 | - | - | - | - | 3 | Life expectancy | 0.492 | - | - |

$I_s$ was correlated to incidence rate of all cancers (all ages) universally in all country groupings (Table 3). However, there was a tendency for the correlations to be stronger in the more developed country groupings than those in the less developed groupings. This trend was revealed in country groupings divided in consideration of geographic locations (5 WHO regions), income classifications (4 groups of the World Bank) and other factors, such as cultural backgrounds (Arab World, countries with English as official language) and international organizations (OECD, APEC, ACD, SADC).

The more developed regions, Americas and Europe, had stronger correlations than those in other regions. Fisher r-to-z transformation revealed that the correlation between $I_s$ and incidence rate of all cancers (all ages) in Europe was significantly stronger than those in the three developing regions, Africa ($z=4.41$, $p<0.001$) and Eastern Mediterranean ($z=3.8$, $p<0.001$) and South-East Asia ($z=2.78$, $p=0.0027$). It was also revealed that in the World Bank income classifications, the correlation between $I_s$ and incidence rate of all cancer (all ages) in the upper middle-income grouping was significantly stronger than that in low income classification ($z=2.48$, $p=0.0066$).

The correlation between $I_s$ and incidence rate of all cancers (all ages) in high income classification was not as strong as that in the upper middle classification. It was almost the same as that in the low middle classification (Table 3).

Table 3 Associations between $I_s$ and cancer incidence (all ages) in different country groupings

| | Pearson's r | | Non-parametric | |
|---|---|---|---|---|
| | Pearson's r | Significance | Spearman | Significance |
| WHO Region | | | | |
| AFRO, n=44 | -0.151 | 0.329 | -0.099 | 0.523 |
| AMRO, n=29 | -0.695 | <0.001 | -0.662 | <0.001 |
| EMRO, n=21 | -0.043 | 0.852 | 0.081 | 0.729 |
| EURO, n=49 | -0.800 | <0.001 | -0.738 | <0.001 |
| SEARO, n=11 | -0.034 | 0.920 | -0.191 | 0.574 |
| WPRO, n=19 | -0.590 | 0.008 | -0.599 | 0.007 |
| The World Bank income | | | | |
| High, n=48 | -0.402 | 0.005 | -0.311 | 0.032 |
| Upper Middle, n=48 | -0.647 | <0.001 | -0.577 | <0.001 |
| Low Middle, n=47 | -0.425 | 0.003 | -0.418 | 0.003 |
| Low, n=30 | -0.166 | 0.381 | -0.104 | 0.586 |
| Other country groupings | | | | |
| Arab World, n=21 | -0.152 | 0.512 | -0.086 | 0.710 |
| English, Official Language, n=50 | -0.692 | <0.001 | -0.625 | <0.001 |
| OECD, n=33 | -0.539 | <0.001 | -0.204 | 0.255 |
| APEC, n=19 | -0.616 | 0.005 | -0.730 | <0.001 |
| ACD, n=27 | -0.447 | 0.019 | -0.373 | 0.055 |
| SADC, n=14 | -0.243 | 0.403 | -0.169 | 0.563 |

## Discussion

Incidence rates of all cancers and of most separate site cancers showed strong and significant correlation to reduced natural selection, measured by $I_s$. It is especially important that this relationship is strong in the younger part of the population which are in pre-reproductive (0-14) and reproductive (15-49) periods.

It is important that we list the limitations, including the intrinsic limitations (conceptualized as ecological fallacy) to this study, before examining the public health implications of our results. Firstly, the data included in this study were for whole nations, so we may only demonstrate the relationships between $I_s$ and cancer incidence at macro-level. We also need to note that it is nearly impossible to test such relationship at the individual or germline level due to very rare cancer occurrence rates. Secondly, data compiled and/or

collected by the major international agencies (WHO, IARC and the World Bank) are fairly crude, and may contain some random errors arising from methods of reporting incidence of specific diseases, reliability of diagnoses and possible administrative errors. Thirdly, not all the contributing factors, such as alcohol consumption, can be included as the potential confounders in data analysis due to data availability or quality. Furthermore, the opportunity for natural selection is only measured with respect to postnatal mortality, while gametic selection and intrauterine mortality are not included [54]. Despite these limitations the findings in this study from different data analysis methods constantly and consistently showed significant correlation between reduced natural selection and all cancer incidence (all ages and 0-49 respectively) and incidence of most of site specific cancer groups, especially those for which genetic background may be expected. Obviously, the changes in the genetic code of the human populations may not fully explain the increasing cancer incidence rate. These changes may be cumulative, each one of minor effect, and may contribute to increasing cancer incidence together with other carcinogenic factors.

Various genes contribute to cancer, e.g. proto-oncogenes can increase proliferation of mutated cells and tumour suppressor genes could inhibit self-regulation of abnormal cells, but their balance may still increase cancer incidence in various ways because these genes have pleiotropic effects. In this study, some of cancer groups have incidences that do not correlate with $I_s$ value, or even show reversed correlations (Table 1, Fig. 4). These include cancers of well-known viral causes – cervical cancer – immune problems related cancers like non-Hodgkin lymphoma, and cancers caused by toxins, like lip and oral cavity cancers. These cancers also have incidence “increases” below zero indicating their greater incidence in countries with larger opportunity for natural selection. This is most likely a result of countries with greater mortality having also poorer hygienic conditions and less medical services.

While specific genes determining risks of specific cancers may be still unknown, the general tendency is clear – relaxation of natural selection allows accumulation of detrimental genetic material, especially if single detrimental alleles have mild effects [7]. Studies have shown that a partially heritable disease, phenylketonuria was only noticeable after being accumulated for several generations [9] with about 2% increase each [55]. Two recent studies also reported that relaxed natural selection has been contributing to the increasing prevalence of two non-communicable diseases, obesity [19] and type 1 diabetes [18] because it may allow detrimental gene accumulation in

human population. However crudely calculated our rates of “incidence increase” (Table 1) indicate rates of increase compatible with alteration of mutation-selection balance. We only have at our disposal recent data, but it can be hypothesised that observed differences among countries in the opportunity for natural selection have existed for a few generations. With a simple accumulation of mutations under zero selection, the incidence rates should double every generation, when selection is not entirely relaxed, but still strongly limited, the increase will be somewhat less than double. Considering that declines in mortality in ‘developed’ countries started in the second half of the 19$^{th}$ century, we can estimate that changes in mutation-selection rates occurred over lifetime of some four, maybe five, generations. Incidence increases of all cancers (2.3-2.4) indicate approximately doubling over that time, while for cancers correlated significantly with $I_s$ the average increase is 5.7. Of course, not the entire incidence increase can be attributed to alteration in mutation-selection balance, because quality of data collection and reporting and presence of carcinogenic external factors may differ between the 10 countries of the lowest opportunity for selection and 10 countries of the highest selection opportunity. Our choice of 10 countries of each kind, instead of only 5 or 20, also influences precision of the numerical indices calculated. What is important here is that the order of magnitude of incidence increases, and their positive relationship to the relaxation of selection, especially in cancers with supposed genetic background, are compatible with expectations of population genetics. In short – such increases in the incidence of cancers are possible upon significant relaxation of natural selection through differential mortality.

Overall, cancer is an inheritable non-communicable disease due to its strong genetic background. Cancer genes may be cumulative at the reduced natural selection. Natural selection had an ample opportunity to eliminate defective genes introduced by mutations [9, 13, 17, 21, 56-59]. However, natural selection has been significantly reduced in the past 100-150 years, and the direct consequence of this process is that nearly every individual born into a population can pass genes to the next generation, while some 150 years ago, only 50% or less of individuals had this chance [13, 59]. Therefore, population allowing more people with cancer genes survive reproduction cycle may boost cancer gene accumulation. For instance, genetic predisposition to childhood leukaemia exists [60]. Patients who survive it will have a chance to pass this predisposition to the next generation. Similar argument may be made with respect to other cancers occurring during pre-reproductive or reproductive period of life. Currently used cancer treatments are not targeting genetic causes of the disease but dealing with its phenotypic expressions – tumours that are surgically removed, or metastatic cell masses whose

proliferation is curtailed by chemotherapy and radiation therapy. Although successful in a portion of cases, these treatments have side effects and do not deal directly with the cause of the disease, therefore, though undoubtedly helpful to a number of patients, they are not optimally effective.

Table 3 showed that country groupings with higher socioeconomic level had stronger associations between $I_s$ and cancer incidence. This finding is consistent with the studies conducted by the WHO cancer research agent, IARC [22, 37, 38, 61]. Similarly, reduced natural selection and type 1 diabetes prevalence also showed stronger association in developed regions [18]. One of the explanations may be that people in developed regions, such as Europe and Americas have been able to access better health services, which has made them to escape natural selection more often and pass their detrimental genes onto their next generation. The long effect from escaping natural selection may allow those genes, including cancer related genes, to accumulate in those populations faster [9, 18, 55].

The association between $I_s$ and cancer incidence was strong and significant in both Upper Middle and High income economic classifications (the World Bank). However, it was stronger in Upper Middle income economic classification. This may be attributable to: 1) Almost all people in the countries in High income grouping may be able to escape natural selection due to high level of health services. This is shown by the extremely low $I_s$ values of these countries, which are close to 0 (Supplemental Document 1). 2) Fast developing GDP in Upper Middle country grouping has driven their medicine level to develop quickly, which may have made more and more people escape natural selection.

## Conclusion

Assuming that the increasing genetic load underlies cancer incidence as one of the contributing factors, the only way to reduce it remains genetic engineering – repair of defective portions of the DNA or their blockage by methylation and similar approaches. These techniques, though theoretically possible, are not yet practically available. They will, however, need to be developed since they provide the only human-made alternative to the disappearing action of natural selection since any eugenics-like approaches are ethically and morally reprehensible.

**Acknowledgments**

The authors express appreciation to Jacques Ferlay from the International Agency for Research on Cancer of WHO for the assistance in locating and defining the data.

**Data Archiving Statement**

All data for this study are publicly available and are ready for the public to download at no cost from the official websites of the World Bank, the WHO and FAO. Use of these data for this research falls within these UN agency's public permission in their terms and conditions. There is no need to have the formal permission to use the data for this study.

## References


1. Global Burden of Disease Cancer, C., et al., *The Global Burden of Cancer 2013.* JAMA Oncol, 2015. **1**(4): p. 505-27.
2. Croce, C.M., *Oncogenes and Cancer.* The New England Journal of Medicine, 2008. **358**(5): p. 502-511.
3. Vogelstein, B. and K.W. Kinzler, *Cancer genes and the pathways they control.* Nat Med, 2004. **10**(8): p. 789-99.
4. Tian, H., et al., *DNA damage response – A double- edged sword in cancer prevention and cancer therapy.* Cancer Letters, 2015. **358**(1): p. 8-16.
5. Conrad, D.F., et al., *Variation in genome- wide mutation rates within and between human families.* Nature Genetics, 2011. **43**(7): p. 712.
6. Crow, J.F., *The origins, patterns and implications of human spontaneous mutation.* Nature Reviews Genetics, 2000. **1**(1): p. 40-47.
7. Henn, B.M., et al., *Estimating the mutation load in human genomes.* Nature Reviews Genetics, 2015. **16**(6): p. 333-343.
8. Lynch, M., *Mutation and Human Exceptionalism: Our Future Genetic Load.* Genetics, 2016. **202**(3): p. 869-75.
9. Stephan, C.N. and M. Henneberg, *Medicine may be reducing the human capacity to survive.* Medical Hypotheses, 2001. **57**(5): p. 633-37.
10. Bodmer, W. and L. Cavalli-Sforza, *Genetoics, Evolution and Man, p289-305*. 1976: Freeman Co. San Francisco.
11. Crow, J.F., *Some possibilities for measuring selection intensities in man.* Human biology, 1958. **30**(1): p. 1.
12. Rühli, F.J. and M. Henneberg, *New perspectives on evolutionary medicine: the relevance of microevolution for human health and disease.* BMC medicine, 2013. **11**(1): p. 115.

13. Rühli, F. and M. Henneberg, *Biological future of humankind – ongoing evolution and the impact of recognition of human biological variation*, in *On Human Nature. Biology, Psychology, Ethics, Politics, and Religion 2016*, M. Tibayrenc and F.J. Ayala, Editors. 2016, Elsevier. p. 263-275.

14. Henneberg, M., *The rate of human morphological microevolution and taxonomic diversity of hominids.* Studies in Historical Anthropology, 2006. **4**(2004): p. 49-59.

15. Fisher, R.A.S., *The genetical theory of natural selection : a complete variorum edition / by R.A. Fisher ; edited with a foreword and notes by J.H. Bennett*, ed. J.H. Bennett. 1999, Oxford: Oxford : Oxford University Press.

16. Henneberg, M. and J. Piontek, *Biological state index of human groups.* Przeglad Anthropologiczny, 1975. **XLI**: p. 191-201.

17. Henneberg, M., *Reproductive possibilities and estimations of the biological dynamics of earlier human populations.* Journal of Human Evolution, 1976. **5**: p. 41-8.

18. You, W.-P. and M. Henneberg, *Type 1 diabetes prevalence increasing globally and regionally: the role of natural selection and life expectancy at birth.* BMJ Open Diabetes Research & Care, 2016. **4**(1): p. e000161.

19. Budnik, A. and M. Henneberg, *Worldwide Increase of Obesity is Related to the Reduced Opportunity for Natural Selection.* PLOS One, 2017. **12**(1).

20. Henneberg, M., *Natural selection through differential fertility in human populations: Quantitative evaluation of selection intensity.* Przeglad Antropologiczny 1980. **46**: p. 21-60.

21. Saniotis, A. and M. Henneberg, *Medicine could be constructing human bodies in the future.* Medical Hypotheses, 2011. **77**(4): p. 560-64.

22. Ferlay, J., et al. *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]*. 2013 28.05.2016]; Available from: http://globocan.iarc.fr.

23. Ramazzini, B.a., *De morbis artificum Bernardini Ramazzini diatriba = Diseases of workers : the Latin text of 1713 revised, with translation and notes / by Wilmer Cave Wright*. Special edition / privately printed for the members of the Classics of Medicine Library. ed. Diseases of workers, ed. W.C.e.t. Wright and i.b. Classics of Medicine Library. 1983: Birmingham, Alabama : The Classics of Medicine Library.

24. Britt, K. and R. Short, *The plight of nuns: hazards of nulliparity.* The Lancet, 2012. **379**(9834): p. 2322-2323.

25. MacMahon, B., et al., *Age at first birth and breast cancer risk.* Bulletin of the World Health Organization, 1970. **43**(2): p. 209-221.

26. The United Nations. *World Fertility Data 2008*. 2012 29.07.2015]; Available from: http://www.un.org.

27. WHO, *World Health Statistics 2012*. Life tables for WHO Member States. 2012, Geneva: World Health Organization.

28. Grafen, A., *A theory of Fisher's reproductive value.* Journal of mathematical biology, 2006. **53**(1): p. 15-60.

29. WHO. *Global Health Observatory, the data repository*. WHO 2015 [11.26.2015]; Available from: http://www.who.int/gho/database/en/.

30. Allender, S., et al., *Quantification of urbanization in relation to chronic diseases in developing countries: a systematic review.* J Urban Health, 2008. **85**(6): p. 938-951.

31. Moore, M., P. Gould, and B.S. Keary, *Global urbanization and impact on health.* International journal of hygiene and environmental health, 2003. **206**(4): p. 269-278.

32. WHO. *Urbanization and health*. WHO 2010 2010-12-07 15:20:05 2 November 2016]; Available from: http://www.who.int/bulletin/volumes/88/4/10-010410/en/.

33. Blot, W.J., J.F. Fraumeni, and B. Stone, *Geographic patterns of breast cancer in the United States.* Journal of the National Cancer Institute, 1977. **59**(5): p. 1407-1411.

34. Nasca, P.C., M.C. Mahoney, and P.E. Wolfgang, *Population density and cancer incidence differentials in New York State, 1978–82.* Cancer Causes & Control, 1992. **3**(1): p. 7-15.

35. Greenberg, M., *Urbanization and Cancer: Changing Mortality Patterns?* International Regional Science Review, 1983. **8**(2): p. 127.

36. The World Bank Group. *World Bank Open Data*. 2016 12.07.2016]; Available from: http://data.worldbank.org/.

37. Jemal, A., et al., *Global cancer statistics.* CA Cancer J Clin, 2011. **61**(2): p. 69-90.

38. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* Int J Cancer, 2015. **136**(5): p. E359-86.

39. United Nations-Department of Economic and Social Affairs-Population Division. *Abridged life table, for females, by major area, region and country, 1950-2100*. 2013 18.07.2016].

40. Breastcancer.org. *Risk of Developing Breast Cancer* 2016 22.10.2016]; Available from: http://www.breastcancer.org.

41. Majeed, A., et al., *Trends in prostate cancer incidence, mortality and survival in England and Wales 1971–1998.* BJU International, 2000. **85**(9): p. 1058-1062.

42. International Agency for Research on Cancer (IARC). *DataSource and methods*. 2012; Available from: http://globocan.iarc.fr/Pages/DataSource_and_methods.aspx.

43. McPherson, K., C. Steel, and J. Dixon, *Breast cancer—epidemiology, risk factors, and genetics.* Bmj, 2000. **321**(7261): p. 624-628.

44. American Cancer Society. *Breast cancer facts & figures 2015-2016*. Breast cancer facts and figures 2015 2 November 2016]; Available from: http://www.cancer.org.

45. Colditz, G.A. and E.K. Wei, *Preventability of cancer: the relative contributions of biologic and social and physical environmental determinants of cancer mortality.* Annual review of public health, 2012. **33**: p. 137.

46. Danaei, G., et al., *Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors.* The Lancet, 2005. **366**(9499): p. 1784-1793.

47. WHO. *WHO regional offices*. [11.26.2015]; Available from: http://www.who.int.

48. The World Bank. *Arab World | Data*. 2015; Available from: http://data.worldbank.org.

49. The OECD. *List of OECD Member countries*. 2015; Available from: http://www.oecd.org.

50. Asia-Pacific Economic Cooperation. *Member Economies-Asia-Pacific Economic Cooperation*. [11.26.2015]; Available from: http://www.apec.org.

51. Asia Cooperation Dialogue. *Member Countries*. Available from: http://www.acddialogue.com.

52. South Africa Development Community. *Southern African Development Community: Member States*. [18.06.2015]; Available from: http://www.sadc.int.

53. (IARC), I.A.f.R.o.C. *GLOBOCAN Cancer Fact Sheets: Cervical cancer*. 2012; Available from: http://globocan.iarc.fr/old/FactSheets/cancers/cervix-new.asp.

54. Rühli, F. and M. Henneberg, *Biological future of humankind – ongoing evolution and the impact of recognition of human biological variation*, in *On Human Nature. Biology, Psychology, Ethics, Politics, and Religion*, M. Tibayrenc and F.J. Ayala, Editors. 2016(accepted 28.12.2015), Elsevier.

55. Medawar, P.B., *Do advances in medicine lead to genetic deterioration?*, in *Natural Selection in Human Populations.*, Bajema C. J., Editor. 1971, Robert E. Krieger Publishing Co.: New York. p. 300-08.

56. Budnik, A., G. Liczbińska, and I. Gumna, *Demographic trends and biological status of historic populations from Central Poland: The Ostrów Lednicki microregion.* American Journal of Physical Anthropology, 2004. **125**(4): p. 369-381.

57. Henneberg, M. and R. Henneberg, *Reconstructing medical knowledge in ancient Pompeii from the hard evidence of bones and teeth.* 2002.

58. Henneberg, M. and R. Henneberg, *Biological characteristics of the population based on analysis of skeletal remains.* 1998.

59. Saniotis, A. and M. Henneberg, *Examining genetic load: an Islamic perspective.* Medical Journal of Islamic World Academy of Sciences, 2012. **20**(3): p. 73-80.

60. Stieglitz, E. and M.L. Loh, *Genetic predispositions to childhood leukemia.* Ther Adv Hematol., 2013. **4**(4): p. 270–290.

61. Ferlay, J., et al., *Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008.* Int J Cancer, 2010. **127**(12): p. 2893-917.

Article 2/10: Type 1 diabetes prevalence increasing globally and regionally: the role of natural selection and life expectancy at birth (Published at BMJ Open Diabetes Research and Care, 2016)

Wen-Peng You[1]*, Maciej Henneberg[1,2]

[1] School of Medicine, The University of Adelaide, Adelaide, SA, Australia 5005

[2] Institute of Evolutionary Medicine, University of Zürich, Switzerland

**Published:** Wenpeng You, Maciej Henneberg. *Meat consumption providing a surplus energy in modern diet contributes to obesity prevalence: an ecological analysis*. BMC Nutrition, 20162:22. DOI: 10.1186/s40795-016-0063-9

✉ **Correspondence:** Wenpeng You, Wenpeng.you@adelaide.edu.au

## Contextual Statement

Natural selection is the key mechanism of evolution, which changes heritable traits characteristic of a population over generations.

In the past 100-150 years, the advanced medical care has reduced natural selection more than previously thought. It has been like a double-edged sword. It saves the lives of those people with type 1 diabetes which is featured with strong heredity. Meanwhile, it offers the opportunities for those people to pass on their genes of type 1 diabetes to their next generation. With 4-5 successive generations subject to the reduced natural selection, the phenotype of the accumulated type 1 diabetes genes/mutations may be noticeable.

We hypothesized and tested that reduced natural selection (indexed by opportunity for reproduction at population level) may be an independent predictor of type 1 diabetes.

## Statement of Authorship

# Statement of Authorship

| | |
|---|---|
| Title of Paper | Type 1 diabetes prevalence increasing globally and regionally: the role of natural selection and life expectancy at birth |
| Publication Status | ☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Publication Style |
| Publication Details | Wenpeng You, Maciej Henneberg. Type 1 diabetes prevalence increasing globally and regionally: the role of natural selection and life expectancy at birth. BMJ Open Diabetes Research and Care, 2016;4:e000161. doi:10.1136/bmjdrc-2015-000161 |

### Principal Author

| | | | |
|---|---|---|---|
| Name of Principal Author (Candidate) | Wenpeng You | | |
| Contribution to the Paper | Conceived the hypothesis, collected data, performed analysis on all samples, interpreted data, wrote manuscript and acted as corresponding author. | | |
| Overall percentage (%) | 60 | | |
| Signature | | Date | 22/12/2017 |

### Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

| | | | |
|---|---|---|---|
| Name of Co-Author | Maciej Henneberg | | |
| Contribution to the Paper | Supervised development of work, formalized and interpreted data analysis, and evaluated and edited manuscript | | |
| Signature | | Date | 22/12/2017 |

| | | | |
|---|---|---|---|
| Name of Co-Author | NA | | |
| Contribution to the Paper | | | |
| Signature | | Date | |

Please cut and paste additional co-author panels here as required.

## Abstract

**Objective:** Prevalence of type 1 diabetes disease (T1D) is increasing worldwide. We aim to test correlation of T1D prevalence to the reduced natural selection measured by Biological State Index ($I_{bs}$).

**Research Design and Methods:** Country-specific estimates of type 1 diabetes prevalence, life expectancy, obesity prevalence rate, urbanization rates, per capita sugars consumption and per capita GDP were obtained. $I_{bs}$ and country-specific longevity ($e_{50}$) increase for each country were self-calculated. These data were then matched to T1D prevalence by country for our ecological study among 118 countries. Countries were also grouped to study the associations in different regions. SPSS v. 22 was used for correlation analysis.

**Results:** Worldwide, both $I_{bs}$ and life expectancy at birth ($I_{bs}$ proxy) were significantly correlated to T1D prevalence in Pearson r ($r=0.713$, $p<0.001$ and $r=0.722$, $p<0.001$ respectively) and Spearman's rho ($r=0.724$, $p<0.001$ and $r=0.689$, $p<0.001$ respectively). T1D prevalence was not correlated to longevity increase measured as life expectancy at 50 years old. T1D prevalence was significantly associated with $I_{bs}$ ($r=0.307$, $p<0.001$) and newborn life expectancy ($r=0.349$, $p<0.001$) independent of per capita total sugar consumption, per capita GDP, urbanization and obesity prevalence in partial correlation. Globally, both life expectancy at birth and $I_{bs}$ exponentially correlated to T1D prevalence. Pearson correlations generally existed in different country categorizations by geographic region, culture background and economic status.

**Conclusions:** Reduced natural selection may have contributed to the increasing T1D prevalence worldwide. T1D epidemiology study in total population may be the practical solution to identify the causes of increasing T1D prevalence.

**Key Words**: Type I diabetes, Biological State Index, Epidemiology, Life expectancy, Insulin

## Background

Type 1 diabetes (T1D) is an autoimmune disease with a strong genetic component [1, 2]. It can occur at any age, but tends to develop in childhood [3], so it has long been called "juvenile diabetes". T1D is characterized by destruction of pancreatic beta cells, culminating in absolute insulin deficiency [4]. As of 2014, an estimated 387 million people have diabetes worldwide [5], of which T1D accounts for between 5% and 10% [6]. Diabetic complications continue to be a major cause of morbidity and mortality in persons

with T1D [7]. Great efforts have been made to assess the incidence and prevalence of T1D. Unfortunately, the exact etiology and pathogenesis of T1D is still unknown. Generally, longitudinal or cross-sectional studies are often locally or regionally performed. Consequently, it is difficult to access generalizable results because the epidemiology of T1D is known to be heterogeneous regarding geography and ethnicity. Genetic predisposition to T1D is only alleged to explain some of the geographic variability in T1D occurrence, but it cannot account for its rapidly increasing frequency [8]. A number of studies have associated gross domestic product (GDP) level with T1D prevalence or incidence [8-11], but GDP does not fully explain variations and trends in T1D prevalence rates observed in many countries, for example Japan. It has been postulated that environmental factors may be able to trigger an autoimmune destruction of the beta cells leading to absolute dependence on insulin treatment [8, 9, 12-17], however, these environmental factors are circumstantial [13].

Natural selection, as one of the basic mechanisms of evolution, is the differential survival and fertility of individuals due to differences in phenotype that reflect genetic differences. In our modern society, natural selection still acts on all members of a population, selecting those individuals that have an increased reproductive success (survival and/or fertility) [18]. The "Biological State Index ($I_{bs}$)" has been proposed to measure the populational reproductive success by taking into account potential loss of reproductive success by dying at age x, summed over all age categories [19, 20]. The $I_{bs}$ is calculated by combining age-specific death frequency (dx variable of a life table) with an age-specific reproductive loss ($s_x$):

$$I_{bs} = 1 - \sum_{x=0}^{x=\omega} \mathrm{dx.Sx}$$

Where: dx is the frequency of death at age x or represents the mortality rate. $s_x$ is the reproductive loss from dying at age x is measured $s_x$, i.e. the estimated probability of not possessing the complete number of births at age x. $s_x$ is based on the cumulative number of births at specific ages [20, 21]. The construction and interpretation of the $I_{bs}$ was predicated upon the assumption that heritability of human fertility variance is negligible [22].

An $I_{bs}$ value of one indicates total adaptation of the population to their environment (ability to overcome selection pressures that are present). An $I_{bs}$ value of zero signifies a total lack of adaptation (inability to overcome selection pressures that are present), and an impossibility to give life to the next generation. An $I_{bs}$ value close to zero indicates large effective natural selection pressures acting on a population, since few individuals are

surviving to produce offspring. In such a scenario there is a possibility for fast evolution, since many genes may not be passed to the next generation. An $I_{bs}$ value close to one indicates that natural selection is not having much effect on the population since many individuals are able to maximally contribute to producing the next generation. Thus, the $I_{bs}$ permits the estimation of the magnitude of the successful reproduction of a population.

The genetic trait of T1D may allow individuals from a population to pass their T1D genes on to their next generation. What fraction of a population had a chance to fully participate in reproducing under a given set of mortality conditions may be associated with the proportion of population carrying T1D genes in the next generation. Previously, Stephan and Henneberg [23] raised a concern that the developed populations may have accumulated more unfavourable genes, such as T1D genes because natural selection (measured by the $I_{bs}$) has been greatly reduced. Recently, a systematic review concluded that the T1D prevalence rate was associated with age increase in population [3]. Therefore, the objective of the current paper was to use country specific data to test, from a global perspective, that population with greater $I_{bs}$ value, fuelled by life expectancy at birth, may have higher T1D prevalence, using empirical, macro-level data collected from the major international organizations.

## Material and Method

The dependent variable in the analysis was the country-specific estimate of T1D prevalence (sum of rates in 0-14 and 15+ years old groups) which were published by the International Diabetes Federation (IDF) in 2000 [24]. We used $I_{bs}$ and life expectancy at birth of each country as the independent variables. The $I_{bs}$ was calculated as proposed by Henneberg [20] and Henneberg and Piontek [19] with the fertility data of each country published by United Nations in 2008 [25] and the mortality data of life tables (2009) published by World Health Organization (WHO) in 2012 [26]. Life expectancy at birth is a proxy for $I_{bs}$ since it expresses an opportunity of a newborn to survive to a specific age, which is usually an age falling into reproductive life span or above it. We extracted the country specific life expectancy (years) at birth by country published by WHO in 2013 [27].

GDP [8-11], urbanization [8] and body weight status [17] have been associated with T1D prevalence. It has been suggested that a population with greater sugar consumption may have greater diabetes (total) prevalence [28] and that sugar consumption may affect health of T1D patients [29]. Therefore, we controlled for market availability of sugars and sweeteners (sugars in short) in g/capita/day in each country for year 2010 from United

Nations Food and Agricultural Organization (FAO) Food Balance Sheet, per capita gross domestic product (GDP expressed in purchasing power parity in 2010 US dollars for comparability among countries) and urbanization (percentage of population living in urban areas in each country in 2010) from the World Bank World Development Indicators Database, and obesity prevalence rate (percentage of the population aged 18+ years old with body mass index greater than or equal to 30 kg/m2 in each country 2010) from the WHO.

No ethical approval or written informed consent for participation was required in this study as all the aforementioned data were freely downloaded from the United Nations agents' websites.

We aligned all independent variables and confounding factors with the country-specific T1D prevalence and obtained a set of data for 118 countries. All country specific data were put in a uniform format. Each country was treated individually, and all of their available information was analysed. Data of calculated $I_{bs}$ and summary statistics are further described in the Supporting Information (Tables S1 & S2).

In order to demonstrate the universal association between T1D prevalence and $I_{bs}$ and life expectancy at birth respectively in different country groupings, we categorized the countries for correlation analyses based on 1) the WHO regional classifications [30]; 2) the strong contrast in terms of geographic distributions, per capita GDP levels and/or cultural backgrounds. We analysed the correlation in the six country groupings: Latin America and the Caribbean (LAC) [31], the Arab World [32], the Organisation for Economic Co-operation and Development (OECD) [33], European Economic Area (EEA) [34], Asia Cooperation Dialogue (ACD) [35] and the Asia-Pacific Economic Cooperation (APEC) [36]. In our analysis, we only included those countries for which we could access their data for the specific groupings.

To a large extent, grouping countries for analysis may also allow us to align our findings against previous local or regional studies regarding heterogeneous T1D epidemiology due to various geographic location and ethnicity.

It might be considered that T1D prevalence is a result of the increase in longevity rather than relaxed natural selection of the genetic background of T1D since general health and advances in medical care improve survival of T1D patients. This consideration is clarified by much stronger correlation between T1D prevalence and $I_{bs}$ than that between T1D and life expectancy (e50) increase (across two periods of 1950-55 and 2005-10). The rationale to use country specific life expectancy (e50) is 1) that the estimate of life

expectancy based on this segment of the population may not be biased with child (0-15 years) mortality, in particular due to deaths caused by T1D disease, and 2) that deaths of adults, especially females, during the reproductive life span (15-50 years) that may differentiate numbers of T1D genes passed on to new generations. Thus, we obtained country specific life expectancies ($e_{50}$) for the periods of 1950-55 and 2005-10 respectively from the WHO life tables [37]. And then we calculated the life expectancy increase from the period of 1950-55 to the period of 2005-10 for each country producing a new variable, which is "life expectancy increase ($e_{50}$, 1950-2010)" for each country across the 55 years. We repeated the above correlation analysis after we replaced the variable of "life expectancy at birth" with life expectancy increase ($e_{50}$, 1995-2010).

Pearson's correlation coefficient, Spearman's rho and partial correlation analyses were conducted using SPSS v. 22 (SPSS Inc., Chicago Il USA). In this study, the data were log transformed for correlation analysis in SPSS.

## Results

Worldwide T1D prevalence was in significant associations with $I_{bs}$ using Pearson r ($r=0.713$, $p<0.001$) and Spearman's rho ($r=0.724$, $p<0.001$) respectively. The similar associations were also observed between T1D and life expectancy in Pearson model ($r=0.722$, $p<0.001$) and Spearman's model ($r=0.689$, $p<0.001$) respectively (Table 1). Further investigation with partial correlation analysis showed that worldwide the association between T1D prevalence was still strongly associated with $I_{bs}$ ($r=0.307$, $p<0.001$) and life expectancy ($r=0.349$, $p<0.001$) when we controlled for per capita total sugars availability, per capita GDP, urbanization and obesity prevalence (Table 1). All confounders were in significant associations with T1D prevalence rate in both Pearson r and Spearman's rho.

Table 1: Global associations between T1D prevalence rate and $I_{bs}$ and life expectancy (years) at birth respectively

| | Pearson's r | | Spearman's rho | | Partial Correlation | |
|---|---|---|---|---|---|---|
| | R | n | r | n | r | df |
| Log $I_{bs}$ | 0.713 | 118 | 0.724 | 118 | 0.307 | 103 |
| Log Life Expectancy at birth | 0.722 | 118 | 0.689 | 118 | 0.349 | 103 |
| Log Sugars per capita | 0.666 | 109 | 0.534 | 109 | - | - |
| Log GDP per capita | 0.720 | 116 | 0.749 | 116 | - | - |
| Log BMI ≥ 30 prevalence | 0.636 | 109 | 0.538 | 118 | - | - |
| Log Urbanization | 0.507 | 118 | 0.567 | 118 | - | - |

** All correlations are significant at the 0.001 level (2-tailed).*

*** Keeping GDP, BMI, urbanization and sugars intake constant.*

Globally, T1D was is noted to be exponentially related with both $I_{bs}$ ($R^2$=0.5203) and life expectancy ($R^2$=0.5302) (Figures 1 & 2) after we removed two outliers of extremely high T1D prevalence in Finland and Sweden respectively.

Figure 1 Relationship between $I_{bs}$ and type 1 diabetes prevalence worldwide

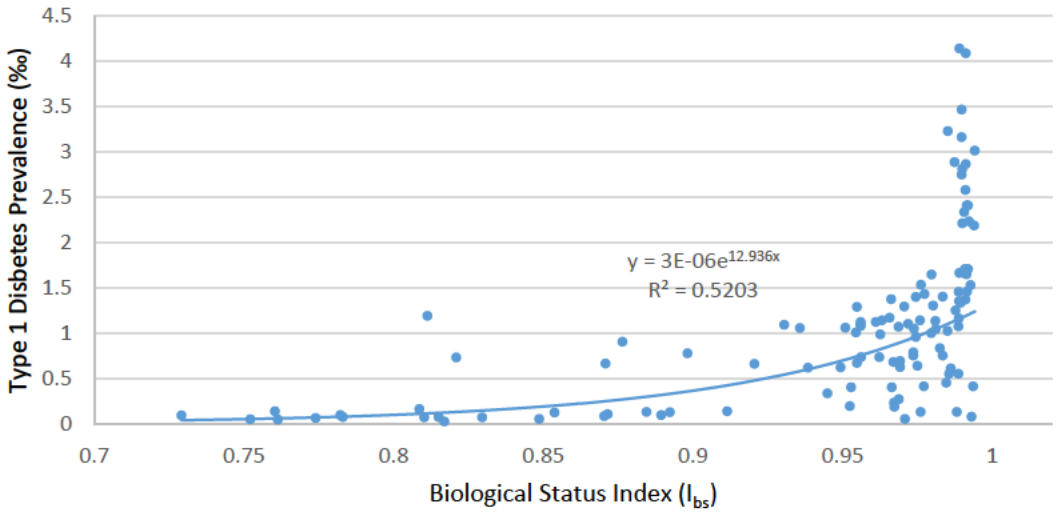
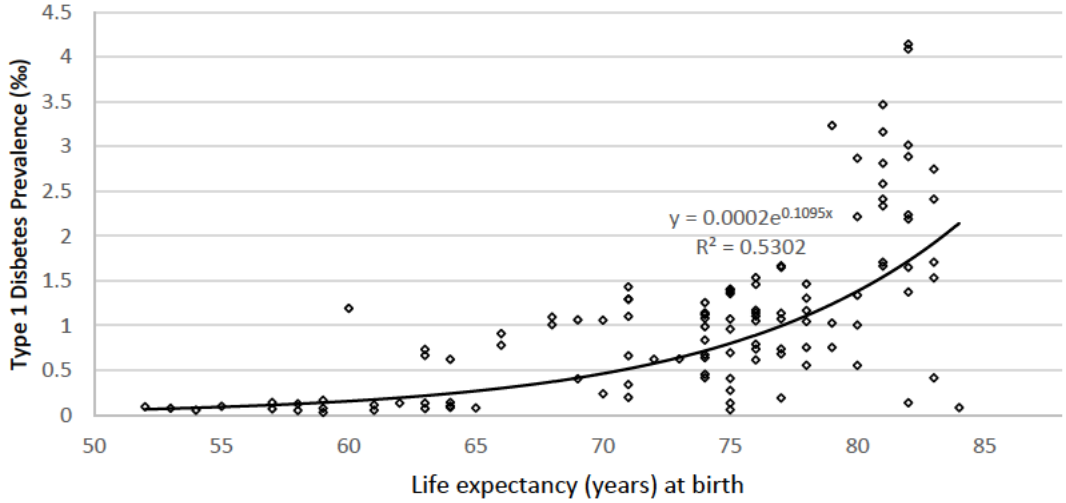



Figure 2 Relationship between life expectancy (years) at birth and type 1 diabetes prevalence worldwide




Worldwide, strong positive Pearson r coefficients were observed in the correlation of $I_{bs}$ to life expectancy (proxy of $I_{bs}$, r=0.908, p<0.001) and to GDP (r=0.781, p<0.001).

Table 2 presented that Europe had the significant associations with $I_{bs}$ (r=0.502, p=0.001) and life expectancy (r=0.610, p<0.001) respectively. The sub-Europe (EEA) also had the very strong associations of T1D prevalence to both $I_{bs}$ (r=0.479, p=0.009) and life expectancy (r=0.574, p=0.001) (Table 3). We observed the slightly negative correlations of T1D prevalence rate to the $I_{bs}$ and life expectancy respectively in two Asian country groupings, SEARO (South-East Asia) and the ACD (Table 3). Both $I_{bs}$ and life expectancy were insignificantly associated with T1D prevalence rate in other four WHO regions, AFRO, AMRO, EMRO and WPRO (Table 2).

Table 2 Associations between T1D prevalence and $I_{bs}$ and life expectancy (years) at birth respectively in the WHO regions

| | $I_{bs}$ | | Newborn Life Expectancy | |
|---|---|---|---|---|
| WHO Region | Pearson's r | Significance | Pearson's r | Significance |
| AFRO (n=21) | 0.343 | 0.128 | 0.214 | 0.351 |
| AMRO (n=28) | 0.145 | 0.461 | 0.173 | 0.380 |
| EMRO (n=15) | 0.783 | 0.001 | 0.541 | 0.037 |
| EURO (n=38) | 0.502 | 0.001 | 0.610 | <0.001 |
| SEARO (n=5) | -0.479 | 0.415 | -0.436 | 0.463 |
| WPRO (n=11) | 0.345 | 0.298 | 0.330 | 0.322 |

Table 3 showed that the T1D prevalence rate was associated with both the $I_{bs}$ and newborn life expectancy in the groupings consisting of countries with similar cultures (Arab World and EEA) and different cultures (APEC and OECD), with similar economy status (OECD) and those with economy status in disparity (APEC), and with the heterogeneous region (LAC) and the homogeneous area (EEA and ACD).

Table 3 Associations between T1D prevalence rate and both the $I_{bs}$ and life expectancy (years) at birth in different country groupings categorized considering their socio-economic, geographic location and cultural backgrounds

| | $I_{bs}$ | | Life Expectancy at birth | |
|---|---|---|---|---|
| Country Grouping | Pearson's r | Significance | Pearson's r | Significance |
| APEC (n=16) | 0.340 | 0.197 | 0.369 | 0.160 |
| Arab World (n=13) | 0.748 | 0.003 | 0.469 | 0.106 |
| EEA (n=29) | 0.479 | 0.009 | 0.574 | 0.001 |
| ACD (n=20) | -0.268 | 0.253 | -0.392 | 0.087 |
| OECD (n=34) | 0.365 | 0.034 | 0.155 | 0.382 |
| LAC (n=26) | 0.524 | 0.006 | 0.044 | 0.831 |

Table 4 showed that in general T1D prevalence is not correlated to longevity increase, but strongly correlated to $I_{bs}$ at country level. Life expectancy increase (e50, 1950-2010) does not correlate significantly with T1D prevalence in Pearson correlation (r=0.165, p=0.079) or Spearman's rho (r=0.166, p=0.077). These contrasted with the correlation between $I_{bs}$ and T1D prevalence in Pearson correlation (r=0.713, p<0.001) and in Spearman rho (r= 0.724, p<0.001) respectively. In partial correlation analysis, when we controlled for GDP, BMI≥30, urbanization and sugars intake, T1D was negatively and insignificantly correlated to life expectancy increase (e50, 1950-2010) (r=-0.070, p=0.487), but it was in strong and significant correlation to $I_{bs}$ (r=0.276, p=0.005).

Table 4: Comparing correlations between type 1 Diabetes prevalence and $I_{bs}$ and life expectancy increase (e50)

| | Pearson's r | | | Spearman's rho | | | Partial Correlation* | | |
|---|---|---|---|---|---|---|---|---|---|
| | r | p | n | r | p | n | r | p | df |
| Log $I_{bs}$ | 0.713 | <0.001 | 118 | 0.724 | <0.001 | 118 | 0.276 | 0.005 | 98 |
| Log Life expectancy (e50) increase | 0.165 | 0.079 | 114 | 0.166 | 0.077 | 114 | -0.070 | 0.487 | 98 |
| Log Sugars per capita | 0.666 | <0.001 | 109 | 0.534 | <0.001 | 109 | - | - | - |
| Log GDP per capita | 0.720 | 0.720 | 116 | 0.749 | <0.001 | 116 | - | - | - |
| Log BMI ≥ 30 prevalence | 0.636 | <0.001 | 109 | 0.538 | <0.001 | 118 | - | - | - |
| Log Urbanization | 0.507 | <0.001 | 118 | 0.567 | <0.001 | 118 | - | - | - |

## Discussion

The worldwide trend of increased T1D prevalence likelihood has multiple etiologies, which may act through multiple mechanisms. By assessing the T1D prevalence rate data for 118 countries we have shown that globally and regionally population which had greater value of $I_{bs}$ (less opportunity for natural selection) may have greater T1D prevalence and secondly, that newborn life expectancy was significantly associated with T1D prevalence rate at population level.

Overall, the operation of natural selection on contemporary populations is declining due to modern medicine [23], but the magnitude of the decline may differ between countries due to their specific level of sanitation, medical interventions and public health measures. Natural selection is still one of the major evolutionary forces that informs changes in gene frequencies in a population through the action of differential fertility and mortality over generations [38]. For example, studies have shown that the increasing prevalence rates of a partially heritable disease, nasal septa and lacrimal bone defects may be attributed to the decreasing effect of natural selection [39]. More than 40 genetic loci located in different chromosomes have been associated with T1D in multiple studies [1, 2]. Although T1D can be fatal, the majority of genetically predisposed people do not develop T1D [40]. This allows for accumulation of genetic predisposition in human populations. This accumulation will increase when fewer persons who developed a disease would die. Differential fertility and mortality are the basic events of natural selection, which operate singly or jointly to determine the fitness (reproductive success) of a particular population in a given environment [38]. The country specific fertility and mortality based $I_{bs}$ at different levels may indicate their different successful reproduction opportunities of individuals in the succeeding countries [23]. The reproduction success opportunity of each population may determine their magnitude of T1D genes accumulation, thus influenced prevalence rate of T1D patients in their next generations. In the present study, the correlation of $I_{bs}$ to the T1D prevalence rate has been observed, which was compatible with suggestion that lower opportunity for selection allows accumulation of unfavourable genes [23, 41]. Our analysis of correlations between T1D prevalence and $I_{bs}$ by region or by WHO grouped countries seem to indicate that in regions where insulin was available earlier and that had better availability of health care the relationship is stronger. This provides the analogue of a snapshot what could happen at different times in the same region as time from insulin introduction and improvement of health care increased. Thus, the distribution across different populations could be interpreted as a surrogate measure of the evolution in time of T1D prevalence after the introduction of insulin. Artificial insulin introduced for T1D treatment and increasing insulin availability

may have played a key role in reducing natural selection as insulin enables countless people with onset T1D to survive [3] and maintain normal reproductive capacity [42]. This may have been boosting T1D genes accumulation and prevalence of T1D. T1D can affect people of any age, but usually occurs in children or sexually mature young adults [3] who have greater potential to reproduce than older adults. T1D has been historically, and continues to be, the most common type of diabetes in children and adolescents [43]. Insulin is the priority for T1D treatment. Otherwise, T1D patient may only live up to one year, some only a week. Several human generations have benefited from insulin since it was discovered and became available in early 1920s [44]. Reduced natural selection boosted by insulin treatment of several generations may have enabled cumulative effect of T1D genes frequency in human population to occur quickly and to be noticeable for a couple of decades [9]. Studies have shown that a partially heritable disease, phenylketonuria was only noticeable after being accumulated for several generations [23] with about 2% increase each [45].

T1D prevalence/incidence is increasing worldwide [46] with special regard to the developed countries [9, 10][47]. This may be partially attributable to earlier and greater affordability of insulin, in addition to relative more reduced natural selection (greater $I_{bs}$ values) in those developed countries. Although exogenous insulin can be obtained from animals (bovine and porcine) [44], production, transportation, storage and administration of such insulin was extremely expensive [48], which may be beyond the affordability of many T1D patients, especially those from developing countries. Biosynthetic insulin based on DNA technology has been commercially available since 1982 [44] and it has been thought that it can continue to accommodate global demand [44] because of low cost from the production to administration. However, unfortunately life-saving insulin is still less accessible, affordable, or both to people diagnosed with diabetes in a developing country than their counterparts in the developed world [49]. This lower survivorship of T1D patients may contribute to lower prevalence figures directly, besides the fact that less predisposing genes have accumulated in the gene pools of those countries.

Our study showed that the relationship between life expectancy and T1D prevalence rate was exponential (Figure 2, $R^2$=0.5266). The Australian Institute of Health and Welfare (AIHW) also indicated the exponential relationship between T1D prevalence rate between age increase of Australian population through the Australian National Diabetes Register [50]. Additionally, Neville *et al.* reported that the increased longevity of diabetic patients contributed to the increasing prevalence of diabetes in Japanese population [51]. The life expectancy gap between patients with T1D and non-diabetic people has reduced

significantly [52] due to developments in sanitation, medical interventions and public health measures. Therefore, the underlying reason for the exponential relationship in our study may be because the number of individual T1D patients have increased in the human population [3]. The American Diabetes Association has also stated that the majority of individuals with T1D are adults even though T1D has been more frequent and a relatively straightforward diagnosis in children [53].

The correlations of T1D prevalence rates to both $I_{bs}$ and life expectancy were not only observed worldwide, but also in different country groupings sharing specific characteristics like geographic locations (Table 2), culture backgrounds (Table 3) and affiliations to international functional organizations (Table 3). Results' highlights indicated that the correlations of $I_{bs}$ and life expectancy to T1D were significant or very strong in European country groupings (WHO-Europe in Table 2 and in EEA in Table 3 respectively), but very weak in Asian country grouping (WHO-SEA in Table 1 and ACD in Table 3 respectively). This may be attributable to high genetic predispositions [13-15, 46] in Europe, but low genetic predisposition in Asia [46, 53].

Although we found that the correlations of both $I_{bs}$ and life expectancy to T1D prevalence rate existed globally and in different country groupings categorized with a variety of criteria, there are several limitations, including the intrinsic limitations (conceptualized as ecological fallacy) to this study.

Firstly, the data analysed were calculated for per capita in each country, so we could only demonstrate the relationships between T1D prevalence rate and $I_{bs}$ and life expectancy at country/population level, which does not necessarily correspond to the same relationships holding true at the individual level. We also need to point out that it would be difficult to test the relationships at the individual level due to very rare T1D occurrence rate.

Secondly, the slow changes in the genetic code of the human populations may not fully explain the increasing T1D prevalence. Nongenetic (environmental) factors partially determine whether, and how risk-associated genotypes may lead to overt T1D disease. Unfortunately, our $I_{bs}$ does not indicate if fitness change at population level is due to evolution of individuals or change/s. It may also be that altered lower natural immunity to infections following decades of using antibiotics may influence increased rates of autoimmune diseases including T1D.

Thirdly, the data compiled and/or collected by the major international agencies (IDF, WHO, FAO and the World Bank) are fairly crude, and may contain some random errors.

Finally, current evidence of the increasing frequency of many heritable genetic disorders, including T1D does not appear to be available. To the best of our knowledge, the T1D prevalence rate for all age groups at country level published by IDF may be the only version to single out T1D prevalence worldwide after consulting the major diabetes research or data collecting institutions. This may be because clinically, adult T1D is difficult to discriminate from certain forms of type 2 diabetes and from Latent Autoimmune Diabetes in Adults (LADA) [53]. Therefore, we don't know how much this set of data was confounded by other forms of diabetes.

The current prevailing paradigm on the increasing prevalence of T1D is that environmental pressures are now able to trigger genotypes [8, 9, 12-17]. Currently, medical gene intervention in modern medicine at this stage cannot remove T1D genes, and eugenics (improvement in the genetic stock) can offer no direction due to ethics issue. Therefore, study of T1D epidemiology based on prevalence/incidence T1D data of all age groups has become imperative as it may offer optimal solution to address or at least slow down T1D genetic load increases in different populations.

## Conclusions

Our study suggested that reduced natural selection ($I_{bs}$) may be the major contributor to the increasing prevalence of T1D worldwide with special regard to European countries. It seems that T1D epidemiology study based on all age groups may be the practical solution to identify the causes of increasing T1D prevalence and to address, or at least slow down, T1D genetic load increases in different populations as modern medicine cannot operate effectively at the gene level yet.

**Acknowledgements**

Financial support to MH was provided by the Wood Jones Bequest to the University of Adelaide.

The authors express appreciation to Dr Arthur Saniotis for his editorial assistance.

**Funding Statement**

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

**Competing Interests Statement**

None of the authors had a personal or financial conflict of interests.

**Contributorship Statement**

MH conceived the idea for this study. WPY extracted the data and WPY will take responsibility of data integrity. MH and WPY interpreted the data. WPY wrote the manuscript. WPY and MH edited and approved the manuscript for submission to the journal.

**Data Sharing Statement**

The corresponding authors could supply the country-specific data used in this research upon request.

## References


1. Noble, J.A. and H.A. Erlich, *Genetics of type 1 diabetes.* Cold Spring Harbor perspectives in medicine, 2012. **2**(1): p. 1-15.
2. Steck, A.K. and M.J. Rewers, *Genetics of type 1 diabetes.* Clinical Chemistry, 2011. **57**(2): p. 176-85.
3. Frese, T. and H. Sandholzer, *The Epidemiology of Type 1 Diabetes Mellitus*, in *Type 1 Diabetes*, A.P. Escher and A. Li, Editors. 2013, InTech: Online. p. 01-22.
4. Maahs, D.M., et al., *Epidemiology of Type 1 Diabetes.* Endocrinology and Metabolism Clinics of North America, 2010. **39**(3): p. 481-97.
5. International Diabetes Federation, *Key findings 2014.* 2015.
6. Melmed, S., et al., *Williams textbook of endocrinology*. 12th ed. Textbook of endocrinology. 2011, Philadelphia: Elsevier/Saunders.
7. Libby, P., et al., *Report of the National Heart, Lung, and Blood Institute- National Institute of Diabetes and Digestive and Kidney Diseases Working Group on Cardiovascular Complications of Type 1 Diabetes Mellitus.* Circulation, 2005. **111**(25): p. 3489-93.
8. Borchers, A.T., R. Uibo, and M.E. Gershwin, *The geoepidemiology of type 1 diabetes.* Autoimmunity Reviews, 2010. **9**(5): p. A355-A365.
9. Gale, E.A.M., *The rise of childhood type 1 diabetes in the 20th century.* Diabetes, 2002. **51**(12): p. 3353-61.
10. International Diabetes Federation, *IDF Diabetes Atlas, 3th edn*. 2006, Brussels, Belgium: International Diabetes Federation.
11. Patterson, C.C., et al., *Is childhood- onset type I diabetes a wealth- related disease? An ecological analysis of European incidence rates.* Diabetologia, 2001. **44 Suppl 3**: p. B9-B16.
12. Eisenbarth, G.S., R.C. Nayak, and S.L. Rabinowe, *Type I diabetes as a chronic autoimmune disease.* Journal of Diabetic Complications, 1988. **2**(2): p. 54-58.
13. Patterson, C., et al., *Diabetes in the young – a global view and worldwide estimates of numbers of children with type 1 diabetes.* Diabetes Research and Clinical Practice, 2013. **103**(2): p. 161-75.
14. Kaprio, J., et al., *Concordance for Type 1 ( insulin- dependent) and Type 2 ( non-insulin- dependent) diabetes mellitus in a population- based cohort of twins in Finland.* Clinical and Experimental Diabetes and Metabolism, 1992. **35**(11): p. 1060-67.

15. Kyvik, K.O., A. Green, and H. Beck-Nielsen, *Concordance rates of insulin dependent diabetes mellitus: a population based study of young danish twins.* BMJ, 1995. **311**(7010): p. 913-17.

16. Gale, E.A.M., *Declassifying diabetes.* Diabetologia, 2006(49(9)): p. 1989-95.

17. Wilkin, T.J., *The accelerator hypothesis: weight gain as the missing link between Type I and Type II diabetes.* Clinical and Experimental Diabetes and Metabolism, 2001. **44**(7): p. 914-22.

18. Byars, S.G., et al., *Colloquium papers: Natural selection in a contemporary human population.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107 Suppl 1**: p. 1787-92.

19. Henneberg, M. and J. Piontek, *Biological state index of human groups.* Przeglad Anthropologiczny, 1975. **XLI**: p. 191-201.

20. Henneberg, M., *Reproductive possibilities and estimations of the biological dynamics of earlier human populations.* Journal of Human Evolution, 1976. **5**: p. 41-8.

21. Henneberg, M., *Notes on the reproduction possibilities of human prehistorical populations.* Przeglad Anthropologiczny, 1975. **41**: p. 75-89.

22. Henneberg, M., *Quantitative evaluation of actual intensity of natural selection through differential fertility in human populations,.* American Journal of Phys Anthropology, 1985. **66**: p. 181.

23. Stephan, C.N. and M. Henneberg, *Medicine may be reducing the human capacity to survive.* Medical Hypotheses, 2001. **57**(5): p. 633-37.

24. International Diabetes Federation, *IDF Diabetes Atlas, 1st Edn*. 2000, Brussels, Belgium: International Diabetes Federation.

25. The United Nations. *World Fertility Data 2008*. 2012 29.07.2015]; Available from: http://www.un.org.

26. WHO, *World Health Statistics 2012*. Life tables for WHO Member States. 2012, Geneva: World Health Organization.

27. WHO. *Life expectancy at birth (years)*. 2013 29.07.2015]; Available from: http://apps.who.int.

28. Basu, S., et al., *Nutritional determinants of worldwide diabetes: an econometric study of food markets and diabetes prevalence in 173 countries.* Public health nutrition, 2013. **16**(1): p. 1-8.

29. Lamb, M., et al., *Sugar intake is associated with progression from islet autoimmunity to type 1 diabetes: the Diabetes Autoimmunity Study in the Young.* Clinical and Experimental Diabetes and Metabolism, 2015. **58**(9): p. 2027-2034.

30. WHO. *WHO regional offices*. 11.26.2015]; Available from: http://www.who.int.

31. The United Nations Educational Scientific and Cultural Organization. *UNESCO Regions-Latin America and the Caribbean*. 2014; Available from: http://www.unesco.org.

32. The World Bank. *Arab World | Data*. 2015; Available from: http://data.worldbank.org.

33. The OECD. *List of OECD Member countries*. 2015; Available from: http://www.oecd.org.

34. The European Free Trade Association. *What is the European Economic Area?* ; Available from: http://www.efta.int/eea.

35. Asia Cooperation Dialogue. *Member Countries*. Available from: http://www.acddialogue.com.

36. Asia-Pacific Economic Cooperation. *Member Economies-Asia-Pacific Economic Cooperation*. Available from: http://www.apec.org.

37. United Nations, Department of Economic and Social Affairs, and Population Division, *World Population Prospects: The 2012 Revision*. 2013: DVD Edition.

38. Livingstone, F.B. and J.N. Sphuler, *Cultural determinants of natural selection*, in *United Nations Educational, Scientific and Cultural Organization: Expert Meeting on the Biological Aspect of Race*. 1964: Mosco. p. 1-3.

39. Post, R.H., *Deformed nasal septa and relaxed selection.* Eugenics Quarterly, 1966. **13**: p. 101-12.

40. Tuomilehto, J., *The Emerging Global Epidemic of Type 1 Diabetes.* Curr Diab Rep, 2013. **13**(6): p. 795-804.

41. Saniotis, A. and M. Henneberg, *Medicine could be constructing human bodies in the future.* Medical Hypotheses, 2011. **77**(4): p. 560-64.

42. Moroni, L., I. Bianchi, and A. Lleo, *Geoepidemiology, gender and autoimmune disease.* Autoimmunity Reviews, 2012. **11**(6-7): p. A386-A392.

43. The Writing Group for the SEARCH for Diabetes in Youth Study Group, *Incidence of diabetes in youth in the United States.* JAMA, The Journal of the American Medical Association, 2007. **297**(24): p. 2716-25.

44. King, K.M., *A history of insulin: from discovery to modern alternatives.* Diabetes, 2003. **Prevalence of Type 1 diabetes in Australian children12**(19): p. 1137-41.

45. Medawar, P.B., *Do advances in medicine lead to genetic deterioration?*, in *Natural Selection in Human Populations.*, Bajema C. J., Editor. 1971, Robert E. Krieger Publishing Co.: New York. p. 300-08.

46. Onkamo, P., et al., *Worldwide increase in incidence of Type I diabetes – the analysis of the data on published incidence trends.* Clinical and Experimental Diabetes and Metabolism, 1999. **42**(12): p. 1395-403.

47. Kawasaki, E. and K. Eguchi, *Is Type 1 Diabetes in the Japanese Population the Same as among Caucasians?* Annals of the New York Academy of Sciences, 2004. **1037**(1): p. 96-103.

48. Suleyman, F., *Landmarks in diabetes care: a historical perspective.* Community Nurse, 1998. **4(6)**: p. 13-6.

49. Ogle, G., et al. *Children and diabetes: success and challenge in the developing world*. 2013; Available from: http://www.idf.org.

50. Australian Institute of Health and Welfare, *Prevalence of Type 1 diabetes in Australian children, 2008*. Diabetes series Number15. Cat. no. CVD 54. 2011, Canberra: AIHW.

51. Neville, S.E., et al., *Diabetes in Japan: a review of disease burden and approaches to treatment.* Diabetes Metab Res Rev., 2009. **25(8)**: p. 705-16.

52. Miller, R.G., et al., *Improvements in the life expectancy of type 1 diabetes: the Pittsburgh epidemiology of diabetes complications study cohort.* Diabetes, 2012. **61**(11): p. 2987-992.

53. Chiang, J.L., et al., *Type 1 diabetes through the life span: a position statement of the American Diabetes Association.* Diabetes care, 2014. **37**(7): p. 2034-54.

## Supporting document

Table S1. $I_{bs}$ values of 118 countries

| Country | Ibs | Country | Ibs | Country | Ibs |
|---|---|---|---|---|---|
| Iceland | 0.994 | Turkey | 0.975 | Nigeria | 0.782 |
| Cyprus | 0.994 | Bahamas | 0.974 | Mali | 0.774 |
| Singapore | 0.994 | Venezuela | 0.974 | Zambia | 0.761 |
| Japan | 0.993 | Colombia | 0.974 | Cameroon | 0.760 |
| Switzerland | 0.993 | Tunisia | 0.972 | Mozambique | 0.752 |
| Sweden | 0.992 | Thailand | 0.971 | Congo, DR | 0.729 |
| Luxembourg | 0.992 | Egypt | 0.971 | New Zealand | 0.988 |
| Germany | 0.992 | Belize | 0.969 | TFYR Macedonia | 0.986 |
| Italy | 0.992 | El Salvador | 0.969 | Chile | 0.986 |
| Czech Republic | 0.992 | China | 0.969 | USA | 0.985 |
| Spain | 0.992 | Sri Lanka | 0.969 | Qatar | 0.985 |
| France | 0.992 | Peru | 0.967 | Malaysia | 0.985 |
| Denmark | 0.991 | Fiji | 0.967 | Costa Rica | 0.984 |
| Norway | 0.991 | Panama | 0.967 | Bulgaria | 0.984 |
| Netherlands | 0.991 | Paraguay | 0.967 | Romania | 0.983 |
| Israel | 0.991 | Libya | 0.966 | Barbados | 0.981 |
| Greece | 0.991 | Saudi Arabia | 0.966 | Bahrain | 0.981 |
| Austria | 0.991 | Iran | 0.963 | Kuwait | 0.980 |
| Finland | 0.991 | Georgia | 0.963 | Uruguay | 0.980 |
| Portugal | 0.991 | Ecuador | 0.962 | Lebanon | 0.980 |
| Belgium | 0.990 | Jordan | 0.961 | Philippines | 0.953 |
| Malta | 0.990 | Suriname | 0.956 | Tonga | 0.953 |
| Ireland | 0.990 | Dominican Republic | 0.956 | Kyrgyzstan | 0.951 |
| Australia | 0.990 | Jamaica | 0.956 | Guyana | 0.949 |
| United Kingdom | 0.990 | Trinidad and Tobago | 0.956 | Indonesia | 0.945 |
| Slovenia | 0.990 | Honduras | 0.955 | Guatemala | 0.939 |
| Estonia | 0.989 | Morocco | 0.955 | Iraq | 0.936 |
| Canada | 0.989 | Kazakhstan | 0.955 | Bolivia | 0.931 |
| Hungary | 0.989 | Pakistan | 0.877 | Bangladesh | 0.921 |
| Slovakia | 0.989 | Kenya | 0.872 | Madagascar | 0.912 |
| Croatia | 0.989 | Haiti | 0.871 | India | 0.898 |
| Cuba | 0.989 | Senegal | 0.870 | Papua New Guinea | 0.892 |
| Poland | 0.989 | Togo | 0.854 | Gabon | 0.890 |
| Republic of Korea | 0.988 | Gambia | 0.849 | | |
| Lithuania | 0.988 | Tanzania | 0.830 | | |
| Ukraine | 0.977 | Sudan | 0.821 | | |
| Mauritius | 0.977 | Uganda | 0.817 | | |
| Argentina | 0.976 | Ethiopia | 0.815 | | |
| Mexico | 0.976 | South Africa | 0.811 | | |
| Syrian Arab Republic | 0.976 | Congo | 0.810 | | |
| Albania | 0.975 | Zimbabwe | 0.809 | | |
| Brazil | 0.975 | Côte d'Ivoire | 0.783 | | |

Table S2: Descriptive Statistics

| | N | Range | Minimum | Maximum | Mean | Std. Deviation | Skewness |
|---|---|---|---|---|---|---|---|
| | Statistic | Statistic | Statistic | Statistic | Statistic | Statistic | Statistic |
| T1DM (Total) Prevalence (‰), IDF, 2000 | 118 | 8.58 | 0.03 | 8.60 | 1.17 | 1.20 | 2.77 |
| $I_{bs}$ self-calculated | 118 | 0.27 | 0.73 | 0.99 | 0.94 | 0.07 | -1.70 |
| Life expectancy increase (e50, 1950-2010) | 116 | 12.14 | -0.57 | 11.57 | 5.87 | 2.46 | -0.15 |
| Life expectancy (e50, 2005-10), UN | 118 | 14.32 | 20.04 | 34.36 | 27.81 | 3.45 | -0.19 |
| Life Expectany at birth (years), WHO, 2010 | 118 | 32.00 | 52.00 | 84.00 | 73.23 | 7.96 | -0.89 |
| BMI≥ 30 prevalence 18+ (%), WHO, 2010 | 118 | 38.70 | 2.90 | 41.60 | 18.34 | 8.80 | -0.11 |
| GDP PPP, the World Bank 2010 | 116 | 124775.41 | 619.47 | 125394.88 | 19773.82 | 19555.61 | 2.15 |
| Sugar availability (g/capita/day), FAO, 2010 | 118 | 167.62 | 0.00 | 167.62 | 86.99 | 46.40 | -0.23 |
| Urban population (% of total) the World Bank, 2010 | 118 | 90.91 | 9.09 | 100.00 | 62.19 | 21.93 | -0.37 |
| Valid N (listwise) | 114 | | | | | | |

Article 3/10: Reduced natural selection contributes to global obesity increase more in males than in females due to more environmental modifications in female body mass

Wenpeng You[1], Maciej Henneberg[1,2]

[1] Adelaide Medical School, the University of Adelaide, Adelaide, South Australia, Australia 5005

[2] Institute of Evolutionary Medicine, University of Zürich, Zürich, Switzerland

**Under review:** Under review by PLOS ONE (after two revisions)

✉ **Correspondence:** Wenpeng You Wenpeng.you@adelaide.edu.au

## Contextual Statement

Natural selection is the key mechanism of evolution, which changes heritable traits characteristic of a population over generations.

In the past 100-150 years, the advanced medical care has reduced natural selection more than previously thought. It has been like a double-edged sword. It saves the lives of those people with the non-communicable diseases which are featured with strong heredity. Meanwhile, it offers the opportunities for those people to pass on their deleterious genes/mutations to their next generation.

Males and females may have inherited obesity related genes/mutations equally over the past 100-150 years. In general, female obesity is more prevalent than male obesity prevalence. However, this ratio of male to female obesity prevalence varies in different regions and different cultures.

We hypothesized and tested that reduced natural selection (indexed by opportunity for reproduction at population level) offers the equal opportunities for males and females to accumulate the obesity related genes/mutations, but it may have more effects on males than on females due to less environmental modification.

**Statement of Authorship**

# Statement of Authorship

| | |
|---|---|
| Title of Paper | Reduced natural selection may contribute to global obesity increase more in males than in females |
| Publication Status | ☐ Published ☐ Accepted for Publication<br>☑ Submitted for Publication ☐ Publication Style |
| Publication Details | PLOSONE PONE-D-17-11108R1; We had completed two revisions as required by the editor office. |

## Principal Author

| | | | |
|---|---|---|---|
| Name of Principal Author (Candidate) | Wenpeng You | | |
| Contribution to the Paper | Conceived the hypothesis, collected data, performed analysis on all samples, interpreted data, wrote manuscript and acted as corresponding author. | | |
| Overall percentage (%) | 60 | | |
| Signature | | Date | 22/12/2017 |

## Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

| | | | |
|---|---|---|---|
| Name of Co-Author | Maciej Henneberg | | |
| Contribution to the Paper | Supervised development of work, formalized and interpreted data analysis, and evaluated and edited manuscript | | |
| Signature | | Date | 22/12/2017 |

| | | | |
|---|---|---|---|
| Name of Co-Author | NA | | |
| Contribution to the Paper | | | |
| Signature | | Date | |

Please cut and paste additional co-author panels here as required.

## Abstract

**Objective:** Relaxed natural selection, measured by Biological State Index (Ibs), results in unfavourable genes/mutations accumulation in population. We aim to examine and compare the effects of reduced natural selection on male and female obesity prevalence.

**Methods:** Country specific data were captured for ecological study. Curvilinear regressions, bivariate and partial correlations, linear mixed models and multivariate linear regression analyses were used to examine the relationship between $I_{bs}$ and sex-specific obesity prevalence. Per capita GDP, urbanization and caloric intake were controlled for as the confounding factors. Fisher r-to-z transformation, $R^2$ increment in multivariate regression and F-test were used to compare the correlations.

**Results:** Curvilinear regressions, bivariate and partial correlations (controlled for GDP, urbanization and calories) revealed that $I_{bs}$ was significantly correlated to obesity prevalence of both sexes, but significantly stronger to male than to female obesity prevalence. Curvilinear regression models also showed strong correlations, but not significantly different between two sexes. Mixed linear models, with effects of GDP, urbanisation and caloric intake controlled for, showed that male and female average obesity prevalences were significantly higher in countries with greater $I_{bs}$ value than their equivalents in countries with lower $I_{bs}$. Between higher and lower $I_{bs}$ countries, the gap of male obesity prevalence is 60% greater than the gap of female obesity prevalence. Stepwise multiple regression identified that $I_{bs}$ was a significant predictor of obesity prevalence of both sexes. Multivariate regression showed that, adding $I_{bs}$ as an obesity predictor, $R^2$ increment in male model was significantly greater than in female model.

**Conclusions:** Reduced natural selection may drive males and females to accumulate metabolic faulty genes equally. Probably due to greater environmental intervention in regulating female body mass, reduced natural selection may show more contributing effects to male obesity prevalence. Gene therapy may be the optimal solution to address the obesity pandemic.

## Introduction

Being overweight was once considered a problem only of high-income countries, but now obesity prevalence is rising worldwide and affects both the developed and developing countries [1]. Indeed, obesity and its sequelae are now so common that they are replacing traditional problems such as undernutrition and infectious diseases as the most significant causes of ill-health [2]. Moreover, people considered overweight or obese have been subject to discrimination and prejudice [2, 3].

The Body Mass Index (BMI) is a common tool to determine body weight status. In WHO statistics [4-6], there are four body weight status definitions regarding individual adult's BMI, i.e. obesity (BMI ≥30 kg/m$^2$), pre-obesity (BMI > 25 kg/m$^2$, but < 30 kg/m$^2$), overweight (BMI ≥ 25 kg/m$^2$) and health (BMI > 18 kg/m$^2$, but < 25 kg/m$^2$).

BMI is a highly heritable human trait [7]. Despite this, in the past three decades, extensive studies explored how non-genetic factors, such as excessive intake of energy, changes of food components, sedentary lifestyle and gut flora imbalance, contributed to body weight increase [8-19]. Traditionally, obesity has been attributable to overeating, but how exactly obese people gain energy surplus is still vigorously debated [11-13, 17]. Recently, a couple of novel studies based on experiments reported that micro flora balance in human guts plays an important role in human energy metabolism [20-23]. Although conclusions of these studies are controversial and/or circumstantial, some researchers questioned the importance of genetics in the aetiology of obesity [7].

Natural selection is a key mechanism of evolution. It has relaxed its effects in shaping humans as a species because modern living conditions, public health and medicine made it reduced sharply [24, 25]. Natural selection, together with mutations controls the frequency of genes, which determine human heritable traits. Population escaping from natural selection over successive generations may make the prevalence of their heritable traits subject to change due to changes in mutation/selection balance [7, 26]. A direct consequence of this process is that *de novo* mutations, including those affecting energy balance and metabolism, recently have accumulated at an unexpectedly significant pace [27-30]. Multiple mutations may be accumulated in genomes quickly, which influences the phenotype [31-33] after only a few generations.

The Biological State Index ($I_{bs}$) measures the populational reproductive success [33-36]. Therefore, it can be used to measure the magnitude of reduced natural selection at population level. The $I_{bs}$ calculation formula [34, 35] is:

$$I_{bs} = 1 - \sum_{x=0}^{x=\omega} \mathrm{dxsx}$$

Where
$d_x$ = the frequency of deaths at age x
$s_x$ =the probability of not having completed fertility at age x
$\omega$: the age at death of the oldest member of the group

The $I_{bs}$ expresses an opportunity for an average individual born into a population to pass on genes to the next generation. The greater $I_{bs}$ value is, the less opportunity for natural selection to act on the population through mortality because all individuals in that

population survive their reproductive period (15-50 years old). Further explanation and calculations of the $I_{bs}$ are described in the Additional File 1 (Additional File: Text -AF 1) and for the $I_{bs}$ value of each country see Additional File 2 (Additional File: Table AF 1).

It was postulated that unfavourable genes may have been accumulating in human populations due to greatly reduced natural selection in the past 100-150 years [33, 36-39]. This hypothesis has been tested in several studies [33, 36, 37, 40] and a very recent study argued that relaxation of natural selection may have been contributing to worldwide obesity prevalence due to accumulation of genes affecting metabolism in human populations [41]. The rationale of the study into the relationship between reduced natural selection and obesity prevalence increase is described as follows:

The probable effect of *de* novo mutations is detrimental. Each population has a segment who carry metabolism and energy balance fault genes. When members of this segment of population participate in the reproduction, they may pass their metabolic fault genes into the next generation [30]. The frequency of metabolic fault genes will increase when a larger fraction of total population have opportunity to participate in reproduction under a given set of mortality conditions [34, 35]. However, only the contribution of relaxed natural selection to obesity prevalence in total population (both sexes) has been studied. No effects of reduced natural selection on obesity prevalence separately in males and females were considered.

The topic of sex disparities in obesity remains largely underresearched, let alone addressed. From the perspective of total population at the country level, males and females in the next generation may share equal opportunities to inherit metabolic fault genes. However, worldwide, obesity is more prevalent in females (23.28%) than in males (15.89%) [42]. Studies of sex disparity in obesity considered differences in fat distribution [43, 44], body fat storage level [45-47], the role of parental investment [48] and the role of estrogen effect on obesity [49]. The interaction between genetic factors and sex in identical twins' BMI has been reported [50, 51]. However, the effects of relaxed natural selection on obesity in different sexes at the population level have not been explored [52]. Due to obvious differences in body composition, fat distribution and hormonal regulation of metabolism, especially during pregnancy, lactation and post-partum periods, expression of different genes in males and females may be influencing energy balance of individuals.

Therefore, the objective of the present study was to evaluate and compare the role of the $I_{bs}$ contribution to male and female obesity prevalence from a global perspective using country sex-specific obesity prevalence data.

## Materials and Methods

### Data Collection and Selection

The WHO Global Health Observatory (GHO) data (2014) on estimated sex-specific obesity prevalence rates by country were obtained and used as the dependent variables [42]. The estimates of sex-specific prevalence rates of obesity are expressed as the percentage of population aged 18+ with BMI equal to or over 30 kg/m$^2$.

We also extracted data on $I_{bs}$ and on obesity prevalence rates of Australian females and males for the years 1976, 1981, 1986, 1991, 1996 and 2009 [53].

Country specific $I_{bs}$ values were used as the independent variable. The $I_{bs}$ calculation [34, 35] was based on the fertility data of each country published by United Nations in 2008 [54] and the mortality data of life tables (2009) published by World Health Organization (WHO) in 2012 [55]. These calculations were the same as in the previous study published by Budnik and Henneberg [30]. Calculations and interpretations of $I_{bs}$ are further described in the Supporting Information (Additional File: Text AF 1). Australian longitudinal $I_{bs}$ was calculated using data published by the Commonwealth Bureau of Census and Statistics. In terms of data availability and quality, for Australia we were only able to calculate the $I_{bs}$ for the years of 1976, 1981, 1986, 1991, 1996 and 2009.

Urbanization (expressed as a percentage of the population living in urban areas in 2010) [56], mean caloric intake in 2011-2013 (expressed in grand total calories per capita per day) [57] and gross domestic product per capita (GDP, expressed in purchasing power parity in 2010 US dollars) [58] were considered and controlled for as the confounding factors. The selection criteria for potential confounding factors include: 1) Due to more affordability of the increases in caloric intake [59], obesity has traditionally been considered as an affluence-related medical condition [60]. 2) Living in urban setting leads to sedentary lifestyle (less physical activity) and poorer diets (more animal products and sugar), which have been considered an important factor to increase the risk of obesity [1, 7, 61-63]. Urban living setting also mirrors the Western lifestyle.

We aligned the $I_{bs}$ with prevalence rates of obesity in females and males and then matched them with GDP, caloric intake and urbanization. Country specific data for 191 countries were put in a uniform format. Each country was treated as an individual subject and all of their available information was analysed. For some countries an estimate of one or the other variable was missing, thus specific analyses have sample sizes varying from 168 to 191.

We also aligned Australian $I_{bs}$ with obesity prevalence of Australian females and males for those years in which we were able to use the data for $I_{bs}$ calculation in order to explore longitudinal trend.

Although the WHO Global Health Observatory (GHO) data repository (2014) [42] defined four levels of BMIs for males and females (obesity, overweight, normal and underweight), we only chose obesity prevalence rates in females and males for modelling, analysing and reporting the correlation and regression results because the results for obesity can be compared with the findings of the previous study conducted by Budnik and Henneberg [30].

**Data Robusticity Check**

The diagnostic test was run to check if there was multicollinearity problem between the data we collected. All the tolerances were less than 0.20 and all the Variance Inflation Factors (VIF) were above 5, which indicates there was not multicollinearity issue [64] (Additional File: Table AF 2).

The Kolmogorof-Smirnov and Shapiro-Wilk tests were performed with SPSS to test the normality of distributions of variables used (Details see Additional File: Table AF 4). All variables analysed here were not normally distributed, thus various data transformations as described below were performed for each method applied.

**Scatter plots**

Worldwide, the relationships between the $I_{bs}$ and each of the male and female obesity prevalence rates were explored and visualized in Microsoft Excel® producing scatter plots. Scatter plots were also used to explore the longitudinal correlation between the Australia-specific $I_{bs}$ and Australian sex-specific obesity prevalence rates. The best fit trendlines were reported respectively.

**Curvilinear Correlation Analysis**

Due to abnormal data distribution detected in the Kolmogorof-Smirnov and Shapiro-Wilk tests, partial correlation analysis was conducted using correlations of residuals, not the standard SPSS procedure. Logarithmic, exponential, power and polynomial regression models were fitted to the data and for each specific regression analysis, the model producing the greatest fit by the least squares criterion (greatest coefficient of determination - $R^2$) was applied. First, best curvilinear regression between GDP and sex-specific obesity prevalence has been obtained, then residuals of individual country points around that line were regressed on urbanisation. Residuals around the best

regression of GDP-residuals on urbanisation were calculated (second-order residuals). These second-order residuals were regressed on the caloric intake and then residuals around this regression line calculated (third-order residuals). First order residuals (sex-specific obesity prevalence standardised on GDP), second order (sex-specific obesity prevalence standardised on GDP and urbanization) residuals and third order residuals (sex-specific obesity prevalence standardised on GDP, urbanization and caloric intake) were regressed on $I_{bs}$ thus obtaining correlations of $I_{bs}$ to sex-specific obesity prevalence corrected for effects of GDP only, GDP and urbanisation, and GDP, urbanisation and caloric intake respectively.

**Data Analysis Based on Linear correlation models**

When data were logarithmed, similar levels of Pearson r correlation and Spearman rho between all variables were obtained. This allows us to consider that the logged data distributions, though not normal, provide homoscedastic distributions as required for linear correlations. Therefore, the data analysis was performed in four steps:

1) Pearson and non-parametric correlation analysis were conducted to examine the strength and direction of the correlations between all variables.

2) Partial correlation analysis was performed to explore the independent linear correlations of $I_{bs}$ to male and female obesity prevalence rates respectively while we controlled for GDP, urbanization and caloric intake.

Fisher's r-to-z transformation was conducted to assess significance level of differences between the Pearson's r and partial correlation coefficientsefficient of $I_{bs}$ to male and female obesity prevalence rates.

Cohen's $f^2$ was used to calculate and report the "effect size" in the partial correlation analysis.

3) Standard multivariate linear regression (Enter) was conducted on log-transformed data to obtain and compare the Beta coefficients between sex-specific obesity prevalence and all independent variables, which included $I_{bs}$, calories, GDP and urbanization.

Standard multivariate linear regression (Stepwise) was performed to assess which non-$I_{bs}$ predictor(s) made substantial contributions to variation in obesity, and then $I_{bs}$ was added to the list of predictors to show improvement in model fits for males and females. The magnitudes of improvements in the two model fits were firstly compared with the absolute improvement values obtained from "the $R^2$ improvement in male prevalence

due to adding $I_{bs}$" and "the $R^2$ improvement in female prevalence due to adding $I_{bs}$" respectively. F-test was used to compare and determine if there is significant difference between the magnitudes of the two improvements. We calculated the ratio (F value) of "the $R^2$ improvement in male prevalence due to adding $I_{bs}$" to "the $R^2$ improvement in female prevalence due to adding $I_{bs}$". The calculated F value was compared with the value of p=0.05 and p=0.01 at degrees of freedom used in regression analyses.

4) The linear Mixed Model Analysis was conducted to summarise the results allowing us to intercept change at the country and regional levels after the data were nested within the WHO regions.

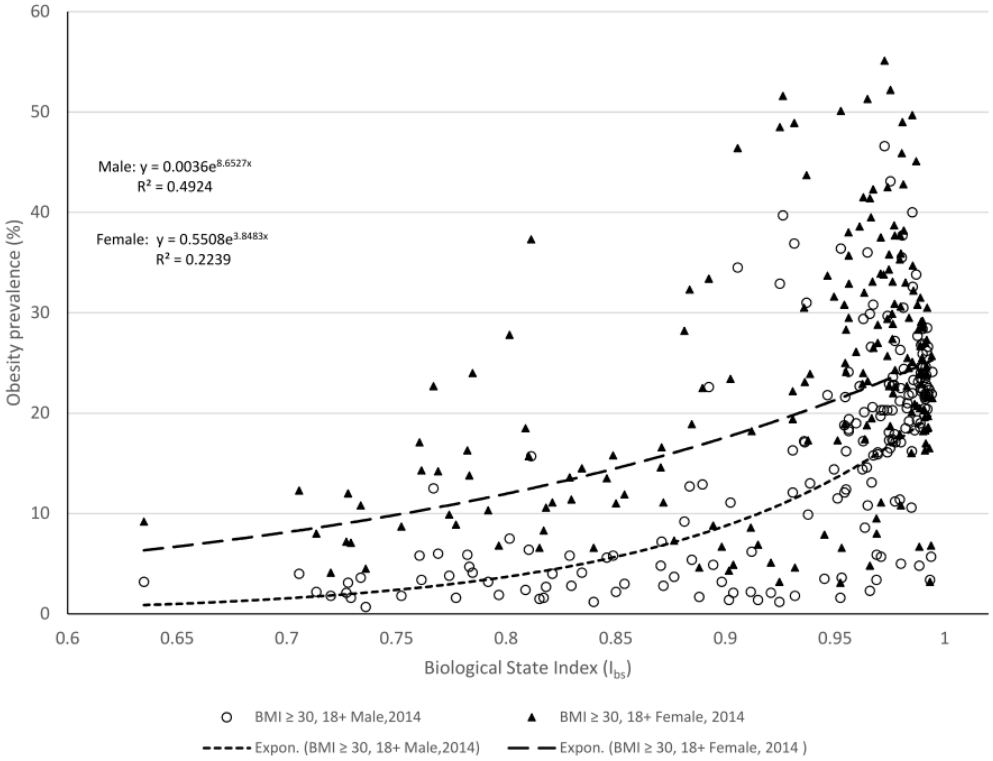
For the application of mixed-effects models that were based on linear relations between variables, scales of GDP, urbanisation and caloric intake were transformed from interval to ordinal. Values of each variable were ordered from the smallest to the largest, ranked and ranks standardised on numbers of observations because numbers of countries for which values of GDP, Urbanisation and Caloric intake were available differed somewhat (from 168 to 191). This way the rank of the country with the maximum value became 100 while the rank of the country with minimum value was 100*1/N that is a fractional number. This procedure produced rectangular distributions of all variables, thus these distributions became homoscedastic and as such acceptable for linear analyses. Averages of ordinally measured variables in the entire sample are 50.0 and thus their averages in variously grouped subsamples are easily interpretable. The mixed model with nested terms fixed and random effects using the Restricted Maximum Likelihood method of estimation was run.

Pearson's r, Spearman's rho coefficient, partial correlation, the linear Mixed Model Analysis and multiple-linear regression analyses were conducted using SPSS v. 24. The statistical significance was set at the 0.05 level, but the significance levels at 0.01 and 0.001 were also reported.

## Results

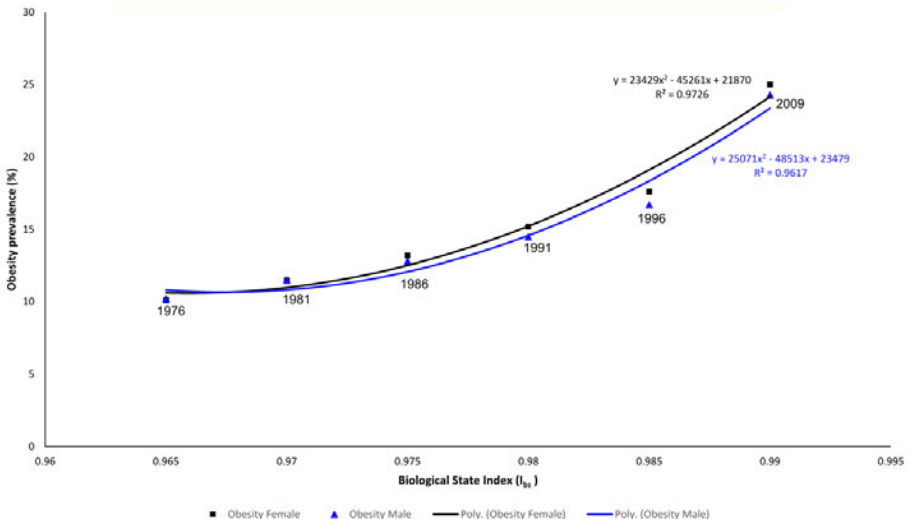
$I_{bs}$ was in strong and significant correlation (exponential) to both male obesity ($r=0.70$, $p<0.001$) and female obesity ($r=0.47$, $p<0.001$). Fisher r-to-z revealed that $I_{bs}$ was in significantly stronger correlation to male obesity than to female obesity ($z=3.46$, $p<0.001$) (Figure 1).

Figure 1 Relationships between $I_{bs}$ and obesity prevalence estimates in males and female




Similar longitudinal trends were revealed between Australia-specific $I_{bs}$ and Australian male and female obesity prevalence (Figure 2).

Figure 2 Longitudinal correlation between $I_{bs}$ and sex-specific obesity prevalence in Australia



Curvilinear residual regressions revealed that $I_{bs}$ was still significantly correlated to both male and female obesity prevalence when corrected for effects of GDP (r= 0.23, p<0.001 and r=0.23, p<0.001 respectively), GDP and urbanisation (r=0.52, p<0.001 and r=0.61, p<0.001 respectively), and GDP, urbanisation and caloric intake (r=0.23, p<0.01 and r=0.20, p<0.01 respectively). Fisher r-to-z transformation was conducted to compare the correlations between $I_{bs}$ and male and female obesity prevalence respectively at the first, second and third order residuals, but there was no significant difference detected. This may suggest that reduced natural selection contributes to male and female obesity equally regardless of the environmental intervention in regulating male and female body mass (Table 1).

Table 1: Curvilinear relationship between $I_{bs}$ and male and female prevalence standardized on individual major obesity contributors in different combinations

| Prevalence % | Regression equation | R | n | Fisher r-to-z |
|---|---|---|---|---|
| Prevalence (Actual) | | | | |
| Male | $y = 0.0036e^{8.6527x}$ | 0.70 | 191 | Z=3.46, p=0.0005 |
| Female | $y = 0.5508e^{3.8483x}$ | 0.47 | 191 | |
| Prevalence Standardized on GDP | | | | |
| Male | $y = 0.1819x^{0.4881}$ | 0.34 | 184 | Z=-1.26, p=0.2077 |
| Female | $y = -360.53x^2 + 639.48x - 278.31$ | 0.23 | 184 | |
| Prevalence Standardized on GDP and Urbanization | | | | |
| Male | $y = 694.64x^3 - 1375.5x^2 + 901.83x - 97.211$ | 0.52 | 184 | Z=-1.31, P=0.1902 |
| Female | $y = 8369x^3 - 19996x^2 + 15903x - 4192.6$ | 0.61 | 184 | |
| Prevalence Standardized on GDP, Urbanization and Calories | | | | |
| Male | $y = -120.34x^3 + 373.64x^2 - 339.93x + 93.492$ | 0.23 | 168 | Z=0.28, P=0.775 |
| Female | $y = 2107.3x^3 - 5332.6x^2 + 4524.5x - 1292.5$ | 0.20 | 168 | |

Sex specific obesity prevalence is the percentage of defined population segment with a body mass index (BMI) of no less than 30 $kg/m^2$.

Data sources: Total calories data from the FAO's FAOSTAT; BMI ≥30 data from the WHO Global Health Observatory; GDP data from the World Bank; Urbanization data from WHO. Biological State Index ($I_{bs}$) was self-calculated with country specific fertility data published by the United Nations and the mortality data published by World Health Organization (WHO).

In Pearson correlation analysis, worldwide, $I_{bs}$ is significantly correlated to both male ($r=0.692$, $p<0.001$) and female ($r=0.470$, $p<0.001$) obesity prevalence rates at levels similar to those for power regressions (Table 2). Similar values of correlation coefficients were observed in Spearman's rho analysis as well indicating that log-transformation is sufficient to avoid substantial deviations from linear regressions in moment-product correlations (Table 2).

Table 2: Pearson r correlation (above the diagonal) Spearman rho (below the diagonal) between all variables*

| | Obesity %, Male | Obesity %, Female | $I_{bs}$ | Calories | GDP | Urbanization |
|---|---|---|---|---|---|---|
| Obesity %, Male | 1 | 0.903** | 0.692** | 0.716** | 0.761** | 0.580** |
| Obesity %, Female | 0.845** | 1 | 0.470** | 0.493** | 0.517** | 0.399** |
| $I_{bs}$ | 0.667** | 0.371** | 1 | 0.639** | 0.710** | 0.513** |
| Calories | 0.742** | 0.451** | 0.765** | 1 | 0.759** | 0.602** |
| GDP | 0.758** | 0.504** | 0.866** | 0.756** | 1 | 0.672 |
| Urbanization | 0.583** | 0.372** | 0.666** | 0.660** | 0.736** | 1 |

Pearson r and Spearman rho is reported. Number of countries included in the analysis ranges from 172 to 191.

*All correlations are significant at the 0.001 level (two-tailed).

Obesity % is percentage of defined population segment with a body mass index (BMI) of no less than 30 $kg/m^2$.

Data sources: Total calories data from the FAO's FAOSTAT; BMI ≥30 data from the WHO Global Health Observatory; GDP data from the World Bank; Urbanization data from WHO. Biological State Index ($I_{bs}$) was self-calculated with country specific fertility data published by the United Nations and the mortality data published by World Health Organization (WHO).

Fisher's r-to-z transformation revealed that, in Pearson correlation analysis, $I_{bs}$ was correlated to male obesity prevalence significantly stronger than to female obesity prevalence ($z= 3.31$, $p<0.001$) (Table 2).

Partial correlation analysis showed that, worldwide, the $I_{bs}$ was still significantly correlated to the male and female obesity prevalence ($r=0.332$, $p<0.001$ and $r=0.147$, $p<0.05$ respectively) while we controlled for caloric intake, GDP and urbanization (Table 3). $I_{bs}$ was in significantly stronger correlation to male obesity prevalence than to female obesity prevalence ($z= 1.76$, $p<0.05$) (Table 2).

The effect size of $I_{bs}$ on male prevalence is 0.124, which is much greater than on female prevalence, 0.022 (Table 3).

Table 3 Correlation coefficients and Fisher's r-to-z transformations of Pearson r and partial correlations between $I_{bs}$ and male and female obesity prevalence

| | Pearson correlation $I_{bs}$ | | | | Partial Correlation $I_{bs}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | N | r | p | Fisher's r-to-z transformation | df | r | p | Effect Size | Fisher's r-to-z transformation |
| BMI 30, 18+M,2014 | 191 | 0.692 | 0.000 | z= 3.31 p=0.0005 | 163 | 0.332 | 0.000 | 0.124 | z=1.76 p=0.039 |
| BMI 30, 18+F,2014 | 191 | 0.470 | 0.000 | | 163 | 0.147 | 0.030 | 0.022 | |
| GDP 2010 USD | 184 | 0.710 | 0.000 | - | - | - | - | - | - |
| Calories, mean 2011-13 | 172 | 0.639 | 0.000 | - | - | - | - | - | - |
| Urbanization | 191 | 0.513 | 0.000 | - | - | - | - | - | - |

Partial correlation (two-tailed) is reported. -, Controlled variable or not relevant.

BMI≥30 and BMI ≥ 25 are percentages of defined population with a body mass index (BMI) of no less than 30 kg/m2 and 25 kg/m2 respectively. Population fall in pre-obesity is indicated as "BMI 25-30, M".

Data sources: Total calories data from the FAO's FAOSTAT. BMI (≥30, ≥25 and mean) data from the WHO Global Health Observatory, and self-calculated pre-obesity (BMI≥25 but <30) by subtracting BMI ≥ 30 from BMI ≥25. GDP data from the World Bank. Urbanization data from WHO.

Mixed linear models revealed that the male and female obesity prevalence rates were significantly different between WHO regions ($F=11.59$, $P<0.001$ & $F=12.18$, $P<0.001$ respectively) when GDP, urbanization and calories were controlled for. When the effects of GDP, urbanisation and calories were kept constant, mixed linear models revealed that male and female average obesity prevalences were significantly higher in countries with greater $I_{bs}$ than their equivalents in countries with lower $I_{bs}$, and that between higher and lower $I_{bs}$ countries, the gap of male obesity prevalence (20.48%-9.54%) is 60% greater than the gap of female obesity prevalence (25.68%-18.85%) (Table 4 and Further details see Additional File: Table AF 3).

Table 4 Results of Mixed Model Analysis with the country specific data nested within WHO regions.

**Males**

| | Countries with $I_{bs}$ ≥0.9658 | | | Countries with $I_{bs}$ <0.9658 | | |
|---|---|---|---|---|---|---|
| WHO Region | N | Mean | Std Deviation | N | Mean | Std Deviation |
| Africa | 1 | 11.20 | NA | 39 | 5.19 | 3.97 |
| Americas | 24 | 20.72 | 4.59 | 11 | 14.95 | 4.71 |
| Eastern Mediterranean | 7 | 27.14 | 5.90 | 10 | 13.23 | 9.48 |
| Europe | 42 | 21.87 | 2.97 | 8 | 15.59 | 4.04 |
| South-East Asia | 3 | 4.70 | 1.18 | 6 | 2.78 | 1.35 |
| West Pacific | 9 | 14.46 | 11.68 | 8 | 17.76 | 14.28 |
| Worldwide | 86 | 20.48 | 6.60 | 82 | 9.54 | 8.23 |

**Females**

| | Countries with $I_{bs}$ ≥0.9658 | | | Countries with $I_{bs}$ <0.9658 | | |
|---|---|---|---|---|---|---|
| WHO Region | N | Mean | Std Deviation | N | Mean | Std Deviation |
| Africa | 1 | 24.30 | NA | 39 | 15.33 | 7.55 |
| Americas | 24 | 31.21 | 5.07 | 11 | 27.33 | 6.62 |
| Eastern Mediterranean | 7 | 39.60 | 4.38 | 10 | 23.02 | 13.26 |
| Europe | 42 | 22.99 | 4.09 | 8 | 21.39 | 3.56 |
| South-East Asia | 3 | 10.47 | 0.85 | 6 | 6.05 | 2.12 |
| West Pacific | 9 | 17.86 | 13.69 | 8 | 26.24 | 19.92 |
| Worldwide | 86 | 25.68 | 8.77 | 82 | 18.85 | 11.11 |

1. Means of prevalence (%) of obesity (>30kg/$m^2$ ) for males and females in countries with $I_{bs}$ values above and below median are shown.
2. Dependent Variable: BMI≥30 prevalence rates in Males and Females in 2014.
3. The mixed model with nested terms fixed and random effects using the Restricted Maximum Likelihood method of estimation was run.
4. $I_{bs}$ Med: Cutoff ponit of 0.9658
5. NA: Not available

Multivariate regression model (Enter) revealed that $I_{bs}$ was a significant (Beta=0.287, p<0.001) predictor of male obesity prevalence when $I_{bs}$, calories, GDP and urbanization were entered as the predicting variables. In contrast, $I_{bs}$ was only a relatively weak and marginally significant (Beta=0.180, p=0.06) predictor of female obesity prevalence (Table 5).

Stepwise multivariate regression model results indicated that $I_{bs}$ was, after GDP, the second strongest and significant predictor of both male and female obesity prevalence. The absolute improvement of $R^2$ value due to adding $I_{bs}$ in male model fit was 0.038 (from 0.680 to 0.642), which was more than double the absolute improvement value 0.016 (from 0.284 to 0.268) due to adding $I_{bs}$ to female model fit (Table 5). This difference was significant (F value 2.375, p<0.01).

Table 5 Results of Enter and Stepwise linear multivariate regression analyses to identify significant predictors of obesity in females and males

**Enter**

| | Male obesity prevalence | | | | Female obesity prevalence | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $I_{bs}$ excluded | | $I_{bs}$ included | | | $I_{bs}$ excluded | | $I_{bs}$ included | |
| Variable | Beta | Sig. | Beta | Sig. | Variable | Beta | Sig. | Beta | Sig. |
| $I_{bs}$ | - | - | 0.287 | 0.000 | $I_{bs}$ | - | - | 0.180 | 0.060 |
| Calories | 0.233 | 0.002 | 0.175 | 0.014 | Calories | 0.131 | 0.209 | 0.095 | 0.366 |
| GDP | 0.515 | 0.000 | 0.360 | 0.000 | GDP | 0.345 | 0.002 | 0.247 | 0.040 |
| Urbanization | 0.135 | 0.035 | 0.126 | 0.037 | Urbanization | 0.117 | 0.194 | 0.112 | 0.212 |

**Stepwise**

| | Male obesity prevalence | | | | Female obesity prevalence | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $I_{bs}$ excluded | | $I_{bs}$ included | | | $I_{bs}$ excluded | | $I_{bs}$ included | |
| Model | Variable | Adjusted $R^2$ | Variable | Adjusted $R^2$ | Model | Variable | Adjusted $R^2$ | Variable | Adjusted $R^2$ |
| 1 | GDP | 0.606 | GDP | 0.606 | 1 | GDP | 0.268 | GDP | 0.268 |
| 2 | Calories | 0.635 | $I_{bs}$ | 0.657 | 2 | Calories | Removed | $I_{bs}$ | 0.284 |
| 3 | Urbanization | 0.642 | Calories | 0.673 | 3 | Urbanization | Removed | Calories | Removed |
| 4 | - | | Urbanization | 0.680 | 4 | - | - | Urbanization | Removed |

Sex specific obesity prevalence is the percentage of defined population segment with a body mass index (BMI) of no less than 30 $kg/m^2$.

Data sources: Total calories data from the FAO's FAOSTAT; BMI ≥30 data from the WHO Global Health Observatory; GDP data from the World Bank; Urbanization data from WHO. Biological State Index ($I_{bs}$) was self-calculated with country specific fertility data published by the United Nations and the mortality data published by World Health Organization (WHO).

## Discussion

The worldwide trend of increased obesity prevalence may be a multi-factorial phenomenon with major contributions from the environmental factors and the genetics. By assessing the data from 191 countries on the prevalence rates of the sex-specific obesity, we have shown that, globally, countries which had greater value of the $I_{bs}$ (less opportunity for natural selection) have greater obesity prevalence rates in both males and females. These trends remained independent of the commonly considered drives (total caloric intake, urbanization and GDP) of obesity Our finding supports a recent study conducted by Budnik and Henneberg that countries with more relaxed natural selection may have greater obesity prevalence in total population [41].

Natural selection is about survival of the fittest individuals through the action of differential fertility and mortality in a population. Medical care service, especially in the developed world has made the selection relaxed abruptly in the last few generations [24, 25, 33]. However, it still acts on phenotype of observable characteristics of human populations [38, 39]. Country specific health care service level and public health policies may determine magnitude of natural selection. Over generations, the phenotype disparities of human populations caused by the different magnitudes of natural selection may be observable [38, 39, 65]. The effects of reduced natural selection on accumulation of genes of partially heritable diseases have explained the increasing prevalence of deformed nasal septa [66], Type 1 diabetes [36] and lacrimal bone defects [67]. Likewise, the mutations producing metabolic faulty genes, which contribute to obesity, may be accumulating due to relaxed natural selection.

Obesity is a morphological trait, but obese people are likely to develop a clustering of complications, such as cancer, diabetes, cardiovascular diseases (CVDs) *etc.* [68-70]. Some of these complications used to be very dangerous or fatal diseases, but they are not in the modern society because the advanced medical/health care can “cure” them [33, 36, 37], which keeps obese people “fat but fit” [71]. Therefore, obese people not only can live as well as other people with healthy body weight [71], but also can participate in the reproduction even if their life expectancy may be reduced [69, 72, 73]. This process may allow “fat but fit” people to pass their faulty genes into the next generation. Over a few generations, the effects of natural selection on metabolic fault genes accumulation in a population may become an observable phenotype (obesity and excessive thinness).

Differential fertility and differential mortality, acting singly or together, are the fundamental events of natural selection to determine the fitness (successful reproduction) of a particular population in a given environment [65]. Country specific value of $I_{bs}$ which

is calculated with the country specific fertility and mortality rates may express different successful average reproduction opportunities of their inhabitants [33]. The reproductive success opportunity (indexed with $I_{bs}$ value) of each population may determine their level of accumulation of unfavourable for metabolism genes, and thus may influence prevalence rate of people of abnormal body mass – too thin and too fat. This may partially explain that countries with high level of medical care for long time may have more obesity issues due to greater accumulation of metabolically faulty genes. This theory has been successfully tested in the relationship between reduced selection and obesity prevalence [30].

The other important finding in this study was that the $I_{bs}$ was in significantly stronger correlation to male obesity prevalence than to female obesity prevalence. Theoretically, metabolic faults may be cumulative in females and males at the same pace in the process of relaxation of natural selection. Accordingly, the $I_{bs}$ should be correlated to the obesity prevalence equally in females and males. The significantly weaker relationship between $I_{bs}$ and female obesity prevalence in some of our analyses may indicate that the effects of reduced natural selection on obesity are moderated by environmental factors more in females than in males. In other words, the same magnitude of metabolic faulty mutations accumulation due to the reduced natural selection in males and females does not lead to the same phenotypes at population level (different obesity prevalence rates in males and females). Multiple environmental factors that may influence the female obesity prevalence in different countries or regions may explain the disparity of obesity prevalence in males and females. Female obesity prevalence, in general, correlates less strongly with country-characteristic variables than male obesity (Table 2). It may be the result of individual females' decisions concerning their body mass being driven by requirements of fashion to a larger extent than those of males. It may, however, also reflect results of industrialisation and economic situation because the ratio of male to female obesity per country shows linear and strong correlation ($r=0.77$, $P<0.001$) to GDP with male/female ratios being less than one in countries with GDP below about 25,000 USD and above 1 in wealthier countries [49]. The authors interpreted this as a result of greater presence of xenoestrogens in environments of wealthier countries, but there may be other reasons.

Fertility is a nutritionally expensive process for women due to gestation and lactation [48]. Therefore, women at reproductive age have been especially susceptible to excessive fat storage from the perspective of evolutionary biology [48]. Birth rates are low in developed countries, but high in developing countries [74, 75]. Nutrition stored in the form of fatness

in females of developed countries, which is supposed to be used for successful reproduction, is simply kept without use, which increases body weight of females in the developed world. This is a result of conscious birth control, unrelated to genetic variation.

Oestrogen is the primary female sex hormone. Higher levels of oestrogen have been associated with greater adiposity in females [49, 76]. It also has been shown that xenoestrogens increase obesity [77]. Low birth rates in developed world [75] may make females exposed to more oestrogen, which may increase fat storage.

Toward the end of the 20$^{th}$ century, there has been a transition away from agricultural labor (both for production and subsistence) to wage labor in many developing countries. This transition has decreased the physical activity of women more than men [78, 79].

Importantly, worldwide, different sociocultural beliefs and practices may also affect female disparities in excessive weight gain [80-84]. To a large extent, females may artificially change their fat accumulation resulting from the genetic endowment. In general, women are socialized to be more appearance-focused than men [80], which makes females more prone to adjust their body weight to meet the expected appearance of the specific sociocultural beliefs. For instance, females have been overprotected and, due to cultural or religious barriers, cannot publicly participate in physical activity in conservative societies, such as in the developing countries in the Middle East and North Africa region and the developed countries of Oman, Kuwait, and Saudi Arabia [85-87]. On the other hand, in the “Western” countries, the female body ideal has been that of a thin person for the last 50 years.

In this study, the curvilinear correlation was applied as the Kolmogorof-Smirnov and Shapiro-Wilk detected that the data distributions are not normal. It is revealed that $I_{bs}$ is correlated to sex-specific obesity prevalence residuals which were obtained by removing the contributing effects of non-genetic (environmental) factors from obesity prevalence, but there is no significant difference between the two correlations within the 1$^{st}$, 2$^{nd}$ and 3$^{rd}$ order residuals. This finding may complement our hypothesis because this may imply that reduced natural selection has increased the frequencies of obesity genes/mutations in males and females equally.

From evolutionary perspective, hypotheses of thrifty gene [88], drifty gene [89] and poor adaptation arising from migration [90] have been proposed to explain modern obesity pandemic. All these three hypotheses would require thousands of years evolution to slowly accumulate the genetic background of obesity. This makes these hypotheses inapplicable to our study as we are advancing a hypothesis that metabolic faults caused

by mutations have been accumulating in human populations at previously unexpected speed [27-29, 33, 91] because natural selection has been reduced sharply in the last 100-150 years [30, 33, 36, 40]. Our hypothesis also implies that modern humans may not be naturally well adapted to the current environment because their survival capacity and "fitness" are maintained by application of advanced technologies and medical services [30, 33, 36, 40]. This implication may not be inferred from the other three hypotheses.

Supported by the drifty gene theory, the Insurance Hypothesis (IH) advanced that food insecurity, instead of food abundance, may contribute to obesity [92], and it was found that in high income populations, perceived food insecurity due to social inequalities was associated more with obesity prevalence among adult women than men [92]. To test this hypothesis, the country specific Gini index (mean of 2008-2012), measuring insecurity level due to social inequality, was captured [93] for exploring the correlation between food insecurity and sex-specific obesity prevalence in the developed countries (those with GDP per capita within the top quartile of the world). The Pearson's r and non-parametric and partial correlation analyses did not establish any strong or significant correlation between Gini index and male or female obesity prevalence (Table 4).

Table 6 Correlation between Gini index and obesity prevalence in the developed world

| | Pearson r | | | Spearman's rho | | | Partial Correlation | | |
|---|---|---|---|---|---|---|---|---|---|
| | r | p | n | r | p | n | r | p | df |
| Male obesity prevalence | -0.039 | 0.837 | 30 | -0.063 | 0.742 | 30 | -0.247 | 0.223 | 24 |
| Female obesity prevalence | 0.086 | 0.652 | 30 | 0.226 | 0.229 | 30 | -0.124 | 0.548 | 24 |
| Biological State Index ($I_{bs}$) | -0.272 | 0.145 | 30 | -0.083 | 0.661 | 30 | - | - | - |
| Calories | 0.162 | 0.393 | 30 | 0.237 | 0.208 | 30 | - | - | - |
| GDP | -0.078 | 0.681 | 30 | -0.078 | 0.682 | 30 | - | - | - |
| Urbanization | -0.190 | 0.314 | 30 | -0.033 | 0.862 | 30 | - | - | - |

Sex specific obesity prevalence is the percentage of defined population segment with a body mass index (BMI) of no less than 30 kg/m$^2$.

Data sources: Total calories data from the FAO's FAOSTAT; BMI ≥30 data from the WHO Global Health Observatory; GDP data from the World Bank; Urbanization data from WHO. Biological State Index ($I_{bs}$) was self-calculated with country specific fertility data published by the United Nations and the mortality data published by World Health Organization (WHO). Gini index from the World Bank.

Population-based prevention strategies targeting ‘‘obesogenic’’ environments have been advocated and adopted as a public health approach [94, 95]. However, unfortunately, no country has achieved their expected results in the past 30 years [96]. The process of natural selection reduction which has driven the accumulation of the energy balance and metabolic faulty genes/mutations in human populations may partially explain this phenomenon [33]. Random mutations are as likely to affect metabolism to produce too much adipose tissue as not to and reduce body mass excessively. There is, however, a simple imbalance between the two directions of metabolic faults – body mass of a living human being cannot be reduced below a certain level determined by the weight of musculo-skeletal, circulatory, urinary, reproductive, nervous and integumentary systems, while it can be doubled, tripled, or even, perhaps, quadrupled by increasing the amount of adipose and muscle tissue. This imbalance produces, on average, increase in body mass and in prevalence of obesity over that of underweight.

Several generations of people in Europe and North America have had the access to advanced medical care earlier and easier than those from the developing areas, such as Africa and Asia. This may be one of the reasons that obesity has become a noticeable pressing issue much earlier in the developed regions. For instance, Olshansky *et al.* reported that the life expectancy in the USA may be reduced if obesity prevalence keeps rising in the future [97].

Several limitations in this study need to be acknowledged:

First, the relationship between $I_{bs}$ and obesity prevalence reported here only shows coincidence, not causality.

Second, we could only demonstrate the relationship between the $I_{bs}$ and the obesity prevalence rate at country/population level, rather than at the individual level because both data analysed [34, 35] and the evolutionary approach [26] are population based.

Third, the changes in the genomes of human populations may be too slow to fully explain the increasing obesity prevalence. Obesity is the result of an unfavourable interaction between our genomes and our current environment which might play more important role in developing obesity in some circumstances.

Fourth, this study analysed the data across 191 countries. However, the results cannot be complemented by the longitudinal data analysis in individual countries, with exception of Australia and Poland [30] due to the fact that obesity only has been an issue in the last few decades. We could not access the combined obesity and $I_{bs}$ data which are older than 30 years.

Finally, the female complexities, adaptation for fertility [48], oestrogen [76] and double x chromosomes in cells [98] may have confounded our analysis of correlation of the $I_{bs}$ to female obesity prevalence, but we could not obtain data to reduce or avoid such confounding effects.

The natural selection has been universally reduced and this trend continues as, worldwide, the medical services keep improving quickly. In the past, eugenics has been proposed to improve genetic stock in humans, but it is unethical [99] and may potentially decrease gene diversity [100]. Instead of "people selection (eugenics)", recent advances in genome editing have made gene therapy possible [101]. For instance. Gendicine and Glybera have been used for treatment of head and neck squamous cell carcinoma [102] and lipoprotein lipase deficiency [103] respectively. The obesity related genes/mutations accumulation in human populations through the process of reduction of natural selection may become more and more imperative. Advances in our knowledge of the molecular basis of obesity and obesity-associated diseases, and development of gene therapy may offer an alternative long-term treatment modality in the near future.

## Conclusions

Recently accumulated high frequency of genes related to metabolic faults in human populations may be one of the important contributors to the increasing prevalence of obesity worldwide. The relaxed natural selection may have accumulated faulty genes in both males and females over successive generations. Reduced natural selection affecting less female obesity prevalence than its male equivalent may be attributable to female-specific physiological mechanisms and various socio-cultural practices. Public health approaches to develop population-based strategies for the prevention of excess weight gain may not be able to achieve expected results. Gene therapy should be considered as a solution to address the global problem of obesity.

## References


1. WHO. *Obesity and overweight*. 2015; Available from: http://www.who.int.
2. WHO, *Obesity: Preventing and Managing the Global Epidemic*. WHO Technical report series No. 894. 2004, Geneva World Health Organization 2000.
3. Puhl, R. and K.D. Brownell, *Bias, discrimination, and obesity.* Obesity research, 2001. **9**(12): p. 788-104.
4. WHO. *Obesity*. WHO 2015 [11.26.2015]; Available from: http://who.int/topics/obesity/en/.
5. WHO. *BMI Classification*. 2016 18/05/2016]; Available from: http://apps.who.int.
6. WHO, *Physical status: The use of and interpretation of anthropometry, Report of a WHO Expert Committee.* 1995.

7. O'Rahilly, S. and I.S. Farooqi, *Genetics of obesity.* Philos Trans R Soc Lond B Biol Sci, 2006. **361**(1471): p. 1095-105.

8. Bleich, S.N., et al., *Why Is the Developed World Obese*. 2008. p. 273-295.

9. Astrup, A. and J. Brand-Miller, *Diet composition and obesity.* Lancet (London, England), 2012. **379**(9821): p. 1100.

10. Drewnowski, A. and B.M. Popkin, *The Nutrition Transition: New Trends in the Global Diet*. 1997: Oxford, UK. p. 31-43.

11. Luke, A. and R.S. Cooper, *Physical activity does not influence obesity risk: time to clarify the public health message.* International Journal of Epidemiology, 2013. **42**(6): p. 1831-1836.

12. Blair, S.N., E. Archer, and G.A. Hand, *Commentary: Luke and Cooper are wrong: physical activity has a crucial role in weight management and determinants of obesity.* International Journal of Epidemiology, 2013. **42**(6): p. 1836-1838.

13. Hill, J.O. and J.C. Peters, *Commentary: Physical activity and weight control.* International Journal of Epidemiology, 2013. **42**(6): p. 1840-1842.

14. Swinburn, B., *Commentary: Physical activity as a minor player in the obesity epidemic: what are the deep implications?* International Journal of Epidemiology, 2013. **42**(6): p. 1838-1840.

15. Prentice, A. and S. Jebb, *Energy Intake/ Physical Activity Interactions in the Homeostasis of Body Weight Regulation.* Nutrition Reviews, 2004. **62**: p. S98-S104.

16. Jotham, S., et al., *Artificial sweeteners induce glucose intolerance by altering the gut microbiota.* Nature, 2014.

17. You, W. and M. Henneberg, *Meat consumption providing a surplus energy in modern diet contributes to obesity prevalence: an ecological analysis.* BMC Nutrition, 2016. **2**(1).

18. You, W. and M. Henneberg, *Cereal Crops Are not Created Equal: Wheat Consumption Associated with Obesity Prevalence Globally and Regionally.* AIMS Public Health, 2016. **3(2)**: p. 313-328.

19. You, W. and M. Henneberg, *Meat in Modern Diet, Just as Bad as Sugar, Correlates with Worldwide Obesity: An Ecological Analysis.* J Nutr Food Sci **6: 517**(4).

20. Peter, J.T., et al., *A core gut microbiome in obese and lean twins.* Nature, 2008. **457**(7228): p. 480.

21. Greenblum, S., P.J. Turnbaugh, and E. Borenstein, *Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease.(MICROBIOLOGY)(Author abstract)(Report).* Proceedings of the National Academy of Sciences of the United States, 2012. **109**(2): p. 594.

22. Tilg, H. and A. Kaser, *Gut microbiome, obesity, and metabolic dysfunction.(Review series).* Journal of Clinical Investigation, 2011. **121**(6): p. 2126.

23. Peter, J.T., et al., *An obesity- associated gut microbiome with increased capacity for energy harvest.* Nature, 2006. **444**(7122): p. 1027.

24. Courtiol, A., et al., *Natural and sexual selection in a monogamous historical human population.* Proceedings of the National Academy of Sciences, 2012. **109**(21): p. 8044-8049.

25. Byars, S.G., et al., *Colloquium papers: Natural selection in a contemporary human population.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107 Suppl 1**: p. 1787-92.

26. Hall, B.K.H., Benedikt, *Strickberger's Evolution (4th ed.)*. 2008: Sudbury, MA: Jones and Bartlett Publishers. ISBN 978-0-7637-0066-9. LCCN 2007008981. OCLC 85814089.

27. Conrad, D.F., et al., *Variation in genome- wide mutation rates within and between human families.* Nature Genetics, 2011. **43**(7): p. 712.

28. Crow, J.F., *The origins, patterns and implications of human spontaneous mutation.* Nature Reviews Genetics, 2000. **1**(1): p. 40-47.

29. Henn, B.M., et al., *Estimating the mutation load in human genomes.* Nature Reviews Genetics, 2015. **16**(6): p. 333-343.

30. Budnik, A. and M. Henneberg, *Worldwide increase of obesity is related to the reduced opportunity for natural selection.* PloS one, 2017. **12**(1): p. e0170098.

31. Gale, E.A.M., *The rise of childhood type 1 diabetes in the 20th century.* Diabetes, 2002. **51**(12): p. 3353-61.

32. Medawar, P.B., *Do advances in medicine lead to genetic deterioration?*, in *Natural Selection in Human Populations.*, Bajema C. J., Editor. 1971, Robert E. Krieger Publishing Co.: New York. p. 300-08.

33. Stephan, C.N. and M. Henneberg, *Medicine may be reducing the human capacity to survive.* Medical Hypotheses, 2001. **57**(5): p. 633-37.

34. Henneberg, M. and J. Piontek, *Biological state index of human groups.* Przeglad Anthropologiczny, 1975. **XLI**: p. 191-201.

35. Henneberg, M., *Reproductive possibilities and estimations of the biological dynamics of earlier human populations.* Journal of Human Evolution, 1976. **5**: p. 41-8.

36. You, W.-P. and M. Henneberg, *Type 1 diabetes prevalence increasing globally and regionally: the role of natural selection and life expectancy at birth.* BMJ Open Diabetes Research & Care, 2016. **4**(1): p. e000161.

37. Saniotis, A. and M. Henneberg, *Medicine could be constructing human bodies in the future.* Medical Hypotheses, 2011. **77**(4): p. 560-64.

38. You, W. and H. M, *Cancer incidence increasing globally: The role of relaxed natural selection.* Evol Appl., 2017. **00:1–13**.

39. Rühli, F. and M. Henneberg, *Biological future of humankind – ongoing evolution and the impact of recognition of human biological variation*, in *On Human Nature. Biology, Psychology, Ethics, Politics, and Religion 2016*, M. Tibayrenc and F.J. Ayala, Editors. 2016, Elsevier. p. 263-275.

40. Henneberg, M., *The rate of human morphological microevolution and taxonomic diversity of hominids.* Studies in Historical Anthropology, 2006. **4**(2004): p. 49-59.

41. Budnik, A. and M. Henneberg, *Worldwide Increase of Obesity is Related to the Reduced Opportunity for Natural Selection.* PLOS One, 2017. **12**(1).

42. WHO. *Global Health Observatory, the data repository*. WHO 2015 [11.26.2015]; Available from: http://www.who.int/gho/database/en/.

43. Camhi, S.M., et al., *The relationship of waist circumference and BMI to visceral, subcutaneous, and total body fat: sex and race differences.* Obesity, 2011. **19**(2): p. 402-408.

44. Schreiner, P.J., et al., *Sex-specific associations of magnetic resonance imaging-derived intra-abdominal and subcutaneous fat areas with conventional anthropometric indices: The Atherosclerosis Risk in Communities Study.* American journal of epidemiology, 1996. **144**(4): p. 335-345.

45. Womersley, J. and J. Durnin, *A comparison of the skinfold method with extent of 'overweight'and various weight-height relationships in the assessment of obesity.* British Journal of Nutrition, 1977. **38**(2): p. 271-284.

46. Jackson, A., et al., *The effect of sex, age and race on estimating percentage body fat from body mass index: The Heritage Family Study.* International journal of obesity, 2002. **26**(6): p. 789.

47. Gallagher, D., et al., *How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups?* American journal of epidemiology, 1996. **143**(3): p. 228-239.

48. Power, M.L. and J. Schulkin, *Sex differences in fat storage, fat metabolism, and the health risks from obesity: possible evolutionary origins.* Br J Nutr, 2008. **99**(5): p. 931-40.

49. Grantham, J.P. and M. Henneberg, *The estrogen hypothesis of obesity.* PLoS One, 2014. **9**(6): p. e99776.

50. Stunkard, A.J., et al., *The body-mass index of twins who have been reared apart.* New England journal of medicine, 1990. **322**(21): p. 1483-1487.

51. Schousboe, K., et al., *Sex differences in heritability of BMI: a comparative study of results from twin studies in eight countries.* Twin Research and Human Genetics, 2003. **6**(5): p. 409-421.

52. Darwin, C., *The descent of man, and selection in relation to sex*. New ed. 1901, London: John Murray.

53. The World Health Organization. *GHO | By category | Obesity (body mass index ≥ 30),age-standardized (%) - Estimates by country*. WHO 05 January 2017 MINERVA PUBLISH DATE 28 August 2017]; Available from: http://apps.who.int/gho/data/view.main.CTRY2450A.

54. The United Nations. *World Fertility Data 2008*. 2012 29.07.2015]; Available from: http://www.un.org.

55. WHO, *World Health Statistics 2012*. Life tables for WHO Member States. 2012, Geneva: World Health Organization.

56. WHO. *Urbanization and health*. WHO 2010 2010-12-07 15:20:05 2 November 2016]; Available from: http://www.who.int/bulletin/volumes/88/4/10-010410/en/.

57. FAO, *Food Balance Sheets. A Handbook*. 2001, Rome: Food and Agriculture Organization.

58. The World Bank: International Comparison Program database: World Development Indicators. *GDP (current US$) per capita per year* 2010 [11.26.2015]; Available from: http://data.worldbank.org.

59. Nestle, M., *Increasing portion sizes in American diets: More calories, more obesity.* Journal of the American Dietetic Association, 2003. **103**(1): p. 39-40.

60. Giskes, K., et al., *Socioeconomic position at different stages of the life course and its influence on body weight and weight gain in adulthood: a longitudinal study with 13-year follow-up.* Obesity (Silver Spring), 2008. **16**(6): p. 1377-1381.

61. Pirgon, Ö. and N. Aslan, *The Role of Urbanization in Childhood Obesity.* J Clin Res Pediatr Endocrinol, 2015. **7**(3): p. 163-167.

62. Fezeu, L., et al., *Waist circumference and obesity-related abnormalities in French and Cameroonian adults: the role of urbanization and ethnicity.* International Journal of Obesity, 2010. **34**(3): p. 446-53.

63. Kjellström, T., C. Håkansta, and C. Hogstedt, *Globalisation and public health - overview and a Swedish perspective.* Scand. J. of Public Hlth., 2007. **35**: p. 2-68.

64. Livingstone, F.B. and J.N. Sphuler, *Cultural determinants of natural selection*, in *United Nations Educational, Scientific and Cultural Organization: Expert Meeting on the Biological Aspect of Race*. 1964: Mosco. p. 1-3.

65. Post, R.H., *Deformed nasal septa and relaxed selection.* Eugenics Quarterly, 1966. **13**: p. 101-12.

66. Post, R.H., *Population differences in tear duct size implications of relaxed selection.* Soc Biol 969 **16**: p. 257-69.

67. WHO, *Global Health Risks Mortality and Burden of Disease Attributable to Selected Major Risks*. 2009, Geneva: Geneva : World Health Organization.

68. Haslam, D. and W. James, *Obesity* The Lancet (Review), 2005. **366**(9492): p. 1197–209.

69. Thompson, D., et al., *Lifetime health and economic consequences of obesity.* Archives of internal medicine, 1999. **159**(18): p. 2177-2183.

70. Lee, C., A. Jackson, and S. Blair, *US weight guidelines: is it also important to consider cardiorespiratory fitness?* International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity, 1998. **22**: p. S2-7.

71. Peeters, A., et al., *Obesity in adulthood and its consequences for life expectancy: a life-table analysis.* Annals of internal medicine, 2003. **138**(1): p. 24-32.

72. Collaboration, P.S., *Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies.* The Lancet, 2009. **373**(9669): p. 1083-1096.

73. Neel, J.V., *Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"?* American journal of human genetics, 1962. **14**(4): p. 353.

74. Speakman, J.R., *Thrifty genes for obesity, an attractive but flawed idea, and an alternative perspective: the'drifty gene'hypothesis.* International journal of obesity, 2008. **32**(11): p. 1611.

75. Sellayah, D., F.R. Cagampang, and R.D. Cox, *On the evolutionary origins of obesity: a new hypothesis.* Endocrinology, 2014. **155**(5): p. 1573-1588.

76. Lynch, M., *Mutation and Human Exceptionalism: Our Future Genetic Load.* Genetics, 2016. **202**(3): p. 869-75.

77. Nettle, D., C. Andrews, and M. Bateson, *Food insecurity as a driver of obesity in humans: The insurance hypothesis.* Behavioral and Brain Sciences, 2016. **40**: p. 1-34.

78. The World Bank. *GINI index (World Bank estimate)*. 2017 27 August 2017]; Available from: http://data.worldbank.org/indicator/SI.POV.GINI.

79. WHO, *Obesity: preventing and managing the global epidemic/report of a WHO Consultation*. 2000, Geneva: World Health Organization.

80. Aranceta, J., et al., *Prevention of overweight and obesity from a public health perspective.* Nutrition Reviews, 2009. **67**: p. S83-S88.

81. Ng, M., et al., *Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013.* The Lancet, 2014. **384**(9945): p. 766-781.

82. Olshansky, S.J., et al., *A Potential Decline in Life Expectancy in the United States in the 21st Century.* The New England Journal of Medicine, 2005. **352**(11): p. 1138-1145.

83. Nargund, G., *Declining birth rate in Developed Countries_A radical policy re-think is required.* Facts Views Vis Obgyn, 2009. **1**(3): p. 191–193.

84. Brown, L.M., et al., *Metabolic impact of sex hormones on obesity.* Brain Research, 2010. **1350**: p. 77-85.

85. Roccisano, D. and M. Henneberg, *Soy Consumption and Obesity.* Food and Nutrition Sciences, 2012. **03**(02): p. 260-266.

86. McGarvey, S.T., *Obesity in Samoans and a perspective on its etiology in Polynesians.* The American journal of clinical nutrition, 1991. **53**(6 Suppl): p. 1586S.

87. Snodgrass, J.J., et al., *The Emergence of Obesity among Indigenous Siberians.* Journal of Physiological Anthropology (JPA) 2006. **25**(1): p. 75-84.

88. Park, L.E., A.M. DiRaddo, and R.M. Calogero, *Sociocultural influence and appearance-based rejection sensitivity among college students.* Psychology of Women Quarterly, 2009. **33**(1): p. 108-119.

89. McLaren, L., *Socioeconomic status and obesity.* Epidemiologic reviews, 2007. **29**(1): p. 29-48.

90. Monteiro, C.A., et al., *Socioeconomic status and obesity in adult populations of developing countries: a review.* Bulletin of the World Health Organization, 2004. **82**(12): p. 940-946.

91. Howe, L.D., R. Patel, and B. Galobardes, *Commentary: Tipping the balance: wider waistlines in men but wider inequalities in women.* International journal of epidemiology, 2010. **39**(2): p. 404-405.

92. Kanter, R. and B. Caballero, *Global gender disparities in obesity: a review.* Advances in Nutrition: An International Review Journal, 2012. **3**(4): p. 491-498.

93. Al-Lawati, J.A. and P.J. Jousilahti, *Prevalence and 10- year secular trend of obesity in Oman.* Saudi medical journal, 2004. **25**(3): p. 346.

94. Al-Riyami, A.A. and M.M. Afifi, *Prevalence and correlates of obesity and central obesity among Omani adults.* Saudi medical journal, 2003. **24**(6): p. 641.

95. Al-Kandari, Y.Y., *Prevalence of obesity in Kuwait and its relation to sociocultural variables.* Obesity Reviews, 2006. **7**(2): p. 147-154.

96. Chen, X., et al., *The number of x chromosomes causes sex differences in adiposity in mice.* PLoS Genet, 2012. **8**(5): p. e1002709.

97. Black, E. and R. Moore, *War Against the Weak; Eugenics and America's Campaign to Create a Master Race.* RADICAL STATISTICS, 2008. **96**: p. 127.

98. Miller, E.M., *Eugenics: economics for the long run.* Research in biopolitics, 1997. **5**: p. 391-416.

99. Hwang, W.Y., et al., *Efficient genome editing in zebrafish using a CRISPR-Cas system.* Nature biotechnology, 2013. **31**(3): p. 227-229.

100. Pearson, S., H. Jia, and K. Kandachi, *China approves first gene therapy.* Nature Biotechnology 2004. **22**(3 - 4).

101. Gallagher, J. *Gene therapy: Glybera approved by European commission*. BBC News. Nov2 2012 31 August 2017]; Available from: http://www.bbc.com/news/health-20179561.

## Additional files

Additional File 1:

Table AF 1: $I_{bs}$ values for 191 countries [1]

| Country | $I_{bs}$ Value | Country | $I_{bs}$ Value | Country | $I_{bs}$ Value | Country | $I_{bs}$ Value |
|---|---|---|---|---|---|---|---|
| Afghanistan | 0.71993 | Dominican Republic | 0.95615 | Luxembourg | 0.99236 | Singapore | 0.99372 |
| Albania | 0.97509 | Ecuador | 0.96240 | Macedonia, FYR | 0.98626 | Slovakia | 0.98890 |
| Algeria | 0.95412 | Egypt | 0.97077 | Madagascar | 0.91151 | Slovenia | 0.98959 |
| Andorra | 0.99196 | El Salvador | 0.96938 | Malawi | 0.77708 | Solomon Islands | 0.94653 |
| Angola | 0.76887 | Equatorial Guinea | 0.76681 | Malaysia | 0.98480 | Somalia | 0.72700 |
| Antigua & Barbuda | 0.97687 | Eritrea | 0.91484 | Maldives | 0.97995 | South Africa | 0.81135 |
| Argentina | 0.97630 | Estonia | 0.98915 | Mali | 0.77401 | Spain | 0.99165 |
| Armenia | 0.97639 | Ethiopia | 0.81512 | Malta | 0.99007 | Sri Lanka | 0.96877 |
| Australia | 0.98989 | Fiji | 0.96733 | Marshall Islands | 0.93130 | St. Kitts and Nevis | 0.97952 |
| Austria | 0.99094 | Finland | 0.99088 | Mauritania | 0.82887 | St. Lucia | 0.97066 |
| Azerbaijan | 0.95952 | France | 0.99155 | Mauritius | 0.97732 | St. Vincent & the Grenadines | 0.97710 |
| Bahamas | 0.97392 | Gabon | 0.88953 | Mexico | 0.97618 | Sudan | 0.82098 |
| Bahrain | 0.98112 | Gambia | 0.84868 | Micronesia, Fed. Sts. | 0.93688 | Suriname | 0.95629 |
| Bangladesh | 0.92070 | Georgia | 0.96274 | Moldova | 0.97951 | Swaziland | 0.80146 |
| Barbados | 0.98129 | Germany | 0.99209 | Mongolia | 0.96440 | Sweden | 0.99242 |
| Belarus | 0.98294 | Ghana | 0.88456 | Montenegro | 0.98737 | Switzerland | 0.99290 |

| Country | $I_{bs}$ Value | Country | $I_{bs}$ Value | Country | $I_{bs}$ Value | Country | $I_{bs}$ Value |
|---|---|---|---|---|---|---|---|
| Belgium | 0.99016 | Greece | 0.99101 | Morocco | 0.95492 | Syrian Arab Republic | 0.97597 |
| Belize | 0.96940 | Grenada | 0.97452 | Mozambique | 0.75217 | Tajikistan | 0.93765 |
| Benin | 0.83445 | Guatemala | 0.93855 | Myanmar | 0.90156 | Tanzania | 0.82970 |
| Bhutan | 0.89433 | Guinea | 0.79173 | Namibia | 0.88120 | Thailand | 0.97093 |
| Bolivia | 0.93065 | Guinea-Bissau | 0.73362 | Nauru | 0.92612 | Timor-Leste | 0.92455 |
| Bosnia & Herzegovina | 0.93065 | Guyana | 0.94943 | Nepal | 0.93168 | Togo | 0.85388 |
| Botswana | 0.88361 | Haiti | 0.87088 | Netherlands | 0.99118 | Tonga | 0.95252 |
| Brazil | 0.97459 | Honduras | 0.95493 | New Zealand | 0.98754 | Trinidad & Tobago | 0.95607 |
| Brunei Darussalam | 0.98497 | Hungary | 0.98901 | Nicaragua | 0.96483 | Tunisia | 0.97206 |
| Bulgaria | 0.98359 | Iceland | 0.99434 | Niger | 0.79660 | Turkey | 0.97458 |
| Burkina Faso | 0.63476 | India | 0.89826 | Nigeria | 0.78223 | Turkmenistan | 0.93614 |
| Burundi | 0.73592 | Indonesia | 0.94507 | Niue | 0.98062 | Tuvalu | 0.90554 |
| Cambodia | 0.88794 | Iran | 0.96320 | Norway | 0.99131 | Uganda | 0.81707 |
| Cameroon | 0.76039 | Iraq | 0.93584 | Oman | 0.97720 | Ukraine | 0.97745 |
| Canada | 0.98911 | Ireland | 0.98989 | Pakistan | 0.87655 | United Arab Emirates | 0.98692 |
| Cape Verde | 0.96350 | Israel | 0.99116 | Palau | 0.97515 | United Kingdom | 0.98988 |
| Central African Republic | 0.71339 | Italy | 0.99201 | Panama | 0.96705 | Uruguay | 0.97986 |
| Chad | 0.70549 | Jamaica | 0.95615 | Papua New Guinea | 0.89240 | USA | 0.98535 |
| Chile | 0.98569 | Japan | 0.99324 | Paraguay | 0.96662 | Uzbekistan | 0.95421 |
| China | 0.96889 | Jordan | 0.96117 | Peru | 0.96741 | Vanuatu | 0.96286 |

| Country | $I_{bs}$ Value | Country | $I_{bs}$ Value | Country | $I_{bs}$ Value | Country | $I_{bs}$ Value |
|---|---|---|---|---|---|---|---|
| Colombia | 0.97377 | Kazakhstan | 0.95456 | Philippines | 0.95294 | Venezuela | 0.97379 |
| Comoros | 0.85001 | Kenya | 0.87165 | Poland | 0.98881 | Viet Nam | 0.96584 |
| Congo | 0.81029 | Kiribati | 0.92477 | Portugal | 0.99074 | Yemen | 0.90227 |
| Congo, Dem. Rep. | 0.72917 | Korea, Dem. Rep. | 0.95241 | Qatar | 0.98518 | Zambia | 0.76135 |
| Cook Islands | 0.97249 | Korea, Rep. | 0.98835 | Romania | 0.98266 | Zimbabwe | 0.80874 |
| Costa Rica | 0.98362 | Kuwait | 0.98038 | Russian Federation | 0.97611 | | |
| Côte d'Ivoire | 0.78317 | Kyrgyzstan | 0.95104 | Rwanda | 0.83990 | | |
| Croatia | 0.98889 | Lao PDR | 0.90351 | Samoa | 0.96475 | | |
| Cuba | 0.98887 | Latvia | 0.98514 | Sao Tome and Principe | 0.91189 | | |
| Cyprus | 0.99406 | Lebanon | 0.97970 | Saudi Arabia | 0.96580 | | |
| Czech Republic | 0.99174 | Lesotho | 0.78464 | Senegal | 0.87041 | | |
| Denmark | 0.99132 | Liberia | 0.81816 | Serbia | 0.98868 | | |
| Djibouti | 0.84585 | Libya | 0.96632 | Seychelles | 0.97996 | | |
| Dominica | 0.98204 | Lithuania | 0.98782 | Sierra Leone | 0.72789 | | |

1. Budnik, A. and M. Henneberg, Worldwide Increase of Obesity is Related to the Reduced Opportunity for Natural Selection. PLOS One, 2017. 12(1).

Additional File 2:

## Table AF 2: Multicollinearity tests amongst the variables

Table AF 2-1: Multicollinearity tests of male and female obesity prevalence to other predictors

| Independent Variables | Tolerance | VIF | Independent Variables | Tolerance | VIF |
|---|---|---|---|---|---|
| Urbanization | 0.535 | 1.869 | Urbanization | 0.535 | 1.869 |
| $I_{bs}$ | 0.471 | 2.124 | $I_{bs}$ | 0.471 | 2.124 |
| Calories | 0.386 | 2.590 | Calories | 0.386 | 2.590 |
| GDP | 0.299 | 3.344 | GDP | 0.299 | 3.344 |

Table AF 2-2: Multicollinearity tests of calories to other predictors in male and female samples respectively

| Independent Variables | Tolerance | VIF | Independent Variables | Tolerance | VIF |
|---|---|---|---|---|---|
| GDP | 0.304 | 3.290 | GDP | 0.358 | 2.793 |
| Urbanization | 0.542 | 1.846 | BMI ≥ 30, Female | 0.697 | 1.434 |
| BMI ≥ 30, Male | 0.324 | 3.084 | Urbanization | 0.558 | 1.791 |
| $I_{bs}$ | 0.424 | 2.360 | $I_{bs}$ | 0.474 | 2.108 |

Table AF 2-3: Multicollinearity tests of GDP to other predictors in male and female samples respectively

| Independent Variables | Tolerance | VIF | Independent Variables | Tolerance | VIF |
|---|---|---|---|---|---|
| Urbanization | 0.553 | 1.809 | BMI ≥ 30, Female | 0.712 | 1.404 |
| BMI ≥ 30, Male | 0.351 | 2.848 | Urbanization | 0.580 | 1.723 |
| $I_{bs}$ | 0.453 | 2.207 | $I_{bs}$ | 0.548 | 1.826 |
| Calories | 0.425 | 2.353 | Calories | 0.472 | 2.119 |

Table AF 2-4: Multicollinearity tests of $I_{bs}$ to other predictors in male and female samples respectively

| Independent Variables | Tolerance | VIF | Independent Variables | Tolerance | VIF |
|---|---|---|---|---|---|
| Calories | 0.376 | 2.659 | Calories | 0.395 | 2.529 |
| GDP | 0.288 | 3.476 | GDP | 0.346 | 2.888 |
| Urbanization | 0.521 | 1.918 | BMI ≥ 30, Female | 0.709 | 1.410 |
| BMI ≥ 30, Male | 0.351 | 2.848 | Urbanization | 0.530 | 1.886 |

Table AF 2-5: Multicollinearity tests of urbanization to other predictors in male and female samples respectively

| Independent Variables | Tolerance | VIF | Independent Variables | Tolerance | VIF |
|---|---|---|---|---|---|
| $I_{bs}$ | 0.419 | 2.385 | $I_{bs}$ | 0.461 | 2.170 |
| BMI ≥ 30, Male | 0.321 | 3.116 | Calories | 0.405 | 2.472 |
| Calories | 0.387 | 2.586 | GDP | 0.319 | 3.134 |
| GDP | 0.282 | 3.542 | BMI ≥ 30, Female | 0.701 | 1.427 |

**Note for the above 5 tables:**

A tolerance of less than 0.20 or a VIF of above 5 indicates a multicollinearity problem.

Sex specific obesity prevalence is the percentage of defined population segment with a body mass index (BMI) of no less than 30 kg/$m^2$.

Data sources: Total calories data from the FAO's FAOSTAT; BMI ≥30 data from the WHO Global Health Observatory; GDP data from the World Bank; Urbanization data from WHO. Biological State Index ($I_{bs}$) was self-calculated with country specific fertility data published by the United Nations and the mortality data published by World Health Organization (WHO). Gini index from the World Bank.

Additional File 3:

Table AF3: Results of Mixed Model Analysis with the data nested within WHO regions and country groupings with greater and lower median of $I_{bs}$

Mixed Model Analysis (Male)

| | | All Countries in the Region | | | Countries with Ibs ≥0.9658 | | | Countries with Ibs <0.9658 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| WHO Region | Variables | Count | Mean | Standard Deviation | Count | Mean | Standard Deviation | Count | Mean | Standard Deviation |
| Africa | BMI ≥ 30 Male | 40 | 5.34 | 4.03 | 1 | 11.20 | | 39 | 5.19 | 3.97 |
| | ROGDP | 40 | 21.89 | 19.11 | 1 | 61.96 | | 39 | 20.86 | 18.21 |
| | ROUrban | 40 | 31.14 | 20.04 | 1 | 31.94 | | 39 | 31.12 | 20.30 |
| | ROCalories | 40 | 26.76 | 20.88 | 1 | 67.44 | | 39 | 25.72 | 20.07 |
| Americas | BMI ≥ 30 Male | 35 | 18.91 | 5.31 | 24 | 20.72 | 4.59 | 11 | 14.95 | 4.71 |
| | ROGDP | 35 | 58.82 | 18.86 | 24 | 66.08 | 14.20 | 11 | 42.98 | 18.54 |
| | ROUrban | 35 | 57.46 | 27.84 | 24 | 62.78 | 29.11 | 11 | 45.84 | 21.65 |
| | ROCalories | 35 | 46.06 | 22.71 | 24 | 53.03 | 22.32 | 11 | 30.87 | 15.39 |
| Eastern Mediterranean | BMI ≥ 30 Male | 17 | 18.96 | 10.65 | 7 | 27.14 | 5.90 | 10 | 13.23 | 9.48 |
| | ROGDP | 17 | 49.94 | 25.19 | 7 | 67.86 | 21.01 | 10 | 37.39 | 20.22 |
| | ROUrban | 17 | 61.75 | 28.64 | 7 | 75.77 | 21.71 | 10 | 51.94 | 29.73 |
| | ROCalories | 17 | 57.35 | 31.71 | 7 | 82.14 | 8.99 | 10 | 40.00 | 30.35 |
| Europe | BMI ≥ 30 Male | 50 | 20.87 | 3.89 | 42 | 21.87 | 2.97 | 8 | 15.59 | 4.04 |
| | ROGDP | 50 | 70.76 | 23.20 | 42 | 76.86 | 18.56 | 8 | 38.72 | 18.79 |
| | ROUrban | 50 | 64.38 | 20.75 | 42 | 70.19 | 16.45 | 8 | 33.84 | 12.65 |
| | ROCalories | 50 | 75.07 | 20.59 | 42 | 79.93 | 16.71 | 8 | 49.56 | 21.15 |
| South-East Asia | BMI ≥ 30 Male | 9 | 3.42 | 1.55 | 3 | 4.70 | 1.18 | 6 | 2.78 | 1.35 |
| | ROGDP | 9 | 31.40 | 15.80 | 3 | 47.46 | 12.36 | 6 | 23.37 | 10.29 |
| | ROUrban | 9 | 20.83 | 12.99 | 3 | 23.73 | 16.64 | 6 | 19.37 | 12.32 |
| | ROCalories | 9 | 32.11 | 15.47 | 3 | 38.76 | 10.80 | 6 | 28.78 | 17.22 |
| West Pacific | BMI ≥ 30 Male | 17 | 16.01 | 12.66 | 9 | 14.46 | 11.68 | 8 | 17.76 | 14.28 |
| | ROGDP | 17 | 50.70 | 28.01 | 9 | 69.69 | 25.27 | 8 | 29.35 | 8.90 |
| | ROUrban | 17 | 47.06 | 33.22 | 9 | 67.83 | 27.97 | 8 | 23.69 | 21.20 |
| | ROCalories | 17 | 49.76 | 21.50 | 9 | 62.66 | 15.37 | 8 | 35.25 | 18.21 |
| Worldwide | BMI ≥ 30 Male | 168 | 15.14 | 9.22 | 86 | 20.48 | 6.60 | 82 | 9.54 | 8.23 |
| | ROGDP | 168 | 50.39 | 28.58 | 86 | 71.17 | 19.00 | 82 | 28.60 | 19.13 |
| | ROUrban | 168 | 50.67 | 28.10 | 86 | 66.27 | 23.87 | 82 | 34.31 | 22.37 |
| | ROCalories | 168 | 50.87 | 28.74 | 86 | 69.21 | 21.96 | 82 | 31.63 | 21.57 |

Mixed Model Analysis (Female)

| | | All Countries in the Region | | | Countries with Ibs ≥0.9658 | | | Countries with Ibs <0.9658 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| WHO Region | Variables | Count | Mean | Standard Deviation | Count | Mean | Standard Deviation | Count | Mean | Standard Deviation |
| Africa | BMI ≥ 30 Female | 40 | 15.55 | 7.59 | 1 | 24.30 | NA | 39 | 15.33 | 7.55 |
| | ROGDP | 40 | 21.89 | 19.11 | 1 | 61.96 | NA | 39 | 20.86 | 18.21 |
| | ROUrbanization | 40 | 31.14 | 20.04 | 1 | 31.94 | NA | 39 | 31.12 | 20.30 |
| | ROCalories | 40 | 26.76 | 20.88 | 1 | 67.44 | NA | 39 | 25.72 | 20.07 |
| Americas | BMI ≥ 30 Female | 35 | 29.99 | 5.80 | 24 | 31.21 | 5.07 | 11 | 27.33 | 6.62 |
| | ROGDP | 35 | 58.82 | 18.86 | 24 | 66.08 | 14.20 | 11 | 42.98 | 18.54 |
| | ROUrbanization | 35 | 57.46 | 27.84 | 24 | 62.78 | 29.11 | 11 | 45.84 | 21.65 |
| | ROCalories | 35 | 46.06 | 22.71 | 24 | 53.03 | 22.32 | 11 | 30.87 | 15.39 |
| Eastern Mediterranean | BMI ≥ 30 Female | 17 | 29.85 | 13.30 | 7 | 39.60 | 4.38 | 10 | 23.02 | 13.26 |
| | ROGDP | 17 | 49.94 | 25.19 | 7 | 67.86 | 21.01 | 10 | 37.39 | 20.22 |
| | ROUrbanization | 17 | 61.75 | 28.64 | 7 | 75.77 | 21.71 | 10 | 51.94 | 29.73 |
| | Rocalories | 17 | 57.35 | 31.71 | 7 | 82.14 | 8.99 | 10 | 40.00 | 30.35 |
| Europe | BMI ≥ 30 Female | 50 | 22.74 | 4.02 | 42 | 22.99 | 4.09 | 8 | 21.39 | 3.56 |
| | ROGDP | 50 | 70.76 | 23.20 | 42 | 76.86 | 18.56 | 8 | 38.72 | 18.79 |
| | ROUrbanization | 50 | 64.38 | 20.75 | 42 | 70.19 | 16.45 | 8 | 33.84 | 12.65 |
| | ROCalories | 50 | 75.07 | 20.59 | 42 | 79.93 | 16.71 | 8 | 49.56 | 21.15 |
| South-East Asia | BMI ≥ 30 Female | 9 | 7.52 | 2.81 | 3 | 10.47 | 0.85 | 6 | 6.05 | 2.12 |
| | ROGDP | 9 | 31.40 | 15.80 | 3 | 47.46 | 12.36 | 6 | 23.37 | 10.29 |
| | ROUrbanization | 9 | 20.83 | 12.99 | 3 | 23.73 | 16.64 | 6 | 19.37 | 12.32 |
| | ROCalories | 9 | 32.11 | 15.47 | 3 | 38.76 | 10.80 | 6 | 28.78 | 17.22 |
| West Pacific | BMI ≥ 30 Female | 17 | 21.80 | 16.91 | 9 | 17.86 | 13.69 | 8 | 26.24 | 19.92 |
| | ROGDP | 17 | 50.70 | 28.01 | 9 | 69.69 | 25.27 | 8 | 29.35 | 8.90 |
| | ROUrbanization | 17 | 47.06 | 33.22 | 9 | 67.83 | 27.97 | 8 | 23.69 | 21.20 |
| | ROCalories | 17 | 49.76 | 21.50 | 9 | 62.66 | 15.37 | 8 | 35.25 | 18.21 |
| Worldwide | BMI ≥ 30 Female | 168 | 22.35 | 10.52 | 86 | 25.68 | 8.77 | 82 | 18.85 | 11.11 |
| | ROGDP | 168 | 50.39 | 28.58 | 86 | 71.17 | 19.00 | 82 | 28.60 | 19.13 |
| | ROUrbanization | 168 | 50.67 | 28.10 | 86 | 66.27 | 23.87 | 82 | 34.31 | 22.37 |
| | ROCalories | 168 | 50.87 | 28.74 | 86 | 69.21 | 21.96 | 82 | 31.63 | 21.57 |

**Type III Tests of Fixed Effects (Male)**

| Source | Numerator df | Denominator df | F | Sig. |
|---|---|---|---|---|
| Intercept | 1 | 156 | 368.08 | <0.001 |
| WHO Region | 5 | 156 | 11.59 | <0.001 |
| Ibs Med | 1 | 156 | 11.98 | <0.001 |
| WHORegion * Ib | 5 | 156 | 3.81 | <0.001 |

**Note:**

1. Dependent Variable: BMI≥30 Male in 2014.
2. RO refers to Relative Order that is the order of countries from the lowest value to the greatest as a percentage of all countries included into the variable.
3. Ibs Med: Cutoff ponit of 0.9658

**Type III Tests of Fixed Effects (Female)**

| Source | Numerator df | Denominator df | F | Sig. |
|---|---|---|---|---|
| Intercept | 1 | 156 | 483.55 | <0.001 |
| WHO Region | 5 | 156 | 12.18 | <0.001 |
| Ibs Med | 1 | 156 | 5.02 | <0.001 |
| WHORegion * Ib | 5 | 156 | 4.25 | <0.001 |

**Note:**

1. Dependent Variable: BMI≥30 Female in 2014.
2. RO refers to Relative Order that is the order of countries from the lowest value to the greatest as a percentage of all countries included into the variable.
3. Ibs Med: Cutoff ponit of 0.9658

Additional File 4:

Table AF4: Tests of normality of distributions of studied variables

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| BMI ≥ 30, Male | 0.138 | 168 | <0.001 | 0.933 | 168 | <0.001 |
| BMI ≥ 30, Female | 0.054 | 168 | 0.200* | 0.984 | 168 | 0.052 |
| Caloric intake | 0.045 | 168 | 0.200* | 0.981 | 168 | 0.023 |
| GDP | 0.249 | 168 | <0.001 | 0.683 | 168 | <0.001 |
| $I_{bs}$ | 0.252 | 168 | <0.001 | 0.767 | 168 | <0.001 |
| Urbanization | 0.067 | 168 | 0.066 | 0.972 | 168 | 0.002 |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Sex specific obesity prevalence is the percentage of a defined population segment with a body mass index (BMI) of no less than 30 kg/m$^2$.

Data sources: Total calories data from the FAO's FAOSTAT; BMI ≥30 data from the WHO Global Health Observatory; GDP data from the World Bank; Urbanization data from WHO. Biological State Index ($I_{bs}$) was self-calculated with country specific fertility data published by the United Nations and the mortality data published by World Health Organization (WHO).

Additional File 5:

**Text AF1: Calculation and significance of Biological State Index ($I_{bs}$)**

Natural selection is a key mechanism of evolution, it may produce the change in heritable traits of a population over successive generations [1]. In our modern society, natural selection still acts on all members of a population, selecting those individuals that have an increased reproductive success (survival and/or fertility) [2]. The "Biological State Index ($I_{bs}$)" has been proposed to measure the populational reproductive success by taking into account potential loss of reproductive success by mortality [3, 4].

The $I_{bs}$ is calculated by combining age-specific death frequency ($d_x$ variable of a life table) with an age-specific reproductive loss ($s_x$):

$$I_{bs} = 1 - \sum_{x=0}^{x=\omega} d_x . s_x$$

Where: $d_x$ is the frequency of deaths at age x, $s_x$ is the reproductive loss from dying at age x, i.e. the estimated probability of not possessing the complete number of births at age x. $s_x$ is based on the cumulative number of births at specific ages [4, 5]. The construction and interpretation of the $I_{bs}$ was predicated upon the assumption that heritability of human fertility variance is negligible [6]. An $I_{bs}$ value of one indicates total adaptation of the population to their environment (ability to overcome mortality selection pressures that are present). An $I_{bs}$ value of zero signifies a total lack of adaptation (inability to overcome selection pressures that are present), and an impossibility to give life to the next generation. An $I_{bs}$ value close to zero indicates large effective natural selection pressures acting on a population, since few individuals are surviving to produce offspring. In such a scenario there is a possibility for fast evolution, since many genes may not be passed to the next generation and little possibility for propagation of deleterious mutations. An $I_{bs}$ value close to one indicates that natural selection does not have much effect on the population since many individuals are able to maximally contribute to producing the next generation.


1. Hall, B.K.H., Benedikt, *Strickberger's Evolution (4th ed.).* . 2008: Sudbury, MA: Jones and Bartlett Publishers. ISBN 978-0-7637-0066-9. LCCN 2007008981. OCLC 85814089.
2. Byars, S.G., et al., *Colloquium papers: Natural selection in a contemporary human population.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107 Suppl 1**: p. 1787-92.
3. Henneberg, M. and J. Piontek, *Biological state index of human groups.* Przeglad Anthropologiczny, 1975. **XLI**: p. 191-201.
4. Henneberg, M., *Reproductive possibilities and estimations of the biological dynamics of earlier human populations.* Journal of Human Evolution, 1976. **5**: p. 41-8.
5. Henneberg, M., *Notes on the reproduction possibilities of human prehistorical populations.* Przeglad Anthropologiczny, 1975. **41**: p. 75-89.
6. Henneberg, M., *Quantitative evaluation of actual intensity of natural selection through differential fertility in human populations,.* American Journal of Phys Anthropology, 1985. **66**: p. 181.

## 2nd Response to Reviewers' comments

**PLOS ONE (PONE-D-17-11108) Review**

PONE-D-17-11108R1
Reduced natural selection may contribute to global obesity increase more in males than in females
PLOS ONE
Wenpeng You, Maciej Henneberg

**Abstract**

I felt your sentence on regression could use some rephrasing with regard to the doubling of Rˆ2. reads a little off.

Authors: The sentence was amended.

$R^2$ increment in multivariate regression due to adding $I_{bs}$ as a predictor of male obesity was twice more than the improvement in $R^2$ due to adding $I_{bs}$ as a predictor of female obesity.

**Introduction**

In your classifications of BMI, you mean 'healthy' not just *health*?

Authors: Since BMI status is described in nouns, "health" should remain in the sentence.

Paragraph on natural selection has many typos.

Authors: We checked and corrected the typos. Thanks

Whilst I do like the concise nature of your introduction, I feel the overall rationale regarding the importance of studying sex differences in the final paragraph lacking. Why is this important?

Authors: We inserted the following explanation to highlight the importance in studying the sex difference in obesity.

Due to obvious differences in body composition, fat distribution and hormonal regulation of metabolism, especially during pregnancy, lactation and post-partum periods, expression of different genes in males and females may be influencing energy balance of individuals.

**Methods**

You use the term 'confounders', this should be **confounding factors/variables* throughout

Authors: We changed the term "confounders" to "confounding factors/variables" throughout the manuscript.

Please rationalise the use of stepwise over standard enter method regression

Authors: The rationale for the use stepwise is:

Comparing to the Enter linear regression, stepwise model is advantageous in selecting and ordering independent variables, which have statistically significant influence on male and female obesity prevalence, from the one that has most influence on the dependent variable down to the one that has least influence. Meantime, stepwise model also removes those variables that have no statistically significant influence on male and female obesity prevalence.

In this study, stepwise model was performed to select and rank variables which had the statistically significant contribution to male and female obesity prevalence when $I_{bs}$ was excluded and included as the independent variable respectively. To make it clearer, we included those variables which were not the significant predictors in the table, but they were denoted as "Removed".

Information regarding how strongly non-normal distributions were before transformation may be useful in an appendix or supplementary information. I would guess this was checked using something like kolmogorov smirnov?

Authors: Supplementary file showing strength of non-normal distributions of all variables checked using kolmogorov smirnov was included as one of the additional files. This was also mentioned in Section of Methods in the manuscript.

**Table AF 4 Tests of normality of variables**

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| BMI ≥ 30, Male | 0.138 | 168 | <0.001 | 0.933 | 168 | <0.001 |
| BMI ≥ 30, Female | 0.054 | 168 | 0.200* | 0.984 | 168 | 0.052 |
| Calories | 0.045 | 168 | 0.200* | 0.981 | 168 | 0.023 |
| GDP | 0.249 | 168 | <0.001 | 0.683 | 168 | <0.001 |
| $I_{bs}$ | 0.252 | 168 | <0.001 | 0.767 | 168 | <0.001 |
| Urbanization | 0.067 | 168 | 0.066 | 0.972 | 168 | 0.002 |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Sex specific obesity prevalence is the percentage of defined population segment with a body mass index (BMI) of no less than 30 kg/m$^2$.

Data sources: Total calories data from the FAO's FAOSTAT; BMI ≥30 data from the WHO Global Health Observatory; GDP data from the World Bank; Urbanization data from WHO. Biological State Index ($I_{bs}$) was self-calculated with country specific fertility data published by the United Nations and the mortality data published by World Health Organization (WHO).

Due to abnormal distribution detected, the curvilinear correlation between $I_{bs}$ and male and female obesity prevalence standardised on the confounding factors was explored respectively. Details see the section of Data Analysis of the manuscript.

**Results**

It appears from your figures you may have some polynomial data. Was this explored?

Authors: Yes, we have realized this, and it has been explored and found that exponential relationship fits better our data than polynomial curve. See Figure 1. Most of our data are polynomial in the sense of having curvilinear relationships with many other data. This has been explored and variously controlled for in partial correlation analyses, mixed linear models and multiple regressions.

Table 1 suggests multicollinearity amongst the predictors. Diagnostic tests should be reported, and controls taken to ensure this does not influence the end model.

Authors: We have run diagnostic tests as shown below, and we have included tolerance figures in the manuscript. A tolerance of less than 0.20 or a VIF of above 5 indicates a multicollinearity problem (O'Brien 2007).

Table AF 2-1: Multicollinearity (diagnostic tests) of male and female obesity prevalence to other predictors

| Independent Variables | Tolerance | VIF | Independent Variables | Tolerance | VIF |
|---|---|---|---|---|---|
| Urbanization | 0.535 | 1.869 | Urbanization | 0.535 | 1.869 |
| $I_{bs}$ | 0.471 | 2.124 | $I_{bs}$ | 0.471 | 2.124 |
| Calories | 0.386 | 2.590 | Calories | 0.386 | 2.590 |
| GDP | 0.299 | 3.344 | GDP | 0.299 | 3.344 |

Table AF 2-2: Multicollinearity (diagnostic tests) of calories to other predictors in male and female samples respectively

| Independent Variables | Tolerance | VIF | Independent Variables | Tolerance | VIF |
|---|---|---|---|---|---|
| GDP | 0.304 | 3.290 | GDP | 0.358 | 2.793 |
| Urbanization | 0.542 | 1.846 | BMI ≥ 30, Female | 0.697 | 1.434 |
| BMI ≥ 30, Male | 0.324 | 3.084 | Urbanization | 0.558 | 1.791 |
| $I_{bs}$ | 0.424 | 2.360 | $I_{bs}$ | 0.474 | 2.108 |

Table AF 2-3: Multicollinearity (diagnostic tests) of GDP to other predictors in male and female samples respectively

| Independent Variables | Tolerance | VIF | Independent Variables | Tolerance | VIF |
|---|---|---|---|---|---|
| Urbanization | 0.553 | 1.809 | BMI ≥ 30, Female | 0.712 | 1.404 |
| BMI ≥ 30, Male | 0.351 | 2.848 | Urbanization | 0.580 | 1.723 |
| $I_{bs}$ | 0.453 | 2.207 | $I_{bs}$ | 0.548 | 1.826 |
| Calories | 0.425 | 2.353 | Calories | 0.472 | 2.119 |

Table AF 2-4: Multicollinearity (diagnostic tests) of $I_{bs}$ to other predictors in male and female samples respectively

| Independent Variables | Tolerance | VIF | Independent Variables | Tolerance | VIF |
|---|---|---|---|---|---|
| Calories | 0.376 | 2.659 | Calories | 0.395 | 2.529 |
| GDP | 0.288 | 3.476 | GDP | 0.346 | 2.888 |
| Urbanization | 0.521 | 1.918 | BMI ≥ 30, Female | 0.709 | 1.410 |
| BMI ≥ 30, Male | 0.351 | 2.848 | Urbanization | 0.530 | 1.886 |

Table AF 2-5: Multicollinearity (diagnostic tests) of urbanization to other predictors in male and female samples respectively

| Independent Variables | Tolerance | VIF | Independent Variables | Tolerance | VIF |
|---|---|---|---|---|---|
| $I_{bs}$ | 0.419 | 2.385 | $I_{bs}$ | 0.461 | 2.170 |
| BMI ≥ 30, Male | 0.321 | 3.116 | Calories | 0.405 | 2.472 |
| Calories | 0.387 | 2.586 | GDP | 0.319 | 3.134 |
| GDP | 0.282 | 3.542 | BMI ≥ 30, Female | 0.701 | 1.427 |

You use both enter and stepwise regressions. What justified the use of stepwise?

Authors: Stepwise analysis automatically selects and ranks the independent variables which contribute most to the variance of dependent variable- most important factors, whereas Enter produces coefficients for all the independent variables and this has been interpreted in the manuscript.

I still think these data may be nested within countries or regions and a linear mixed model of some sort might neatly summarise your data allowing for intercept change at the country level.

Authors: The linear Mixed Model Analysis was conducted to summarise the data quality after the data were nested within the WHO regions. After controlling for Calories, GDP and Urbanization, the male and female obesity prevalence rates were significantly different between WHO regions ($F=35.95$, $P<0.001$ & $F=12.18$, $P<0.001$ respectively). Further details see Additional File: Table AF 4.

**Discussion**

In your discussion of the possible causes of the male/female differences you use bullet points - Change these to normal paragraphs. I think overall this section should take precedence to the rest of the discussion as this formed the main rationale behind the study and I felt it was skimmed over too quickly. This disparity is an interesting one and certainly one I would really like to hear more about.

Authors: This is a good point. The important finding of this study that female body weight varies more has been highlighted after we moved forward the possible rationale of female obesity variation in different countries/regions.

There was not so much discussion around the longitudinal data in Australia in Fig2

Authors: The following paragraph has been included in the discussion as one of the study limitation.

Fourth, this study analysed the data across 191 countries. However, the results cannot be complemented by the longitudinal data analysis in the individual country due to the fact that obesity only has been an issue in the last three decades. We could not access the combined obesity and $I_{bs}$ prevalence data which are older than 30 years

**General Comments**

Overall I think this is a much improved manuscript on an interesting topic. I still think the main issues surround the analysis section of the paper and do suggest to the authors a multilevel model may be appropriate and help summarise the data in a neater fashion taking into account possible clustering of country/region. It might be worth trying to vary the intercept by these factors and seeing if this results in better model fit. The standard method of adding variables one by one and comparing model fit might provide the authors some nice statistics for comparing the influence of each predictor on model fit. Also, lot of the predictors are highly correlated with one another causing multicollinearity. Clarity on how this is tackled should be included.

Authors: In this paper, we only explored one predictor, which is $I_{bs}$ and obesity of males and females.

It is inappropriate for us in this paper to explore influence of other factors beyond controlling for them. The reviewer's comment is valuable as an idea for a new and different paper to explore influence of a number of factors on obesity. In the process of studying the issue of obesity, we have accumulated data on a large of number of variables such like, nutrients, food products, macronutrients and *etc*. it would be better to include all variables into such a new study rather than just a few in this paper. We would be happy to invite the reviewer as the co-author of this new study

I felt the other area of weakness was a lack of depth regarding the male/female sex differences. I see the authors make an attempt to summarise some possible factors, but a better discussion I feel is required as it does form the basis of the main rationale (and title of the study).

Authors: We included the multiple environmental factors in the discussion of the male/female sex differences. However, it seems that the interaction between genetic trait of obesity and environmental factors should have been highlighted although it is obvious that reduced natural selection may increase genetic background of obesity. The following paragraph has been inserted.

In other words, the same magnitude of metabolic faulty mutations accumulation due to the reduced natural selection in males and females does not lead to the same phenotypes at population level (different obesity prevalence rates in males and females). Multiple environmental factors that may influence the female obesity prevalence in different countries or regions may explain the disparity of obesity prevalence in males and females.

And later on in the text we have expanded on those possible non-genetic factors.

## Chapter 2: Nutrients/diets & global public health challenges

Article 4/10: Meat consumption providing a surplus energy in modern diet contributes to obesity prevalence: an ecological analysis (Published at BMC Nutrition, 2016)

Wen-Peng You[1], Maciej Henneberg[1,2,]

[1] Biological Anthropology and Comparative Anatomy Unit, School of Medical Sciences, the University of Adelaide, Adelaide, SA, Australia 5000.

[2] Institute of Evolutionary Medicine, University of Zürich, Switzerland.

**Published:** Wenpeng You, Maciej Henneberg. *Meat consumption providing a surplus energy in modern diet contributes to obesity prevalence: an ecological analysis*. BMC Nutrition, 20162:22. DOI: 10.1186/s40795-016-0063-9

✉ **Correspondence:** Wenpeng You Wenpeng.you@adelaide.edu.au

### Contextual Statement

Meat has been constantly associated with obesity. The prevailing explanation is that meat contains saturated fat. Misled by this dogma, lean meat has been included as the healthy food component in healthy dieting guidelines published by the authorities. A few previous studies have suggested that that it actually may be the main macronutrient of the meat, protein that contributes to obesity.

We postulated and tested that, in modern diet, meat protein, instead of meat fat, may be a direct obesity contributor as carbohydrates and fats may provide enough energy for human daily life needs before meat is digested.

## Statement of Authorship

# Statement of Authorship

| | |
|---|---|
| Title of Paper | Meat consumption providing a surplus energy in modern diet contributes to obesity prevalence: an ecological analysis. |
| Publication Status | ☑ Published ☐ Accepted for Publication<br>☐ Submitted for Publication ☐ Publication Style |
| Publication Details | Wenpeng You, Maciej Henneberg. Meat consumption providing a surplus energy in modern diet contributes to obesity prevalence: an ecological analysis. BMC Nutrition, 20162:22. DOI: 10.1186/s40795-016-0063-9 |

### Principal Author

| | | | |
|---|---|---|---|
| Name of Principal Author (Candidate) | Wenpeng You | | |
| Contribution to the Paper | Conceived the hypothesis, collected data, performed analysis on all samples, interpreted data, wrote manuscript and acted as corresponding author. | | |
| Overall percentage (%) | 60 | | |
| Signature | | Date | 22/12/2017 |

### Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

| | | | |
|---|---|---|---|
| Name of Co-Author | Maciej Henneberg | | |
| Contribution to the Paper | Supervised development of work, formalized and interpreted data analysis, and evaluated and edited manuscript | | |
| Signature | | Date | 22/12/2017 |

| | | | |
|---|---|---|---|
| Name of Co-Author | NA | | |
| Contribution to the Paper | | | |
| Signature | | Date | |

Please cut and paste additional co-author panels here as required.

## Abstract

**Background:** Excessive energy intake has been identified as a major contributor to the global obesity epidemic. However, it is not clear whether dietary patterns varying in their composition of food groups contribute. This study aims to determine whether differences in per capita availability of the major food groups could explain differences in global obesity prevalence.

**Methods:** Country-specific Body Mass Index (BMI) estimates (mean, prevalence of obesity and overweight) were obtained. BMI estimates were then matched to mean of three year-and country-specific availability of total kilocalories per capita per day, major food groups (meat, starch, fibers, fats and fruits). The per capita Gross Domestic Product (GDP) and prevalence of physical inactivity for each country were also obtained. SPSS was used for log-transformed data analysis.

**Results:** Spearman analyses of the different major food groups shows that meat availability is most highly correlated with prevalence of obesity ($r=0.666$, $p<0.001$) and overweight ($r=0.800$, $p<0.001$) and mean BMI ($r=0.656$, $p<0.001$) and that these relationships remain when total caloric availability, prevalence of physical inactivity and GDP are controlled in partial correlation analysis. Stepwise multiple linear regression analysis indicates that meat availability is the most significant predictors of prevalence of obesity and overweight and mean BMI among the food groups. Scatter plot diagrams show meat and GDP adjusted meat are strongly correlated to obesity prevalence.

**Conclusion:** High meat availability is correlated to increased prevalence of obesity. Effective strategies to reduce meat consumption may have differential effects in countries at different stages of the nutrition transition.

**Keywords**: Obesity, Food group, Meat, Macronutrient, Meat protein, Carbohydrates, Adaptation

## Background

The global prevalence of obesity and its associated metabolic syndrome has increased markedly in adults and children over the past 20 years [1-6]. Once considered a problem only in high income countries, obesity is now dramatically on the rise in low- and middle-income countries, particularly in urban settings. Obesity has been considered as one of major risk factors for a number of chronic diseases, including diabetes, cardiovascular diseases and cancer [7]. The World Health Organization (WHO) describes obesity as one of the most blatantly visible, yet most neglected, public health problems [8].

Body weight status is determined with reference to the body mass index (BMI). Those with a BMI ranging between 18-24.99 kg/m$^2$ are considered healthy. In WHO statistics, population segment consisting of individuals with a BMI equal to 25 kg/m$^2$ or higher is classified as overweight whilst obesity is reserved for those reaching or exceeding a BMI of 30 kg/m$^2$ [9]. WHO also publishes the country-level estimate of mean BMI in kg/m$^2$ to reflect its general body weight status.

It is well recognised that diet and lifestyle are the major contributing factors, yet previous population based dietary interventions that focus on one dietary factor such as reducing fat intake have been ineffective in combating the increasing rates of obesity [10-12]. Although energy intake is recognised as a major contributing factor to the growing obesity rates, there is increasing evidence that some dietary patterns have a greater influence on promoting body weight gain than others [13]. Food production modernization and rising income levels in last decades have made a range of foods easily available and affordable with less seasonal variation [14].To combat obesity a common approach has been to limit energy intake, although weight loss is often achieved in the short term, studies are unable to show that this weight loss is maintained in the long term [15]. Of the food groups, meat when consumed at high levels has been shown to increase weight gain due to its high energy density and/or fat content [16-20]. Whether and how nutrients provided by other food groups contribute to this effect is not known. In addition, there is little evidence that diet containing different composition of food groups or macronutrients may also be important in determining the development of obesity, yet this has yet to be evaluated at the population level.

Our group recently suggested that the portion size of animal and plant products in the modern diet has contributed to obesity prevalence [21]. People from different countries have different availability of meat due to their affordability and dietary habits. We hypothesise that the persistent consumption of high quantities of meat contributes to increasing adiposity and thus obesity when carbohydrates and fats consumed are sufficient or overabundant to satisfy caloric needs. Here we test this hypothesis using three country specific variables defined by BMI values (prevalence of obesity and overweight and mean BMI) and per capita availability data of various major groups of foodstuffs (meat, starch crop, fruits, fats and fibers) and the three macronutrients (fats, proteins and carbohydrates).

## Methods

The country specific data were collected for this ecological study:

**The WHO Global Health Observatory (GHO) data**

The WHO Global Health Observatory (GHO) data on estimated prevalence rates of obesity and overweight (percent of population aged 18+ with BMI ≥ 30 and 25 kg/m$^2$ respectively) and on mean BMI of the population aged 18+ by country was obtained for the year 2010 [22]. We did not use the most recent version of three levels of BMI (BMI=30, BMI=25 and mean BMI) in 2014, but used the 2010 year data because of other key variables of interest (described below). We included overweight prevalence and mean BMI in our study in case meat availability was a late-stage predictor of obesity.

We also captured the estimated prevalence rate of physical inactivity for each country for the population aged 18+ [22]. The estimated prevalence rate of physical inactivity is defined as percent of defined population attaining less than 150 minutes of moderate-intensity physical activity per week, or less than 75 minutes of vigorous-intensity physical activity per week, or equivalent.

The GHO is an initiative of the WHO to share data on global health, including statistics by country and information about specific diseases and health measures. The GHO specifically assembles prevalence data of the biological risk factors, including obesity, overweight and mean BMI for WHO Member States using standardized protocols (http://www.who.int/gho/ncd/methods/en/).

**The FAOSTAT Food Balance Sheet (FBS) data**

The FAOSTAT Food Balance Sheet (FBS) data on major food group availability per capita per day of: i) total meat; ii) starch crops (mixed cereals and starchy root); iii) fibers (vegetables and pulses); iv) fats (plant oils and animal fats) and v) fruits [23]. The food items in each food group are indicated in the Supporting Information (Table S1).

We also extracted the availability of grand total calories and macro-nutrients of fats (animal and plant, in g/capita/day) and proteins (animal, plant and meat, in g/capita/day) from FBS for our study. As animal protein includes meat protein, we subtracted meat protein from the animal protein to obtain the variable, “Animal protein, excluding meat protein” for more precise data analysis. Following the Atwater system [24], we calculated the energy from carbohydrates using the formula: carbohydrates energy per day = total calories- fat (grand total, in gram/day) x 9 – protein (total, in gram/day) x 4. For carbohydrates availability in g/capita/day, we used the energy in kilocalories (kcal) divided by 4. Because obesity develops after cumulative exposure to dietary risks (i.e. high intake of risk food groups today does not lead to immediate obesity, but a prolonged exposure to high intake of risk food type(s) is required.), we calculated the mean grams

per person per day over a 3-year period (2007-2009) in each of these food categories to represent typical long-term exposure to each of these dietary components. The rationale for this decision is that studies have shown that three years is a practical period to develop metabolic syndrome leading to obesity after exposure to dietary risks (i.e. high intake of meat today does not lead to immediate obesity) [25-27]. Using the mean of three years of nutrients and food groups may also reduce the random errors during the data collection and calculation by FAO.

The FAOSTAT database disseminates statistical data collected and maintained by the FAO. FAOSTAT data are provided as a time-series from 1961 in most domains through the Food Balance Sheet (FBS, http://faostat3.fao.org/home/E). The FBS presents a comprehensive picture of the pattern of a country's food supply during a specified reference period. The FBS shows for each food item i.e. each primary commodity availability for human consumption which corresponds to the sources of supply and its utilisation. The total quantity of foodstuffs produced in a country added to the total quantity imported and adjusted to any change in stocks that may have occurred since the beginning of the reference period gives the supply available during that period. On the utilisation side a distinction is made between the quantities exported, fed to livestock + used for seed, losses during storage and transportation, and food supplies available for human consumption. The per capita supply of each such food item available for human consumption is then obtained by dividing the respective quantity by the related data on the population actually partaking in it [28].

Minimum Dietary Energy Requirements, expressed as kcal per person per day, is the weighted average of the minimum energy requirements of the different gender-age groups in the population with light activity. Grantham *et al.* reported that when a mixed meal of protein, carbohydrate and fat is consumed, carbohydrates and fats are digested faster and metabolised to satisfy body's energetic needs while slower digested protein is ultimately and stored as fat [29]. Therefore, we extracted the Minimum Dietary Energy Requirements from the FAO website (http://www.fao.org/) and compared it and with the energy from carbohydrates and fats by country to see if the energy from the proteins is the surplus.

**The World Bank data**

The World Bank dataset measures progress on aggregate outcomes for member countries for selected indicators. GDP PPP is gross domestic product converted to international dollars using purchasing power parity rates (http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD) [30]. GDP PPP is the measure of

average income in constant 2010 $US adjusted for purchasing power parity for cross-country comparability.

WHO, FAO and the World Bank are intergovernmental organizations using specialized information relevant to their respective fields. Their professional personnel should have evaluated these data in consideration of their possible use, e.g. for scientific research and decision making, before they were published. Therefore, the data reporting is as free of bias and error as it can be with government statistics. This means that errors are reduced but some inaccuracies related to reporting quality may still be present in the data. Similar data from the same sources were recently used to analyse the relationships between nutrients and obesity [31, 32] and diabetes [33-35] in a number of publications.

We obtained data for 170 countries after we matched the prevalence estimates of obesity and overweight and mean BMI to the year-and country-specific food and other variables. Each country was treated individually as the subject and all their availability for other variables information was analysed. The detailed information of country-level estimates is in the Supporting Information (Table S2).

For particular analyses, the number of countries included may have differed somewhat because all information on other variables was not uniformly available for all countries due to unavailability from relevant UN agencies. All the data were extracted and saved in Microsoft Excel® for analysis. Data sources and summary statistics are further described in the Supporting Information (Table S3).

**Statistical analysis**

The prevailing dogma of obesity is that obesity is an affluence related medical conditions [36], which is generally caused by eating too much (too much calories intake) [37] and moving too little (physically inactive) [38]. Therefore, in this study we used GDP PPP, total calories and prevalence of physical inactivity as the potential confounders and the other variables are divided into two sets, i.e. major food group and macronutrient for data analysis in 5 steps.

Spearman rank correlation analyses was used to evaluate the strength and direction of the associations between food group and macronutrient availability for consumption and prevalence estimates of overweight and obesity and mean BMI.

Partial correlation was used to find the unique variance between each food group and macronutrient and prevalence of obesity and overweight and mean BMI respectively while eliminating the variance from total calories, GDP PPP and physical inactivity. In

order to show the independent correlation of meat and meat protein to the three variables defined by BMI (BMI≥30, BMI≥25 and mean BMI) respectively, we controlled for three potential confounders (total calories, GDP PPP and physical inactivity) plus all other food groups and all other macronutrient variables respectively for partial analysis.

Stepwise multiple linear regression modelling was performed to identify and rank predictors (independent variables) of prevalence of obesity, overweight and mean BMI respectively from two sets of data of food groups and macronutrients respectively.

Scatter plots were used to explore the relationship between meat and meat protein (both GDP adjusted) and three variables defined by BMI. Scatter plots were also used to explore the relationship between prevalence of obesity and each food group and macronutrient respectively.

Human diet patterns varying in different food components may be affected by the types of food availability in a particular region, socio-economic status and cultural beliefs. In order to demonstrate that correlation universally exists between meat availability and obesity regardless of these factors, countries were grouped for correlation analyses. The criteria for grouping countries the World Bank income classifications [39], WHO regions [40], countries sharing specific characteristics like geography, culture, development role or socio-economic status, like Latin America and the Caribbean (LAC) [41], Organisation for Economic Co-operation and Development (OECD) [42], Asia-Pacific Economic Cooperation (APEC) [42], Southern African Development Community (SADC) [43], the Arab World [42], Latin America (LA), and Asia Cooperation Dialogue (ACD) [44]. All the country listings are sourced from their official websites for matching except LA which is self-classified based on region primarily speaking romance languages. Countries included in LA are listed in the Supporting Information (Table S4).

SPSS v. 22 (SPSS Inc., Chicago Il USA) was used for data analysis and the statistical significance was set at the 0.01 level (two-tailed). Prior to analysis data were log-transformed to bring their distributions close to normal.

## Results

Spearman rank correlation analyses of the different major food groups shows that meat availability is most highly correlated with prevalence of obesity ($r=0.666$, $p< 0.001$) and overweight ($r= 0.800$, $p< 0.001$) and mean BMI ($r= 0.656$, $p< 0.001$) and that these relationships remain when total caloric availability, prevalence of physical inactivity and GDP PPP are kept statistically constant in partial correlation analysis (Table 1). Starch crop availability is strongly in negative correlation with prevalence of obesity ($r=-0.205$,

p< 0.01) and overweight (r=-0.228, p< 0.01) and mean BMI (r= -0.318, p< 0.001), but the relationship does not remain in our partial correlation analysis (Table 1). Interestingly, in Spearman rank correlation analyses fats group is second to meat in significant correlation with prevalence of obesity (r=0.517, p< 0.001) and overweight (r= 0.728, p< 0.001) and mean BMI (r= 0.438, p< 0.001). However, these relationships nearly disappear in the succeeding partial correlation analysis with controlling for total caloric availability, prevalence of physical inactivity and GDP (Table 1).

Table 1 also presents the strongest significant correlation between meat protein availability and prevalence of obesity (r=0.673, p< 0.001) and overweight (r= 0.793, p< 0.001) and mean BMI (r=0.660, p<0.001). This correlation is sustained when total caloric availability, prevalence of physical inactivity and GDP PPP are kept statistically constant in partial correlation analysis (Table 1). Animal protein (excluding meat protein) shows quite high nonparametric correlation coefficients with prevalence of obesity (r=0.522, p< 0.001) and overweight (r= 0.741, p< 0.001) and mean BMI (r=0.516, p<0.001), but this correlation is not sustained in succeeding partial analysis (Table 1). Plant protein group shows slightly negative correlation with all the three stages of body weight (BMI ≥ 30, BMI≥25 and mean BMI) in Spearman rank correlation analyse, but the relationships are relative strong (not at significance level of p<0.001 yet) in partial correlation analysis with controlling for total caloric availability, prevalence of physical inactivity and GDP (Table 1). Both animal fat and plant oil food types are correlated with prevalence of obesity (r=0.581, p<0.001 and r=0.440, p<0.001respectively) and overweight (r=0.803, p<0.001 and r=0.570, p<0.001 respectively) and mean BMI (r=0.574, p<0.001 and r=0.371, p< 0.001respectively) in Spearman rank correlation analyses. However, in the succeeding partial correlation analysis the significance either does not remain or becomes weak except the correlation between animal fats group and prevalence overweight (r=0.358, p<0.001). Carbohydrates energy shows the relative significant correlation with prevalence of obesity (r=0.230, p<0.01) and overweight (r=0.202, p<0.01) and mean BMI (r=0.208, p<0.01), but this relationship becomes slightly negative in partial correlation analysis (Table 1).

Meat and meat protein are in significant correlation with prevalence of obesity (r=0.356, p<0.001 and r=0.392, p<0.001 respectively) and overweight (r=0.421, p<0.001 and r=0.431, p<0.001 respectively) and mean BMI (r=0.380, p<0.001 and r=0.400, p<0.001 respectively) when we control for the potential confounders, total calories, GDP and physical inactivity in partial analysis (Table 1). Meat availability is also significantly correlated to prevalence of obesity (r=0.357, p<0.001) and overweight (r=0.415, p<0.001)

and mean BMI (r=0.339, p<0.001) when we controlled for the four other food groups and the three potential confounders in partial correlation. We have the similar correlation of meat protein to three variables defined by BMI respectively when we controlled for the other five macronutrients and the three potential confounders (Table 1).

Table 1 Spearman and partial correlation between food groups and three variables defined by BMI (obesity, overweight and mean BMI)

| | Spearman | | | Partial | | |
|---|---|---|---|---|---|---|
| Variables | BMI≥30 | BMI≥25 | BMI mean | BMI≥30 | BMI≥25 | BMI mean |
| **Food group** | | | | | | |
| Meat, total | 0.666*** | 0.800*** | 0.656*** | 0.356*** | 0.421*** | 0.380*** |
| Meat, total, all variable controlled+ | - | - | - | 0.357*** | 0.415*** | 0.339*** |
| Fats (plant oil + animal fat) | 0.517*** | 0.728*** | 0.483*** | 0.077 | 0.166 | -0.005 |
| Fruits, total | 0.467*** | 0.521*** | 0.461*** | 0.173 | 0.197* | 0.258** |
| Fibers (vegetables + pulses) | 0.315*** | 0.516*** | 0.330*** | -0.197* | -0.035 | -0.107 |
| Starch (cereals + starchy root) | -0.205** | -0.228** | -.318*** | 0.078 | -0.011 | -0.085 |
| **Macronutrient** | | | | | | |
| Meat protein | 0.673*** | 0.793*** | 0.660*** | 0.392*** | 0.431*** | 0.400*** |
| Meat protein, all variable controlled++ | - | - | - | 0.316*** | 0.183* | 0.299*** |
| Animal protein, excluding meat protein | 0.522*** | 0.741*** | 0.516*** | 0.017 | 0.214* | 0.029 |
| Plant protein, total | -0.094 | -0.063 | -0.094 | -0.227* | -0.333*** | -0.248* |
| Animal fats, total | 0.581*** | 0.803*** | 0.574*** | 0.196* | 0.379*** | 0.222* |
| Plant fats, total | 0.440*** | 0.570*** | 0.371*** | 0.252* | 0.230** | 0.201* |
| Carbohydrates | 0.230** | 0.202** | 0.208** | -0.193* | -0.324*** | -0.166 |
| **Potential confounder** | | | | | | |
| Calories, total | 0.623*** | 0.805*** | 0.563*** | - | - | - |
| GDP PPP | 0.642*** | 0.808*** | 0.610*** | - | - | - |
| Physical Inactivity | 0.438*** | 0.384*** | 0.460*** | - | - | - |

Spearman's rho of correlation and partial correlation are reported. Numbers of countries (df) included in the two correlation analyses are 161-170 and 115-123 respectively. * P< 0.05, **P< 0.01; ***P< 0.001.

BMI≥30 and BMI ≥ 25 are percentages of defined population with a body mass index (BMI) of no less than 30 kg/m2 and 25 kg/m$^2$ respectively. BMI mean is the mean body mass index (BMI) in kg/m$^2$ of defined population.

Availabilities of food types (meat, fats, fruits, fibers and starch) and macronutrients (meat protein, animal protein (excl. meat protein), plant protein, animal fats, plant fats and carbohydrates) are expressed in g/capita/day.

Total calories is in kcal/capita/day. GDP PPP is in per capita USD per year. Physical inactivity is defined as the percent of defined population attaining less than 150 minutes of moderate-intensity physical activity per week, or less than 75 minutes of vigorous-intensity physical activity per week, or equivalent.

+ Partial analysis with controlling for fats (plant oil + animal fat), Fruits, total (total), Fibers (vegetables + pulses) and Starch (cereals + starchy root) and the three potential confounders, calories, GDP PPP and physical activity.

++ Partial analysis with controlling for Animal protein (excluding meat protein), Plant protein (total), Animal fats (total), Plant fats (total) and Carbohydrate energy and the three potential confounders, calories, GDP PPP and physical activity.

Table 2 presents that meat and meat protein availability are the most significant predictors of prevalence of obesity (R2=0.468 and R2=0.472 respectively) and overweight (R2=0.628 and R2=0.614 respectively) and mean BMI (R2=0.507 and R2=0.498 respectively) when all food groups and macronutrients were entered into the regression model respectively for stepwise multiple linear regression analysis.

Table 2 Results of stepwise multiple linear regression analyses to identify food group and macronutrient predictors of three variables defined by BMI

| | BMI ≥ 30 | | BMI ≥ 25 | | BMI, Mean | |
|---|---|---|---|---|---|---|
| Rank | Variables entered | Adjusted $R^2$ | Variables entered | Adjusted $R^2$ | Variables entered | Adjusted $R^2$ |
| **Food groups** | | | | | | |
| 1 | Meat, total | 0.468 | Meat, total | 0.628 | Meat, total | 0.507 |
| 2 | Fruits | 0.483 | Fibers (vegetables + Pulses) | 0.667 | Fruits | 0.538 |
| 3 | Fats (animal fat + plant oil) | 0.494 | Fats (animal fat + plant oil) | 0.687 | - | - |
| 4 | - | - | Fruits | 0.701 | - | - |
| **Macronutrients** | | | | | | |
| 1 | Meat protein | 0.472 | Meat protein | 0.614 | Meat protein | 0.498 |
| 2 | Plant oil | 0.522 | Animal protein, excl. meat protein | 0.666 | Plant oil | 0.526 |
| 3 | Carbohydrates | 0.549 | Plant oil | 0.694 | Carbohydrates | 0.548 |
| 4 | - | - | Carbohydrates | 0.714 | - | - |

Stepwise multiple linear regression modelling is reported. Number of countries included in the analysis range from 157 to 166.

BMI≥30 and BMI ≥ 25 are percentages of defined population with a body mass index (BMI) of no less than 30 kg/m2 and 25 kg/m2 respectively. BMI mean is the mean body mass index (BMI) in kg/m2 of defined population.

Availabilities of food types (meat, fats, fruits, fibers and starch) and macronutrients (meat protein, animal protein (excl. meat protein), plant protein, animal fats, plant fats and carbohydrates) are expressed in g/capita/day.

The relationship between GDP adjusted meat availability and prevalence of obesity and overweight and mean BMI is noted to be logarithmic with strong correlations (Figure 1). Meanwhile relationship between GDP adjusted meat protein and the three levels of BMIs shows polynomial relationship with the three variables describing weight status (Figure 2).

Figure 1. Relationships between meat availability adjusted for GDP and prevalence of obesity and overweight and mean BMI by country

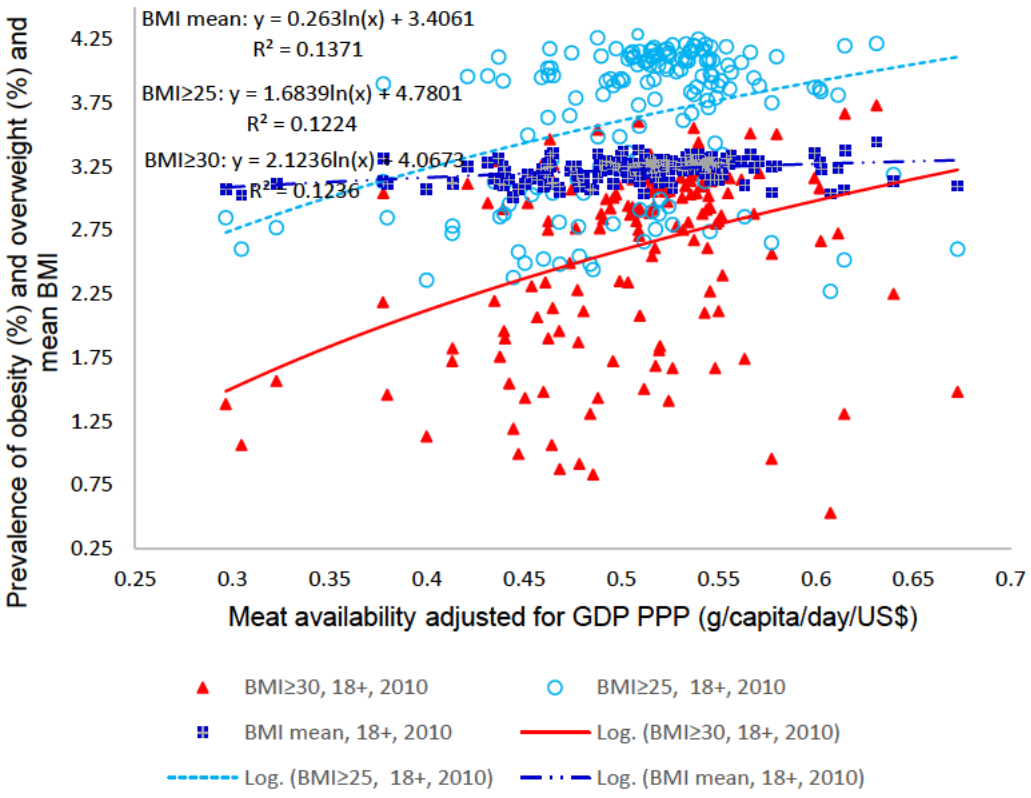
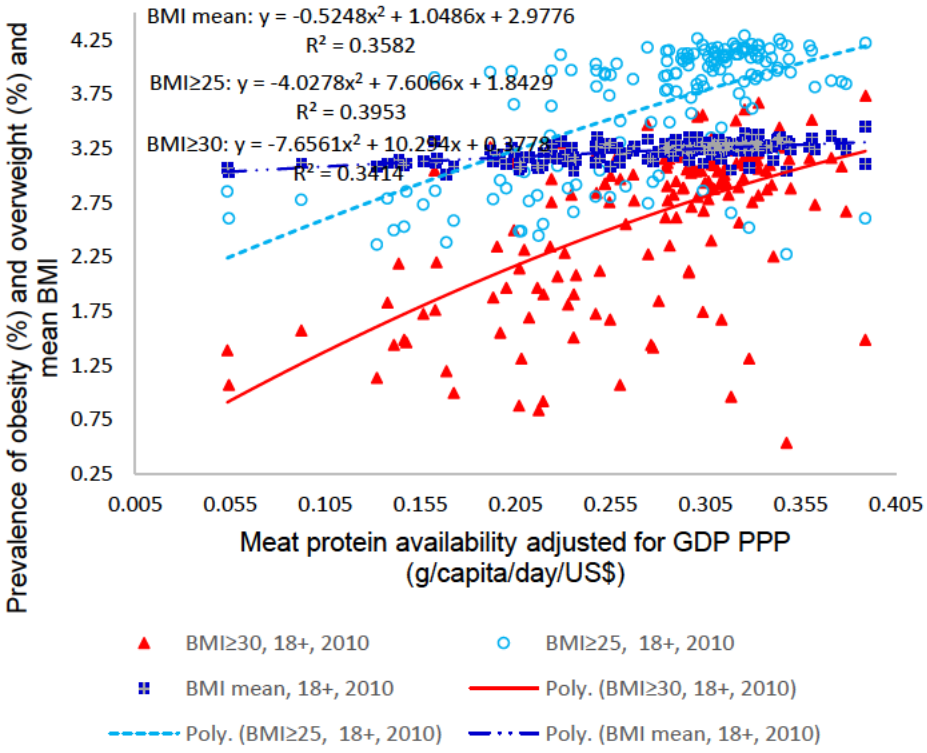



Figure 2. Relationships between meat protein availability adjusted for GDP and prevalence of obesity and overweight and mean BMI by country



We also used scatter plots to show the relationship between prevalence of obesity and each food group and macronutrient. See the Supporting Information (Figures S 1 and 2).

Table 3 shows that generally meat availability is positively correlated with prevalence of obesity and overweight and mean BMI can be observed in different country groupings regardless of cultural backgrounds, economic levels and geographic locations of the clustered countries.

Table 3 Correlation of meat availability to prevalence of obesity and overweight and mean BMI in different country groupings

| Country groupings | BMI ≥30 | BMI ≥25 | BMI, Mean |
|---|---|---|---|
| Worldwide (n=167) | 0.666*** | 0.800*** | 0.656*** |
| World Bank income classifications | | | |
| Low (n=31) | 0.167 | 0.254 | 0.196 |
| Low middle (n=41) | 0.439** | 0.537*** | 0.465** |
| Upper middle (n=47) | 0.167 | 0.149 | 0.209 |
| High (n=48) | 0.241 | 0.631*** | 0.288* |
| WHO regions | | | |
| AFRO (n=40) | 0.585*** | 0.612*** | 0.552*** |
| AMRO (n=35) | 0.671*** | 0.606*** | 0.546*** |
| EMRO (n=15) | 0.857*** | 0.879*** | 0.634* |
| EURO (n=50) | 0.429** | 0.751*** | 0.128 |
| SEARO (n=10) | -0.267 | -0.097 | 0.322 |
| WPRO (n=17) | 0.309 | 0.478 | 0.447 |
| Countries grouped based on various factors | | | |
| APEC (n=17) | 0.773*** | 0.858*** | 0.789*** |
| Arab World (n=13) | 0.687** | 0.687** | 0.426 |
| LAC (n=26) | 0.609*** | 0.519** | 0.487** |
| OECD (n=34) | 0.243 | 0.607*** | 0.285 |
| SADC (n=14) | 0.890*** | 0.952*** | 0.802*** |
| ACD (n=26) | 0.593*** | 0.720*** | 0.707*** |
| LA (n=20) | 0.557* | 0.675*** | 0.433 |

Spearman's rho of correlation is reported. Number of countries included in the analysis range from 161 to 170.
* P< 0.05, **P< 0.01; ***P< 0.001

BMI≥30 and BMI ≥ 25 are percentages of defined population with a body mass index (BMI) of no less than 30 kg/m2 and 25 kg/m2 respectively. BMI mean is the mean body mass index (BMI) in kg/m2 of defined population.

Availabilities of food types (meat, fats, fruits, fibers and starch) and macronutrients (meat protein, animal protein (excl. meat protein), plant protein, animal fats, plant fats and carbohydrates) are expressed in g/capita/day.

Based on the WHO region classifications, the positive correlation is observed in every region except in SEARO.

The correlation between meat availability and three variables defined by BMI can also be observed in the country groupings of the Arab World (geographically scattered in Asia and Africa) and LAC (located in Americas only) featured with the similar cultures respectively. The trends also present in two functional alliances, OECD and APEC although the former comprises developed countries only and the latter is comprised of both developing and developed countries.

We subtracted grand total protein energy from grand total calories to allow us to obtain the energy from grand total fats and carbohydrates in kcal/capita/day [28], which is more than the minimum dietary energy requirements in all countries except Haiti (-29.3 kcal/capita/day) and Zambia (-90.9 kcal/capita/day).

## Discussion

The worldwide secular trend of increased obesity prevalence likely has multiple aetiologies, which may act through multiple mechanisms. By examining the per capita availability of the major food groups and macronutrients for 170 countries we have shown that populations with the highest availability levels of meat (meat protein) have the highest prevalence of overweight and obesity and greatest mean BMI. Meat is most significant predictor of prevalence of obesity and overweight and mean BMI at country level, and this relationship is independent of total calories availability, GDP and prevalence of physical inactivity. Our finding of the relationship between meat availability and body weight increase is consistent with data from Belgium [45] and USA [46-48] that showed a positive association between obesity prevalence among adults and children and meat consumption. Studies in China also showed that high intakes of meat products, including red meat were associated with the prevalence of obesity [49, 50]. A survey in Ireland showed that young girls avoided meat because they concluded that "'meat is a fattening food" [51]. The association for the Chinese population is particularly striking as the changes in dietary patterns and obesity rates have occurred very rapidly [52]. All these studies based on the individual level held the view that fat in meat contributed to obesity or body weight increase even though fresh meat has been leaner than ever over the past few decades due to leaner animals being bred and improved butchery and feeding techniques that make fat content fall significantly [53, 54]. The correlation we found in this study between the three major macronutrients or their proxy food groups and three variables defined by BMI is compatible with Grantham *et al.*'s finding that, in

modern diet, carbohydrates and fats are digested to satisfy body's energetic needs while protein is converted and stored as fat [29].

The human metabolic system has been adapting to forager diet for millions of years [56], and adaptations to an agriculture-based diet only started a few thousand years ago in most populations [29, 57]. An evolutionary mismatch between modern dietary constituents and the food available prior to the agricultural revolution has long been considered a factor in the obesity epidemic [58]. In the Palaeolithic age our ancestors' diet comprised of what could be extracted from natural environments through gathering, scavenging and hunting and thus predominately consisted of animal protein [59]. In addition to hunting large animals, the main food sources included smaller animals such as amphibians, reptiles, invertebrates and their eggs, but also plant products, such as tubers, fruits and nuts that could be collected seasonally. In general, there was limited availability of animal and plant food, but plant sources were often least available [21]. Fats do not occur in large quantities in plants or wild animals. In the foraging situation ingested protein was mainly used for energy production as available carbohydrates from plants would be too scarce to satisfy human energy needs [56]. This use of protein was possible as humans have efficient deaminases that can convert amino acids to carbon skeletons that, when broken down to pyruvate can be processed in the citric acid cycle, or de novo lipogenesis, or gluconeogenesis [21]. Occasionally, when there was an abundant meat source, e.g. a large mammal, surplus ingested protein was efficiently stored in the human body as adipose tissue [60]. Thus, the human metabolic system has evolved over thousands of years to predominately rely on animal protein and to a lesser degree carbohydrate and fats to satisfy our energy needs and to store surplus food intake into the adipose tissue [21]. Further support of human adaptation and dependence on protein for energy, comes from similarities in total energy intake (standardised by body mass) and intestinal tract morphology between modern humans and extant carnivores [21]**.**

In the current study animal products provided less than half (3.1% - 44.5%) of the individual daily energy requirement for all countries examined [23], and a majority of energy came from plant products. Interestingly, there are a number of different weight loss diets that are high in animal and low in plant products such as the Atkins Nutritional Approach [61-63]. Although these diets can be effective in reducing weight in the short term, energy restriction is difficult to maintain long term and a majority of people regain any weight that was lost [15]. Daily energy requirements of modern humans may be quickly and easily satisfied by digesting plant products rich in carbohydrates [21, 29, 50]

whereas consumed concurrently animal products, including meat that are more costly and slower to digest, will be metabolised into fat and stored [21]. The FAO/WHO currently recommends that our dietary protein should make up 10-15 percent of calorie intake [64]. It has been reported that consuming an amount of protein above the FAO/WHO recommendation may be deleterious for weight maintenance through adult life [65]. In support of this, the PANACEA project which used data from the EPIC cohort [66] showed that participants consuming more than 22% of energy from protein had 23-24% higher risk of becoming overweight or obese than participants consuming a diet low in protein (≤14%) [66]. Additionally, a 5% higher proportion of protein at the expense of carbohydrates was associated with a 247g weight gain in men (95% CI = (160,334)) and a 388g weight gain (296,480) in women after 5 years [66]. Furthermore, increasing the proportion of fat by 5% at the expense of carbohydrates during the same period showed no association with body weight increase [66]**.**

Experiments among young males and rats undertaken by Mikkelsen *et al.* [67] and Toden *et al.* [68] respectively did not show the high meat protein quantity was associated with body weight increase. The underlying reasons may be that the used diets contained too much meat protein which was over FAO/WHO recommended level and/or that these experiments focused on one or two sources of proteins, which did not reflect the actual protein metabolism within human body. Two case-controlled studies have shown that adults and children consuming vegetarian diets have lower BMI values and a lower prevalence of obesity [69, 70]. A medical and performance testing of 46,684 Swiss showed that obesity rates were also markedly lower in vegetarian adults [29] and epidemiological studies have consistently shown that vegetarians are thinner than comparable non-vegetarians [71]. A meta-analysis of adult vegetarian diet studies estimated a reduced weight difference of 7.6 kg for men and 3.3 kg for women, which resulted in a 2-point lower BMI [69]. Although there are some animal data suggesting that diets low in protein may increase the prevalence of obesity [72], evolutionary differences between humans and other animal species may explain our different metabolic response to dietary protein [73]. Rats [74] and mice [75] model experiments have shown that dairy protein rich diet reduces adiposity, which might be interpreted that the associations between dairy protein and overweight and obesity are not as strong as meat protein in this study. Our results show animal protein (excluding meat protein) is associated with the three stages of BMIs, but not as significantly as meat protein does may be because protein from dairy [74] and fish products [76] don't contribute to body weight increase.

There is a growing body of evidence which suggests that increased plant protein intakes are protective of body weight gain. A longitudinal association study in the US showed that people with the highest levels of plant protein intake had a reduced risk of being obese [48]**.** A similar association was found in the Belgian population using a food consumption survey [45]. These findings are consistent with the current study which showed that plant protein consumption rates were inversely associated with prevalence of both overweight and obesity [50] and mean BMI**.** Plant and meat protein may have different effects on body weight [48] because of their differences in amino acid composition [77]. Generally, dietary plant protein in food is mixed with indigestible carbohydrate (fiber) that can reduce plant protein digestibility. Therefore, plant protein varies in its digestibility and may provide considerably less energy compared to meat proteins.

The current study shows an inverse association between starch food group (mixed cereals and starchy root) and carbohydrates availability and prevalence of overweight and obesity and mean BMI. Cereals and starchy roots are grown in greater quantities and provide more food energy worldwide than any other type of crop. Carbohydrates are not an essential nutrient in humans [78, 79] even though they are a common source of energy. For instance, carbohydrate content in foods provide 70 percent or more of the energy intake of the population in the developing countries and about 40 percent in the United States and Europe [80]. Humans are the only large mammal that derives a majority of its energy by absorbing and metabolising carbohydrate. Because carbohydrate metabolism primarily concentrates on the oxidation of carbohydrates in the direct production of energy, this rarely produces fat [78, 81].

Our results show that both plant oils and animal fats are significantly associated with mean BMI, overweight and obesity in Spearman analysis, but the significance of this relationship disappears or is reduced because we controlled for total calories, GDP and prevalence of physical inactivity in partial correlation analysis. Numerous studies have shown increased intakes of dietary fat increase obesity risk/development [82-86]. However, a causal relationship between fat intake and obesity prevalence based on these studies [87-89] is difficult to demonstrate. Furthermore, the third American National Health and Nutrition Examination Survey showed that in the past two decades in United States, the prevalence of obesity has increased whereas the fat consumption was reduced [90, 91]. Therefore, the increase in obesity cannot be explained by changes in dietary fat alone.

A strength of this study is that we used per capita availability data from 170 countries which enabled us to examine relationships in food group and macronutrient intake and how they may explain differences in the rates of prevalence of obesity and overweight and mean BMI at population level. However, there are several limitations in this study. Firstly, although we attempted to remove confounding effects of variables such as GDP, caloric *etc.* by means of partial correlation analysis, some confounding factors may still influence correlation we found. Secondly, there may be some variables not included in our analysis that influence the correlation found in this study. It is however difficult to see what such variables may be. Thirdly, we could only use an international food database that tracks the general market availability of different food types, not the actual human consumption. There are no direct measures of actual human consumption that can account for food wastage and provide precise measures of food consumption internationally. Fourthly, we were unable to analyze associations of food groups with obesity by each individual food item at country level. One of the main reasons is that some country may not access some particular food item due to its availability in their region, socio-economic status or cultural beliefs. For instance, pig meat (pork) is not consumed in Muslim countries or less consumed in countries with Muslim population, but they consumed mutton and lamp and other animal meat which share similar nutritional properties. Finally, the data analysed are calculated per capita in each country, so we can only demonstrate a relationship between food group availability and obesity, overweigh and mean BMI at a country level, which does not necessarily correspond to the same relationships holding true at the individual level. Prospective cohort studies are proposed to explore these associations further.

## Conclusion

By examining the per capita availability of macronutrients and the major food groups for 170 countries we are able to identify that countries with dietary patterns that are higher in meat have greater rates of obesity and overweight and higher mean BMI. Considering the findings of adverse effect of obesity on the risk of other chronic diseases revealed by other studies as well as the environmental impact of meat production, the country authorities may advise people not to adopt a high-meat diet for long-term healthy weight management.

### Declarations

List of abbreviations

FAO: The Food and Agriculture Organization of the United Nations

PANACEA: Physical Activity, Nutrition, Alcohol, Cessation of Smoking, Eating Out of Home and Obesity

SPSS: Statistical Package for the Social Sciences

LAC: Latin America and the Caribbean

OECD: The Organisation for Economic Co-operation and Development

EPIC: The European Prospective Investigation into Cancer and Nutrition

APEC: Asia-Pacific Economic Cooperation

SADC: Southern African Development Community

GDP: Gross Domestic Product

LA: Latin America

ACD: Asia Cooperation Dialogue

BMI: Body Mass Index

GHO: Global Health Observatory

FBS: Food Balance Sheet

UN: The United Nations

CI: Confidence Interval

GDP PPP: Gross domestic product per capita based on purchasing power parity

USA: United States of America

WHO: World Health Organization

AFRO: WHO Regional Office for Africa

AMRO: WHO Regional Office for the Americas

SEARO: WHO Regional Office for South-East Asia

EURO: WHO Regional Office for Europe

EMRO: WHO Regional Office for the Eastern Mediterranean

WPRO: WHO Regional Office for the Western Pacific

### Ethics

All the data supporting our findings in this paper were freely downloaded from the United Nations (UN) agencies' websites. No ethical approval or written informed consent for participation was required.

### Consent to publish

Not applicable.

### Competing interests

The authors (WPY & MH) declare that there are no conflicts of interest, financial or otherwise, related to this paper.

### Funding

There was no funding for this project.

### Authors' contributions

WPY reviewed the literature and obtained the data. MH and WPY formulated the hypothesis relating food groups to obesity, analysed the data and interpreted results and wrote the text. Both authors approved the final version of the manuscript.

### Availability of data and materials

All data for this study are publicly available and are ready for the public to download at no cost from the official websites of the World Bank, the WHO and FAO. Use of these data for this research falls within these UN agency's public permission in their terms and conditions. There is no need to have the formal permission to use the data for this study. The sources and data robustness have been described in the section of "Methods". Furthermore, all the data supporting our findings are contained within the Supplemental File titled, Table S2 Detailed information of country-level estimates.

### Acknowledgments

The authors express appreciation to the Statistics Department of FAO for the assistance in locating and defining the data.

## References

1. de Onis, M., M. Blossner, and E. Borghi, *Global prevalence and trends of overweight and obesity among preschool children.* Am J Clin Nutr, 2010. **92**(5): p. 1257-64.

2. Jang, M. and D. Berry, *Overweight, obesity, and metabolic syndrome in adults and children in South Korea: a review of the literature.* Clin Nurs Res, 2011. **20**(3): p. 276-91.

3. Ng, M., et al., *Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013.* The Lancet, 2014. **384**(9945): p. 766-781.

4. Grundy, S.M., *Metabolic syndrome pandemic.* Arterioscler Thromb Vasc Biol, 2008. **28**(4): p. 629-36.

5. Cornier, M.A., et al., *The metabolic syndrome.* Endocr Rev, 2008. **29**(7): p. 777-822.

6. Friend, A., L. Craig, and S. Turner, *The prevalence of metabolic syndrome in children: a systematic review of the literature.* Metab Syndr Relat Disord, 2013. **11**(2): p. 71-80.

7. WHO, *Global Health Risks Mortality and Burden of Disease Attributable to Selected Major Risks*. 2009, Geneva: Geneva : World Health Organization.

8. WHO. *Controlling the global obesity epidemic*. WHO [11.26.2015]; Available from: http://www.who.int/nutrition/topics/obesity/en/.

9. WHO. *WHO | Obesity*. WHO 2015 [11.26.2015]; Available from: http://who.int/topics/obesity/en/.

10. Jéquier, E. and G.A. Bray, *Low-fat diets are preferred.* The American Journal of Medicine, 2002. **113**(9): p. 41-46.

11. Willett, W. and R. Leibel, *Dietary fat is not a major determinant of body fat.* American Journal Of Medicine, 2002. **113**: p. 47-59.

12. Hession, M., et al., *Systematic review of randomized controlled trials of low-carbohydrate vs. low-fat/low-calorie diets in the management of obesity and its comorbidities.* Obes Rev, 2009. **10**(1): p. 36-50.

13. Mozaffarian, D., et al., *Changes in Diet and Lifestyle and Long-Term Weight Gain in Women and Men.* New England Journal Of Medicine, 2011. **364**(25): p. 2392-2404.

14. Brown, P.J. and M. Konner, *An Anthropological Perspective on Obesity.* Annals of the New York Academy of Sciences, 1987. **499**(1): p. 29-46.

15. Sumithran P and P. J., *The defence of body weight: a physiological basis for weight regain after weight loss.* Clin Sci (Lond), 2013.

16. Bes-Rastrollo, M., et al., *Predictors of weight gain in a Mediterranean cohort: the Seguimiento Universidad de Navarra Study.* American Journal Of Clinical Nutrition, 2006. **83**(2): p. 362-370.

17. French SA , et al., *Predictors of weight change over two years among a population of working adults: the Healthy Worker Project.* Int J Obes Relat Metab Disord, 1994. **18(3)**: p. 145-54.

18. Rosell, M., et al., *Weight gain over 5 years in 21,966 meat-eating, fish-eating, vegetarian, and vegan men and women in EPIC-Oxford.* Int. J. Obes., 2006. **30**(9): p. 1389-1396.

19. Schulz, M., et al., *Food groups as predictors for short-term weight changes in men and women of the EPIC-Potsdam cohort.* Journal Of Nutrition, 2002. **132**(6): p. 1335-1340.

20. Vergnaud, A., et al., *Meat consumption and prospective weight change in participants of the EPIC-PANACEA study.* Am. J. Clin. Nutr., 2010. **92**(2): p. 398-407.

21. Henneberg, M. and J. Grantham, *Obesity - a natural consequence of human evolution.* Anthropological Review, 2014. **77**(1): p. 1-10.

22. WHO. *Global Health Observatory, the data repository*. WHO 2015 [11.26.2015]; Available from: http://www.who.int/gho/database/en/.

23. FAO. *FAOSTAT-Food Balance Sheet*. 2015 [11.26.2015]; Available from: http://faostat3.fao.org/.

24. D.A.T. Southgate, A.R.C.F.R.I., Norwich, UK, , *Joint FAO/WHO/UNU Expert Consultation on Energy and Protein Requirements-The Rellationship between Food Composition and Available Energy*. October 1981, FAO/WHO/UNU: Rome.

25. Davis, B. and B. Wansink, *Fifty years of fat: news coverage of trends that predate obesity prevalence.* BMC Public Health, 2015. **15:629**: p. 1-6.

26. den Engelsen, C., et al., *Development of metabolic syndrome components in adults with a healthy obese phenotype: a 3-year follow-up.* Obesity (Silver Spring), 2013. **21**(5): p. 1025-1030.

27. Trøseid, M., et al., *Arterial stiffness is independently associated with interleukin-18 and components of the metabolic syndrome.* Atherosclerosis, 2010. **209**(2): p. 337-339.

28. FAO, *Food Balance Sheets. A Handbook*. 2001, Rome: Food and Agriculture Organization.

29. Grantham, J.P., et al., *Modern diet and metabolic variance--a recipe for disaster?* Nutrition journal, 2014. **13:15**: p. 01-10.

30. The World Bank: International Comparison Program database: World Development Indicators. *GDP (current US$) per capita per year* 2010 [11.26.2015]; Available from: http://data.worldbank.org.

31. Siervo, M., et al., *Sugar consumption and global prevalence of obesity and hypertension: an ecological analysis.* 2014. **17**(3): p. 587-596.

32. Roccisano, D. and M. Henneberg, *Soy Consumption and Obesity.* Food and Nutrition Sciences, 2012. **03**(02): p. 260-266.

33. Basu, S., et al., *The relationship of sugar to population-level diabetes prevalence: an econometric analysis of repeated cross-sectional data.* PLoS One, 2013. **8**(2): p. e57873: 1-8.

34. Basu, S., et al., *Nutritional determinants of worldwide diabetes: an econometric study of food markets and diabetes prevalence in 173 countries.* Public health nutrition, 2013. **16**(1): p. 1-8.

35. Weeratunga, P., et al., *Per capita sugar consumption and prevalence of diabetes mellitus – global and regional associations.* BMC Public Health, 2014. **14**: p. 186-91.

36. Giskes, K., et al., *Socioeconomic position at different stages of the life course and its influence on body weight and weight gain in adulthood: a longitudinal study with 13-year follow-up.* Obesity (Silver Spring), 2008. **16**(6): p. 1377-1381.

37. Nestle, M., *Increasing portion sizes in American diets: More calories, more obesity.* Journal of the American Dietetic Association, 2003. **103**(1): p. 39-40.

38. Berentzen, T. and T.I. Sorensen, *Physical inactivity, obesity and health.* Scand J Med Sci Sports, 2007. **17**(4): p. 301-2.

39. The World Bank. *Country and Lending Groups | Data*. 2015; Available from: http://data.worldbank.org/about/country-and-lending-groups.

40. WHO. *WHO regional offices*. [11.26.2015]; Available from: http://www.who.int.

41. The United Nations Educational Scientific and Cultural Organization. *UNESCO Regions-Latin America and the Caribbean*. 2014 [11.26.2015]; Available from: http://www.unesco.org.

42. Asia-Pacific Economic Cooperation. *Member Economies-Asia-Pacific Economic Cooperation*. [11.26.2015]; Available from: http://www.apec.org.

43. South Africa Development Community. *Southern African Development Community: Member States*. [18.06.2015]; Available from: http://www.sadc.int.

44. Asia Cooperation Dialogue. *Member Countries*. Available from: http://www.acddialogue.com.

45. Lin, Y., et al., *Plant and animal protein intake and its association with overweight and obesity among the Belgian population.* Br J Nutr, 2011. **105**(7): p. 1106-1116.

46. Wang, Y. and M.A. Beydoun, *Meat consumption is associated with obesity and central obesity among US adults.* International Journal of Obesity, 2009. **33**(6): p. 621.

47. Bradlee, M., et al., *Food group intake and central obesity among children and adolescents in the Third National Health and Nutrition Examination Survey (NHANES III).* Public Health Nutr., 2010. **13**(6): p. 797-805.

48. Bujnowski, D., et al., *Longitudinal association between animal and vegetable protein intake and obesity among men in the United States: the Chicago Western Electric Study.* J Am Diet Assoc, 2011. **111**(8): p. 1150-1155 e1.

49. Liu, J., et al., *Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese multi-provincial cohort study.* Jama-Journal Of The American Medical Association, 2004. **291**(21): p. 2591-2599.

50. Wang, Z., et al., *Fatty and lean red meat consumption in China: Differential association with Chinese abdominal obesity.* Nutr. Metab. Carbiovasc. Dis., 2014. **24**(8): p. 869-876.

51. Ryan, Y.M., *Meat avoidance and body weight concerns: nutritional implications for teenage girls.* Proceedings of the Nutrition Society 1997. **56**: p. 519-24.

52. Li, J., et al., *Incidence of obesity and its modifiable risk factors in Chinese adults aged 35-74 years.* Zhonghua Liu Xing Bing Xue Za Zhi, 2014. **35(4)**: p. 349-53

53. Pearce, K.L., H.C. Norman, and D.L. Hopkins, *The role of saltbush-based pasture systems for the production of high quality sheep and goat meat.* Small Ruminant Research, 2010. **91**(1): p. 29-38.

54. Lawrie, R.A. and D.A. Ledward, *Lawrie's meat science (7th ed.)*. 2006, Cambridge: Woodhead Publishing Limited.

55. Cordain, L., et al., *Plant-animal subsistence ratios and macronutrient energy estimations in worldwide hunter-gatherer diets.* American Journal Of Clinical Nutrition, 2000. **71**(3): p. 682-692.

56. Cordain, L., et al., *Plant-animal subsistence ratios and macronutrient energy estimations in worldwide hunter-gatherer diets.* American Journal Of Clinical Nutrition, 2000. **71**(3): p. 682-692.

57. Roy, R.P., *A Darwinian View of Obstructed Labor.* The American College of Obstetricians and Gynecologists, 2003. **101**(2): p. 397-401.

58. Eaton, S.B., M. Konner, and M. Shostak, *Stone-agers in the fast lane_chronicdegenerative diseases in evolutionary perspective.* The American Journal of Medicine, 1988. **84**(4): p. 739-49.

59. Richards, M.P., et al., *FOCUS: Gough's Cave and Sun Hole Cave human stable isotope values indicate a high animal protein diet in the British Upper Palaeolithic.* Journal Of Archaeological Science, 2000. **27**(1): p. 1-3.

60. Tarnopolsky, M.A., et al., *Evaluation of Protein-requirements for trained strength athletes.* Journal Of Applied Physiology, 1992. **73**(5): p. 1986-1995.

61. Hearth Foundation, *Position Statement on Very Low Carbohydrate Diets (April 2004).* 2015.

62. Layman, D., et al., *A Moderate-Protein Diet Produces Sustained Weight Loss and Long-Term Changes in Body Composition and Blood Lipids in Obese Adults.* J. Nutr., 2009. **139**(3): p. 514-521.

63. Samaha, F., et al., *A low-carbohydrate as compared with a low-fat diet in severe obesity.* New England Journal Of Medicine, 2003. **348**(21): p. 2074-2081.

64. WHO, *Diet, Nutrition and the Prevention of Chronic Diseases. Report of a Joint WHO/FAO Expert Consultation*, in *Technical Report Series No. 916*. 2003: Geneva: World Health Organization.

65. Elmadfa, I., *European Nutrition and Health Report 2009*. 2009, Basel; London: Karger.

66. Vergnaud, A.C., et al., *Macronutrient composition of the diet and prospective weight change in participants of the EPIC-PANACEA study.* PLoS One, 2013. **8**(3): p. e57300.

67. Mikkelsen, P., S. Toubro, and A. Astrup, *Effect of fat-reduced diets on 24-h energy expenditure: comparisons between animal protein, vegetable protein, and carbohydrate.* American Journal Of Clinical Nutrition, 2000. **72**(5): p. 1135-1141.

68. Toden, S., et al., *High red meat diets induce greater numbers of colonic DNA double-strand breaks than white meat in rats: attenuation by high-amylose maize starch.* Carcinogenesis, 2007. **28**(11): p. 2355-62.

69. Sabate, J. and M. Wien, *Vegetarian diets and childhood obesity prevention.* Am J Clin Nutr, 2010. **91**(5): p. 1525S-1529S.

70. Tuso, P.J., et al., *Nutritional update for physicians: plant-based diets.* Perm J, 2013. **17**(2): p. 61-6.

71. Key, T., et al., *Mortality in vegetarians and non-vegetarians: a collaborative analysis of 8300 deaths among 76,000 men and women in five prospective studies.* Public Health Nutrition, 1998. **1**(1): p. 33-41.

72. Raubenheimer, D., et al., *Nutritional ecology of obesity: from humans to companion animals.* The British journal of nutrition, 2015. **113 Suppl**: p. S26.

73. Wells, J.C., *The evolution of human fatness and susceptibility to obesity: an ethological approach.* Biol Rev Camb Philos Soc, 2006. **81**(2): p. 183-205.

74. Belobrajdic, D., G. McIntosh, and J. Owens, *A high-whey-protein diet reduces body weight gain and alters insulin sensitivity relative to red meat in Wistar rats.* Journal Of Nutrition, 2004. **134**(6): p. 1454-1458.

75. Schwarz, J., et al., *Dietary Protein Affects Gene Expression and Prevents Lipid Accumulation in the Liver in Mice.* PLoS ONE, 2012. **7**(10).

76. Ramel, A., M.T. Jonsdottir, and I. Thorsdottir, *Consumption of cod and weight loss in young overweight and obese adults on an energy reduced diet for 8-weeks.* Nutrition, Metabolism and Cardiovascular Diseases, 2009. **19**(10): p. 690-696.

77. Elliott, P., et al., *Association between protein intake and blood pressure: the INTERMAP Study.* Archives of internal medicine, 2006. **166**(1): p. 79.

78. Westman, E.C., *Is dietary carbohydrate essential for human nutrition?* American Journal Of Clinical Nutrition, 2002. **75**(5): p. 951-953.

79. Sarwar, H., *The importance of cereals (Poaceae: Gramineae) nutrition in human health: A review.* Journal of Cereals and Oilseeds, 2013. **4**(3): p. 32-35.

80. Latham, M.C., *Human nutrition in the developing world* 1997, Rome: Food and Agriculture Organization of the United Nations.

81. Jensen, H. and N.B.N.J. The Squibb Institute for Medical Research, *Our Present Knowledge of Carbohydrate Metabolism.* Transactions of the New York Academy of Sciences, 2015. **2**: p. 103-108.

82. Pi-sunyer, F.X., *Effect of the Composition of the Diet on Energy Intake.* Nutrition Reviews, 1990. **48**(2): p. 94-105.

83. Astrup, A., *Dietary composition, substrate balances and body fat in subjects with a predisposition to obesity.* Int J Obes Relat Metab Disord, 1993. **17 Suppl 3**: p. S32-6; discussion S41-2.

84. Romieu, I., et al., *Energy intake and other determinants of relative weight.* Energy intake and other determinants of relative weight, 1988. **47**(3): p. 406-412.

85. Drewnowski, A., et al., *Sweet tooth reconsidered: Taste responsiveness in human obesity.* Physiology & Behavior, 1985. **35**(4): p. 617-622.

86. Drewnowski, A., et al., *Food preferences in human obesity- carbohydrates versus fats.* Appetite, 1992. **18**(3): p. 207-221.

87. Lissner, L. and B.L. Heitmann, *Dietary fat and obesity: evidence from epidemiology.* Eur J Clin Nutr, 1995. **49**: p. 79-90.

88. Chen, J., et al., *Diet, life-style, and mortality in China: a study of the characteristics of 65 Chinese counties*. 1990, Oxford: Oxford: Oxford University Press ; Ithaca, N.Y : Cornell University Press.

89. Willett, W., *Is dietary fat a major determinant of body fat?* American Journal Of Clinical Nutrition, 1998. **67**(3): p. 556S-562S.

90. Centers for Disease Control and Prevention, *Daily dietary fat and total food-energy intakes - NHANES III, Phase 1, 1988-91.* JAMA, The Journal of the American Medical Association, 1994. **271**(17): p. 1309.

91. Kuczmarski, R.J., et al., *Increasing prevalence of overweight among US adults. The National Health and Nutrition Examination Surveys, 1960 to 1991.* JAMA, 1994. **272**(3): p. 205.

## Supplemental File

Table S1: Food items included in each food group for this study

The following food groups are based on the major nutrient content and the role of the foodstuffs in human nutrition.

| Food group | | Items included |
|---|---|---|
| This study | FAO definition | |
| Meat | Meat, total | Beef and veal, Buffalo meat, Pig meat, Mutton and lamb, Goat meat, Horse meat, Chicken meat, Goose meat, Duck meat, Turkey meat, Rabbit meat, Game meat and Offal |
| Fruits | Fruits | Melons, Watermelons, Apples, Apricots, Avocados, Cherries, Figs, Grapes, Mangoes, Papaya, Peaches, Pears, Persimmons, Pineapples, Plums, Quinces, Blueberries, Gooseberries, Raspberries, Strawberries, Kiwi, Dates, Figs (dried), Prunes, Currants, Raisins and Other fresh and dried fruits |
| Fibers | Vegetables | Beets, Carrots, Turnips, Rutabagas or swedes, Onions (green), Onions (dry), Artichokes, Tomatoes, Asparagus, Cabbage, Cauliflower, Celery, Kale, Lettuce, Spinach, Beans (green), Broad beans (green), Chilli peppers, Garlic, Cucumbers, Mushrooms, Eggplant, Peas (green), Pumpkins, Squash, Gourds, Okra, Radishes and Other vegetables |
| | Pulses | Beans (dry), Broad beans (dry), Peas (dry), Chick peas, Cow peas, Pigeon peas, Lentils, Vetches, Lupins and Other pulses |
| Starch | Starchy root | Potatoes, Sweet potatoes, Cassava, Taro, Yams and Other roots and tubers |
| | Cereals | Wheat, Rye, Barley, Oats, Maize, Rice, Mixed grains, Buckwheat, Sorghum, Millet, Quinoa and Other cereals |
| Fats | Vegetable oil | Sunflower seed oil, Cottonseed oil, Linseed oil, Hempseed oil, Sesame seed oil, Copra and coconut oil, Palm kernel oil, Palm oil, Soybean oil, Olive oil and Maize oil |
| | Animal fats | Butter, Ghee, Fish liver oil, Whale oil and Other animal fats |

*Source: [http://www.fao.org/docrep/003/x9892e/x9892e02.htm](http://www.fao.org/docrep/003/x9892e/x9892e02.htm)*

Table S2: Country-level estimates of variable included in this study

| Country | BMI≥30,18+,2010, WHO | BMI≥25, 18+, 2010, WHO | BMI mean,18+, 2010 WHO | Meat, total 2007-09 mean (g/capita/day), FAO | Fat (plant oil+ anima fat)(g/capita/day) 07-09, FAO | Fruits (g/capita/day) 2007-09 mean, FAO | Fiber (Vgetables + Pulses) (g/capita/day), 2007-09 FAO | Starch (Cereals Starchy root) (g/capita/da 2007-09 mean,FAO | Meat +protein(g/capita/day), 2007-09 mean, FAO | Animal protein, excl meat protein (g/capita/day) 2007-09 mean,FAO | Plant protein (g/capita/day) 2007-09 mean, FAO | Animal Fat Total (g/capita/day), 2007-09 mean, FAO | Plant oil total (g/capita/day),2007-09 mean, FAO | Carbohydrate (g/capita/day) 2007-09 mean, calculated | Calories Total (kcal/capita/day) 2007-09 mean, FAO | Minimum Dietary Energy Requirement (kcal/person/day) 2006-08, FAO | Physicak Inactivity %, 2010 18+, | GDP ppp (current US$), the World Bank, 2010 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Afghanistan | 2.4 | 12.0 | 21.2 | 31.7 | 12.6 | 54.3 | 100.4 | 523.4 | 4.6 | 7.0 | 45.3 | 14.9 | 15.9 | 387.7 | 2056.0 | | | 1604.2 |
| Albania | 16.1 | 54.5 | 25.9 | 124.9 | 29.8 | 331.7 | 606.8 | 498.3 | 16.5 | 33.5 | 47.9 | 59.6 | 34.0 | 416.3 | 2899.0 | 1930 | | 9297.5 |
| Algeria | 22.6 | 52.4 | 25.9 | 52.5 | 38.1 | 210.6 | 354.3 | 745.8 | 6.8 | 15.2 | 62.4 | 19.5 | 47.1 | 532.4 | 3066.3 | 1840 | 34.4 | 12241.3 |
| Angola | 8.3 | 21.0 | 23.6 | 69.9 | 25.7 | 189.5 | 174.4 | 930.2 | 9.1 | 6.2 | 35.0 | 14.6 | 33.6 | 401.4 | 2240.3 | 1750 | | 6904.5 |
| Antigua and Barbuda | 28.0 | 52.7 | 27.9 | 225.6 | | 515.9 | 152.7 | 225.6 | 29.9 | 30.0 | 24.7 | 51.1 | 32.3 | 320.2 | 2369.3 | 1880 | | 20151.3 |
| Argentina | 23.7 | 60.8 | 27.2 | 264.1 | 48.2 | 197.6 | 187.1 | 461.6 | 40.5 | 22.9 | 35.3 | 66.6 | 44.6 | 436.3 | 3141.3 | 1900 | 39.2 | |
| Armenia | 17.8 | 52.6 | 26.3 | 115.0 | 30.7 | 296.0 | 861.1 | 551.9 | 15.0 | 22.2 | 47.9 | 46.2 | 31.8 | 440.1 | 2802.7 | 1880 | | 6376.3 |
| Australia | 26.0 | 67.6 | 26.9 | 312.3 | 81.1 | 277.0 | 276.4 | 409.2 | 38.4 | 30.7 | 36.2 | 73.3 | 70.9 | 372.8 | 3210.0 | 1950 | 23.8 | 39086.4 |
| Austria | 16.7 | 59.3 | 25.3 | 294.6 | 120.5 | 398.9 | 283.6 | 472.4 | 34.5 | 29.5 | 42.7 | 91.4 | 78.2 | 459.8 | 3791.7 | 1970 | 23.8 | 41804.2 |
| Azerbaijan | 19.4 | 52.7 | 26.7 | 64.3 | 14.6 | 158.0 | 457.0 | 863.3 | 8.7 | 14.9 | 61.4 | 29.3 | 19.5 | 527.5 | 2889.0 | 1910 | | 15627.6 |
| Bahamas | 33.5 | 63.5 | 28.4 | 291.2 | 34.1 | 606.0 | 302.3 | 265.4 | 37.2 | 19.5 | 27.8 | 63.1 | 31.6 | 352.7 | 2600.7 | 1900 | 43.0 | 22451.8 |
| Bangladesh | 2.9 | 13.5 | 20.6 | 10.7 | 17.3 | 64.2 | 82.7 | 617.0 | 1.5 | 7.2 | 45.0 | 5.2 | 23.3 | 484.3 | 2409.3 | 1770 | 26.8 | 2409.5 |
| Barbados | 28.3 | 54.1 | 28.2 | 210.5 | 36.2 | 338.8 | 233.7 | 407.8 | 27.0 | 26.4 | 36.5 | 51.1 | 45.8 | 433.0 | 2963.3 | 1950 | 37.6 | 14987.4 |
| Belarus | 21.0 | 57.2 | 26.4 | 210.2 | 62.9 | 183.3 | 405.2 | 837.4 | 25.7 | 25.8 | 40.0 | 59.2 | 56.9 | 440.3 | 3172.3 | 1940 | | 15385.5 |
| Belgium | 18.7 | 63.8 | 25.4 | 218.7 | 137.9 | 235.6 | 353.6 | 559.8 | 25.1 | 38.0 | 39.1 | 92.8 | 72.8 | 466.4 | 3765.0 | 1980 | 33.2 | 39211.3 |
| Belize | 20.8 | 46.5 | 28.4 | 123.6 | 35.5 | 725.3 | 151.8 | 384.5 | 14.9 | 12.7 | 43.0 | 38.8 | 30.6 | 462.9 | 2758.3 | 1760 | | 8042.9 |
| Benin | 8.0 | 18.4 | 23.1 | 42.9 | 20.9 | 94.0 | 150.1 | 1056.5 | 5.7 | 5.4 | 51.2 | 6.3 | 42.5 | 468.6 | 2563.0 | 1740 | 6.9 | 1603.4 |
| Bolivia | 15.3 | 45.3 | 25.5 | 186.4 | 16.0 | 182.5 | 100.1 | 522.8 | 22.2 | 7.1 | 33.3 | 33.4 | 15.4 | 363.4 | 2143.3 | 1740 | | 5172.1 |

| Country | BMI≥30,18+,2010, WHO | BMI≥25, 18+, 2010, WHO | BMI mean,18+, 2010 WHO | Meat, total 2007-09 mean (g/capita/day), FAO | Fat (plant+ anima fat)(g/capita/day) 07-09, FAO | Fruits (g/capita/day) 2007-09 mean, FAO | Fiber (Vgetables + Pulses) (g/capita/day), 2007-09 FAO | Starch (Cereals Starchy root) (g/capita/da 2007-09 mean,FAO | Meat +protein(g/capita/day), 2007-09 mean, FAO | Animal protein, excl meat protein (g/capita/day) 2007-09 mean,FAO | Plant protein (g/capita/day) 2007-09 mean, FAO | Animal Fat Total (g/capita/day), 2007-09 mean, FAO | Plant oil total (g/capita/day),2007-09 mean, FAO | Carbohydrate (g/capita/day) 2007-09 mean, self-a calculated | Calories Total (kcal/capita/day) 2007-09 mean, FAO | Minimum Dietary Energy Requirement (kcal/person/day) 2006-08, FAO | Physicak Inactivity 2010 %, 18+,WHO | GDP ppp (current US$), the World Bank, 2010 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bosnia and Herzegovina | 16.8 | 53.0 | 26.0 | 66.5 | 25.8 | 246.7 | 546.7 | 693.7 | 8.4 | 22.0 | 59.6 | 32.8 | 38.3 | 521.0 | 3083.7 | 1960 | 18.1 | 8746.1 |
| Botswana | 19.4 | 33.1 | 24.4 | 73.2 | 35.4 | 140.3 | 108.1 | 438.9 | 11.8 | 11.9 | 36.8 | 18.0 | 38.4 | 358.5 | 2183.0 | 1840 | 27.2 | 13366.5 |
| Brazil | 17.8 | 51.9 | 25.6 | 230.1 | 57.0 | 311.4 | 183.9 | 487.9 | 28.5 | 16.7 | 42.3 | 50.6 | 58.9 | 453.5 | 3150.0 | 1860 | 27.8 | 14363.4 |
| Brunei Darussalam | 15.9 | 44.3 | 26.0 | 202.6 | 37.7 | 242.6 | 162.6 | 426.5 | 24.0 | 20.7 | 38.2 | 33.8 | 46.7 | 461.0 | 2900.3 | 1890 |  | 69207.6 |
| Bulgaria | 21.2 | 61.6 | 25.8 | 140.9 | 52.8 | 110.0 | 214.5 | 499.0 | 17.6 | 20.4 | 40.9 | 44.5 | 47.8 | 401.3 | 2751.3 | 1950 | 21.0 | 14690.5 |
| Burkina Faso | 5.3 | 16.4 | 21.8 | 45.3 | 16.7 | 15.3 | 86.8 | 603.2 | 6.3 | 5.1 | 66.7 | 11.6 | 46.0 | 440.1 | 2591.0 | 1740 | 18.4 | 1401.8 |
| Cabo Verde | 11.0 | 28.4 | 24.3 | 120.5 | 35.8 | 193.9 | 216.1 | 470.8 | 14.5 | 17.9 | 41.0 | 45.4 | 35.5 | 404.4 | 2639.7 | 1820 | 19.6 | 5883.5 |
| Cambodia | 2.3 | 11.5 | 21.5 | 44.3 | 7.7 | 76.5 | 103.1 | 557.7 | 5.5 | 12.5 | 43.6 | 16.1 | 19.2 | 460.0 | 2403.7 | 1760 | 10.3 | 2462.2 |
| Cameroon | 9.8 | 23.3 | 24.2 | 42.0 | 24.5 | 259.2 | 325.8 | 704.7 | 6.1 | 6.9 | 52.5 | 8.6 | 44.2 | 425.4 | 2439.0 | 1790 | 30.7 | 2518.8 |
| Canada | 25.9 | 67.2 | 27.0 | 264.2 | 117.1 | 373.9 | 340.6 | 510.8 | 32.8 | 25.4 | 46.2 | 67.8 | 77.7 | 430.2 | 3448.7 | 1950 | 23.2 | 40055.3 |
| Central African Republic | 4.4 | 13.5 | 22.2 | 96.0 | 37.5 | 128.5 | 78.7 | 842.1 | 14.0 | 5.8 | 31.5 | 17.8 | 54.1 | 316.9 | 2120.7 | 1750 | 12.0 | 882.9 |
| Chad | 7.1 | 16.7 | 22.0 | 34.3 | 11.4 | 26.9 | 46.8 | 558.1 | 5.1 | 4.6 | 51.5 | 7.7 | 42.5 | 332.1 | 2025.3 | 1760 | 24.6 | 1913.0 |
| Chile | 25.3 | 60.4 | 27.5 | 209.4 | 33.4 | 165.6 | 213.2 | 521.2 | 25.8 | 18.1 | 40.9 | 54.5 | 32.4 | 453.2 | 2934.3 | 1890 | 21.3 | 18111.7 |
| China | 5.3 | 31.1 | 23.4 | 149.2 | 27.6 | 190.2 | 873.1 | 588.9 | 17.5 | 17.8 | 56.0 | 53.0 | 36.5 | 346.6 | 2964.4 | 1910 | 24.1 | 9238.8 |
| Colombia | 19.0 | 51.3 | 25.6 | 120.0 | 40.9 | 355.5 | 121.5 | 481.8 | 14.6 | 17.8 | 31.1 | 32.3 | 43.5 | 459.2 | 2773.0 | 1790 | 63.6 | 10558.9 |
| Congo | 9.5 | 24.3 | 23.0 | 61.2 | 27.4 | 165.1 | 105.6 | 898.5 | 8.9 | 9.0 | 27.2 | 7.6 | 38.8 | 382.6 | 2128.7 | 1810 | 25.4 | 619.5 |
| Costa Rica | 21.9 | 55.4 | 26.5 | 136.8 | 51.0 | 246.8 | 172.4 | 353.5 | 16.2 | 23.0 | 36.0 | 38.8 | 48.7 | 439.8 | 2847.7 | 1890 |  | 12072.5 |
| Côte d'Ivoire | 7.9 | 21.9 | 23.4 | 37.7 | 33.8 | 218.6 | 112.2 | 1162.8 | 6.1 | 7.7 | 42.4 | 5.3 | 46.9 | 514.6 | 2752.7 | 1780 | 22.6 | 2833.7 |
| Croatia | 21.4 | 61.7 | 25.3 | 175.6 | 65.8 | 294.5 | 242.4 | 492.5 | 18.4 | 28.2 | 36.9 | 63.8 | 56.9 | 426.0 | 3124.3 | 1980 | 16.2 | 18968.9 |

| Country | BMI≥30, 18+, 2010, WHO | BMI≥25, 18+, 2010, WHO | BMI mean, 18+, 2010 WHO | Meat, total 2007-09 mean (g/capita/day), FAO | Fat (plant oil+ animal fat)(g/capita/day) 07-09, FAO | Fruits (g/capita/day) 2007-09 mean, FAO | Fiber (Vgetables + Pulses) (g/capita/day), 2007-09 FAO | Starch (Cereals Starchy root) (g/capita/day 2007-09 mean,FAO | Meat +protein(g/capita/day), 2007-09 mean, FAO | Animal protein, excl meat protein (g/capita/day) 2007-09 mean,FAO | Plant protein (g/capita/day) 2007-09 mean, FAO | Animal Fat Total (g/capita/day), 2007-09 mean, FAO | Plant oil total (g/capita/day),2007-09 mean, FAO | Carbohydrate (g/capita/day) 2007-09 mean, self-calculated | Calories Total (kcal/capita/day) 2007-09 mean, FAO | Minimum Dietary Energy Requirement (kcal/person/day) 2006-08, FAO | Physicak Inactivity 2010 %, 18+, WHO | GDP ppp (current US$), the World Bank, 2010 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cuba | 22.5 | 50.5 | 25.7 | 132.3 | 17.9 | 397.4 | 449.2 | 623.1 | 16.3 | 12.9 | 56.9 | 31.8 | 29.2 | 577.1 | 3201.7 | 1910 | | 18103.2 |
| Cyprus | 22.0 | 62.0 | 26.8 | 221.4 | 49.2 | 269.8 | 307.3 | 311.2 | 25.3 | 23.3 | 32.0 | 49.6 | 65.6 | 320.8 | 2642.0 | 1940 | 34.7 | 33934.4 |
| Czech Republic | 25.3 | 67.7 | 26.8 | 231.0 | 92.8 | 203.5 | 210.2 | 507.4 | 26.1 | 29.4 | 38.2 | 65.6 | 70.6 | 417.3 | 3270.0 | 1990 | 23.8 | 27051.0 |
| Dem. Republic of Korea | 2.2 | 18.0 | 21.8 | 36.2 | 15.9 | 148.7 | 410.7 | 566.8 | 5.0 | 4.8 | 47.0 | 9.3 | 24.6 | 391.3 | 2097.3 | 1870 | | |
| Denmark | 17.7 | 62.6 | 25.1 | 214.5 | 89.9 | 315.5 | 308.0 | 579.7 | 26.9 | 41.2 | 39.7 | 95.0 | 34.5 | 443.2 | 3370.0 | 1980 | 24.3 | 41806.9 |
| Djibouti | 8.3 | 24.4 | 23.2 | 75.7 | 40.6 | 56.3 | 191.3 | 453.0 | 10.3 | 8.3 | 43.7 | 17.5 | 49.5 | 397.3 | 2441.0 | 1830 | | 2611.4 |
| Dominica | 23.0 | 50.0 | 26.5 | 154.4 | | 964.7 | 200.7 | 591.7 | 21.5 | 30.5 | 40.4 | 44.2 | 35.4 | 497.5 | 3075.3 | 1870 | 21.8 | 10165.1 |
| Dominican Republic | 21.0 | 48.2 | 26.2 | 149.3 | 53.3 | 427.0 | 153.0 | 274.7 | 16.7 | 12.0 | 26.8 | 27.8 | 56.3 | 351.1 | 2382.3 | 1860 | 35.9 | 10850.9 |
| Ecuador | 16.8 | 48.7 | 26.6 | 150.4 | 57.3 | 386.8 | 75.8 | 294.6 | 18.0 | 21.9 | 23.9 | 51.6 | 47.7 | 303.5 | 2362.3 | 1780 | 25.2 | 9122.6 |
| Egypt | 26.2 | 52.8 | 28.7 | 73.7 | 22.1 | 293.3 | 634.4 | 777.9 | 9.9 | 12.7 | 77.6 | 21.7 | 39.3 | 627.6 | 3460.0 | 1850 | 32.3 | 10399.9 |
| El Salvador | 20.0 | 50.6 | 27.0 | 78.5 | 24.6 | 217.5 | 232.7 | 410.1 | 9.6 | 15.3 | 45.7 | 26.7 | 34.1 | 434.9 | 2570.0 | 1780 | | 7090.6 |
| Estonia | 21.2 | 59.1 | 25.4 | 166.2 | 40.5 | 224.1 | 305.7 | 626.2 | 18.9 | 32.9 | 43.0 | 52.3 | 40.6 | 502.9 | 3226.3 | 1990 | 11.9 | 21085.1 |
| Ethiopia | 3.3 | 10.8 | 20.3 | 22.1 | 8.3 | 20.3 | 89.6 | 573.9 | 3.2 | 4.4 | 51.0 | 8.2 | 15.9 | 389.2 | 2009.0 | 1750 | 18.9 | 1059.4 |
| Fiji | 35.0 | 66.9 | 27.1 | 115.9 | 46.5 | 121.3 | 138.9 | 666.8 | 14.7 | 16.6 | 42.3 | 39.3 | 54.9 | 446.2 | 2927.3 | 1860 | 17.0 | 6954.0 |
| Finland | 19.0 | 60.2 | 25.9 | 202.8 | 60.6 | 247.1 | 228.0 | 495.3 | 25.4 | 42.7 | 42.5 | 93.2 | 38.3 | 405.3 | 3246.7 | 1990 | 23.5 | 38296.5 |
| France | 22.0 | 65.2 | 25.2 | 242.2 | 104.9 | 317.1 | 284.3 | 483.4 | 30.4 | 41.0 | 40.1 | 95.0 | 69.3 | 398.1 | 3516.7 | 1950 | 23.8 | 35868.9 |
| Gabon | 15.7 | 37.1 | 25.1 | 170.0 | 14.8 | 409.6 | 120.0 | 770.3 | 24.1 | 16.8 | 40.5 | 20.8 | 33.2 | 472.5 | 2702.0 | 1810 | 26.0 | 15659.0 |
| Gambia | 9.0 | 22.9 | 23.5 | 24.9 | 42.8 | 17.9 | 97.6 | 524.2 | 3.4 | 12.4 | 46.5 | 9.1 | 61.1 | 427.2 | 2589.0 | 1770 | 21.5 | 1633.4 |
| Georgia | 18.6 | 51.5 | 26.8 | 72.6 | 29.0 | 110.5 | 165.2 | 681.4 | 8.9 | 18.6 | 49.9 | 30.4 | 35.7 | 498.4 | 2897.7 | 1930 | 20.6 | 5818.2 |

| Country | BMI≥30,18+,2010, WHO | BMI≥25, 18+, 2010, WHO | BMI mean,18+, 2010 WHO | Meat, total 2007-09 mean (g/capita/day), FAO | Fat (plant oil+ animal fat)(g/capita/day) 07-09, FAO | Fruits (g/capita/day) 2007-09 mean, FAO | Fiber (Vgetables + Pulses) (g/capita/day), 2007-09 FAO | Starch (Cereals Starchy root) (g/capita/day 2007-09 mean,FAO | Meat +protein(g/capita/day), 2007-09 mean, FAO | Animal protein, excl meat protein (g/capita/day) 2007-09 mean, FAO | Plant protein (g/capita/day) 2007-09 mean, FAO | Animal Fat Total (g/capita/day), 2007-09 mean, FAO | Plant oil total (g/capita/day),2007-09 mean, FAO | Carbohydrate (g/capita/day) 2007-09 mean, self-calculated | Calories Total (kcal/capita/day) 2007-09 mean, FAO | Minimum Dietary Energy Requirement (kcal/person/day) 2006-08, FAO | Physicak Inactivity 2010 %, 18+,WHO | GDP ppp (current US$), the World Bank, 2010 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Germany | 18.5 | 60.7 | 26.1 | 235.4 | 102.5 | 217.8 | 249.1 | 497.2 | 27.8 | 34.2 | 40.1 | 80.4 | 64.0 | 452.3 | 3516.7 | 1980 | 21.1 | 39553.2 |
| Ghana | 10.1 | 20.7 | 23.8 | 38.0 | 22.6 | 456.5 | 93.4 | 1345.3 | 5.4 | 10.6 | 41.9 | 6.2 | 41.3 | 549.6 | 2858.0 | 1790 | 15.6 | 3028.3 |
| Greece | 21.3 | 64.2 | 27.3 | 213.0 | 82.4 | 409.9 | 654.4 | 569.7 | 25.5 | 37.8 | 50.7 | 59.7 | 93.3 | 428.6 | 3547.3 | 1950 | 12.9 | 28900.8 |
| Grenada | 23.6 | 47.6 | 26.4 | 175.6 | 8.8 | 368.8 | 120.7 | 219.5 | 20.6 | 26.2 | 27.3 | 40.4 | 49.6 | 330.5 | 2428.0 | 1860 | 30.5 | 10952.2 |
| Guatemala | 17.1 | 45.5 | 26.1 | 75.3 | 26.4 | 184.2 | 219.2 | 383.9 | 8.9 | 9.4 | 43.5 | 17.1 | 40.9 | 414.8 | 2428.0 | 1690 | 13.3 | 6710.7 |
| Guinea | 5.8 | 17.4 | 22.5 | 21.9 | 38.7 | 266.3 | 155.1 | 679.5 | 3.2 | 4.9 | 45.3 | 5.5 | 52.5 | 446.6 | 2522.0 | 1760 | 9.9 | 1161.5 |
| Guinea-Bissau | 6.1 | 17.8 | 22.8 | 42.1 | 37.9 | 136.6 | 59.9 | 562.3 | 5.3 | 3.1 | 34.9 | 12.3 | 46.7 | 395.2 | 2284.7 | 1760 | | 1340.2 |
| Guyana | 20.5 | 41.1 | 25.8 | 108.3 | 10.6 | 111.8 | 197.0 | 478.1 | 12.9 | 18.7 | 41.9 | 23.5 | 30.2 | 460.2 | 2618.0 | 1850 | | 5430.0 |
| Haiti | 10.4 | 28.9 | 23.6 | 39.5 | 22.5 | 206.5 | 104.8 | 485.5 | 5.1 | 3.7 | 37.2 | 9.0 | 31.7 | 366.0 | 2014.7 | 1860 | | 1487.0 |
| Honduras | 16.5 | 44.4 | 26.0 | 97.0 | 36.9 | 195.4 | 172.5 | 384.5 | 12.3 | 14.2 | 39.8 | 28.1 | 44.7 | 417.9 | 2592.0 | 1730 | | 4183.7 |
| Hungary | 22.5 | 64.6 | 26.2 | 221.5 | 112.9 | 207.7 | 282.1 | 458.0 | 24.6 | 23.3 | 38.5 | 78.7 | 62.3 | 390.3 | 3175.3 | 1970 | 18.1 | 21477.6 |
| Iceland | 21.3 | 63.2 | 25.9 | 254.0 | 47.2 | 372.3 | 199.3 | 377.9 | 33.6 | 65.3 | 36.6 | 105.4 | 43.8 | 385.2 | 3425.3 | 1980 | | 38659.0 |
| India | 4.0 | 17.3 | 21.6 | 12.1 | 31.4 | 132.2 | 236.5 | 496.5 | 1.6 | 9.9 | 46.6 | 14.3 | 36.5 | 431.3 | 2414.0 | 1780 | 13.4 | 4454.6 |
| Indonesia | 4.3 | 17.3 | 22.5 | 30.7 | 25.6 | 178.4 | 110.0 | 643.4 | 3.8 | 11.6 | 41.5 | 9.6 | 42.6 | 464.9 | 2556.3 | 1820 | 23.7 | 8326.3 |
| Iran | 23.9 | 56.1 | 25.9 | 90.7 | 39.4 | 450.8 | 618.9 | 642.5 | 12.3 | 10.9 | 61.0 | 24.5 | 50.2 | 506.6 | 3035.3 | 1830 | 33.5 | 16636.2 |
| Iraq | 21.0 | 49.5 | 27.5 | 34.8 | 44.3 | 71.9 | 374.9 | 510.7 | 4.6 | 5.3 | 48.8 | 9.3 | 52.8 | 404.9 | 2412.7 | | 49.3 | 12080.3 |
| Ireland | 23.1 | 64.6 | 27.2 | 246.1 | 77.9 | 394.4 | 254.7 | 615.0 | 31.9 | 34.1 | 44.0 | 74.1 | 56.9 | 483.4 | 3552.0 | 1950 | 35.1 | 42900.1 |
| Israel | 23.5 | 66.2 | 26.1 | 271.6 | 77.0 | 372.6 | 488.2 | 514.9 | 40.2 | 31.0 | 55.9 | 47.9 | 98.8 | 423.4 | 3521.7 | 1820 | | 28762.6 |
| Italy | 19.6 | 62.7 | 25.9 | 238.6 | 107.9 | 419.7 | 439.1 | 535.9 | 29.9 | 30.1 | 50.9 | 69.6 | 85.3 | 435.8 | 3580.3 | 1950 | 33.2 | 34716.1 |
| Jamaica | 24.5 | 48.7 | 26.9 | 171.7 | 38.4 | 322.4 | 222.1 | 451.6 | 19.8 | 19.8 | 38.1 | 33.9 | 46.3 | 432.6 | 2762.3 | 1870 | 27.9 | 8201.1 |
| Japan | 2.9 | 27.2 | 22.6 | 126.4 | 46.7 | 150.4 | 283.7 | 371.9 | 15.0 | 34.2 | 40.0 | 34.2 | 53.4 | 399.0 | 2740.7 | 1890 | 33.8 | 33740.5 |

| Country | BMI≥30, 18+, 2010, WHO | BMI≥25, 18+, 2010, WHO | BMI mean, 18+, 2010 WHO | Meat, total 2007-09 mean (g/capita/day), FAO | Fat (plant+ animal fat)(g/capita/day) 07-09, FAO | Fruits (g/capita/day) 2007-09 mean, FAO | Fiber (Vgetables + Pulses) (g/capita/day), 2007-09 FAO | Starch (Cereals Starchy root) (g/capita/day 2007-09 mean,FAO | Meat +protein(g/capita/day), 2007-09 mean, FAO | Animal protein, excl meat protein (g/capita/day) 2007-09 mean,FAO | Plant protein (g/capita/day) 2007-09 mean, FAO | Animal Fat Total (g/capita/day), 2007-09 mean, FAO | Plant oil total (g/capita/day),2007-09 mean, FAO | Carbohydrate (g/capita/day) 2007-09 mean, self-calculated | Calories Total (kcal/capita/day) 2007-09 mean, FAO | Minimum Dietary Energy Requirement (kcal/person/day) 2006-08, FAO | Physicak Inactivity 2010 %, 18+, WHO | GDP ppp (current US$), the World Bank, 2010 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Jordan | 28.1 | 59.0 | 28.6 | 117.1 | 50.8 | 132.2 | 325.9 | 534.3 | 15.8 | 13.3 | 53.3 | 26.8 | 64.9 | 487.2 | 3103.7 | 1760 | 15.6 | 11028.4 |
| Kazakhstan | 21.2 | 57.0 | 27.0 | 173.5 | 60.2 | 98.8 | 498.3 | 675.1 | 22.6 | 30.4 | 41.5 | 59.6 | 60.0 | 429.4 | 3171.0 | 1910 | 20.6 | 19204.6 |
| Kenya | 5.6 | 16.5 | 22.6 | 47.9 | 22.5 | 172.1 | 172.2 | 565.0 | 6.9 | 9.8 | 42.4 | 18.3 | 31.5 | 360.7 | 2127.0 | 1760 | 19.2 | 2451.9 |
| Kiribati | 39.2 | 66.8 | 29.4 | 96.7 |  | 173.7 | 159.7 | 598.0 | 11.9 | 25.7 | 35.8 | 23.1 | 79.0 | 428.6 | 2926.7 | 1770 | 41.1 | 1694.4 |
| Kuwait | 36.8 | 72.8 | 29.5 | 299.4 | 60.5 | 191.9 | 489.4 | 491.7 | 38.5 | 22.5 | 51.5 | 55.2 | 68.5 | 489.3 | 3519.7 | 1950 | 56.6 | 73841.0 |
| Kyrgyzstan | 13.0 | 42.7 | 25.8 | 96.4 | 20.0 | 96.9 | 418.5 | 726.8 | 12.9 | 21.0 | 49.7 | 36.5 | 28.0 | 456.4 | 2739.7 | 1890 | 13.3 | 2733.7 |
| Laos | 2.5 | 12.8 | 22.1 | 51.7 | 7.1 | 168.3 | 342.4 | 589.9 | 6.1 | 6.6 | 48.7 | 15.0 | 19.2 | 432.1 | 2282.0 | 1700 | 10.3 | 3821.8 |
| Latvia | 22.2 | 59.0 | 25.7 | 182.3 | 97.4 | 150.6 | 318.3 | 624.8 | 20.6 | 36.8 | 39.3 | 70.3 | 55.5 | 433.7 | 3253.3 | 1930 | 22.0 | 17488.8 |
| Lebanon | 28.4 | 64.5 | 27.3 | 145.9 | 58.2 | 299.4 | 543.2 | 593.3 | 19.8 | 14.5 | 49.0 | 35.0 | 74.0 | 457.6 | 3144.0 | 1880 | 38.8 | 15934.0 |
| Lesotho | 12.8 | 18.1 | 24.6 | 52.8 | 4.6 | 50.9 | 83.1 | 768.9 | 7.6 | 3.0 | 62.0 | 10.7 | 22.1 | 492.7 | 2556.7 | 1780 | 7.2 | 2189.3 |
| Liberia | 5.4 | 15.8 | 23.7 | 29.1 | 46.8 | 127.4 | 75.7 | 716.1 | 4.0 | 1.9 | 31.0 | 4.7 | 52.7 | 391.1 | 2229.0 | 1760 | 27.5 | 674.6 |
| Libya | 30.0 | 63.4 | 28.0 |  |  |  |  |  |  |  |  |  |  |  |  | 1850 | 38.0 | 29648.5 |
| Lithuania | 23.8 | 59.5 | 26.4 | 225.5 | 53.2 | 199.8 | 309.5 | 698.6 | 26.0 | 50.4 | 49.6 | 65.7 | 34.1 | 525.6 | 3505.3 | 1930 | 18.4 | 20052.2 |
| Luxembourg | 20.8 | 65.3 | 26.4 | 281.4 | 60.0 | 517.8 | 260.6 | 472.6 | 34.1 | 36.2 | 38.2 | 84.8 | 62.7 | 449.2 | 3558.0 | 1990 | 28.5 | 84200.6 |
| Madagascar | 4.5 | 14.3 | 20.9 | 40.1 | 9.2 | 125.0 | 56.2 | 774.7 | 5.5 | 5.4 | 37.0 | 10.7 | 16.4 | 409.4 | 2074.0 | 1700 | 17.9 | 1362.3 |
| Malawi | 4.2 | 12.1 | 22.5 | 19.4 | 10.8 | 181.3 | 99.1 | 925.2 | 2.5 | 2.8 | 55.6 | 6.4 | 30.8 | 432.2 | 2307.3 | 1700 | 7.5 | 722.4 |
| Malaysia | 10.5 | 32.7 | 25.0 | 140.2 | 44.9 | 124.1 | 165.0 | 446.4 | 17.0 | 25.1 | 36.1 | 30.5 | 53.6 | 439.6 | 2828.3 | 1820 | 52.3 | 19985.7 |
| Maldives | 6.7 | 22.9 | 24.8 | 61.1 | 8.7 | 282.4 | 222.6 | 386.6 | 7.8 | 55.9 | 33.1 | 29.8 | 31.4 | 403.3 | 2550.7 | 1800 | 30.7 | 11437.7 |
| Mali | 5.7 | 17.4 | 22.4 | 64.5 | 24.9 | 79.1 | 180.5 | 673.4 | 9.4 | 16.5 | 53.6 | 24.4 | 39.8 | 471.0 | 2780.3 | 1770 | 23.7 | 1630.0 |
| Malta | 24.7 | 66.4 | 27.1 | 240.8 | 47.7 | 260.3 | 574.2 | 570.6 | 29.6 | 31.8 | 50.1 | 65.5 | 46.4 | 478.0 | 3365.7 | 1910 | 42.9 | 26671.5 |
| Mauritania | 8.2 | 23.0 | 24.5 | 80.4 | 40.3 | 42.5 | 107.3 | 472.6 | 10.9 | 20.7 | 49.3 | 28.5 | 46.0 | 447.2 | 2782.3 | 1800 | 45.1 | 3232.6 |

| Country | BMI≥30,18+,2010, WHO | BMI≥25, 18+, 2010, WHO | BMI mean,18+, 2010 WHO | Meat, total 2007-09 mean (g/capita/day), FAO | Fat (plant oil+ animal fat)(g/capita/day) 07-09, FAO | Fruits (g/capita/day) 2007-09 mean, FAO | Fiber (Vgetables + Pulses) (g/capita/day), 2007-09 FAO | Starch (Cereals Starchy root) (g/capita/day 2007-09 mean,FAO | Meat +protein(g/capita/day), 2007-09 mean, FAO | Animal protein, excl meat protein (g/capita/day) 2007-09 mean,FAO | Plant protein (g/capita/day) 2007-09 mean, FAO | Animal Fat Total (g/capita/day), 2007-09 mean, FAO | Plant oil total (g/capita/day),2007-09 mean, FAO | Carbohydrate (g/capita/day) 2007-09 mean, self-calculated | Calories Total (kcal/capita/day) 2007-09 mean, FAO | Minimum Dietary Energy Requirement (kcal/person/day) 2006-08, FAO | Physicak Inactivity 2010 %, 18+,WHO | GDP ppp (current US$), the World Bank, 2010 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mauritius | 15.0 | 35.5 | 25.3 | 135.1 | 55.3 | 144.8 | 233.5 | 511.3 | 17.6 | 20.9 | 48.8 | 24.3 | 63.4 | 482.5 | 3068.0 | 1880 | 25.2 | 15283.5 |
| Mexico | 25.9 | 60.4 | 27.8 | 166.4 | 37.5 | 299.7 | 189.5 | 487.8 | 19.9 | 20.9 | 47.1 | 44.4 | 46.2 | 476.7 | 3074.3 | 1860 | 26.0 | 14690.0 |
| Mongolia | 14.4 | 46.6 | 25.4 | 217.2 | 27.0 | 81.0 | 114.0 | 486.2 | 29.5 | 18.9 | 30.7 | 62.0 | 25.2 | 322.3 | 2389.3 | 1860 | 21.4 | 7552.4 |
| Montenegro | 18.7 | 59.5 | 25.9 | 177.1 | 47.2 | 552.1 | 647.5 | 543.2 | 19.9 | 36.8 | 53.1 | 73.5 | 61.5 | 492.7 | 3624.7 | 1950 | | 13339.0 |
| Morocco | 20.2 | 48.8 | 25.2 | 76.7 | 37.9 | 186.9 | 392.1 | 802.8 | 10.3 | 10.4 | 69.9 | 18.0 | 46.7 | 581.3 | 3270.0 | 1860 | | 6334.9 |
| Mozambique | 4.4 | 12.5 | 22.1 | 23.0 | 21.6 | 61.9 | 80.5 | 935.7 | 2.7 | 2.5 | 38.4 | 7.2 | 30.3 | 406.3 | 2137.3 | 1800 | 5.8 | 910.7 |
| Myanmar | 2.1 | 11.2 | 22.3 | 82.1 | 24.9 | 97.5 | 258.4 | 415.2 | 9.8 | 16.5 | 45.8 | 23.5 | 38.9 | 375.8 | 2353.0 | 1800 | 9.9 | |
| Namibia | 16.8 | 27.4 | 23.9 | 97.4 | 22.9 | 65.4 | 102.1 | 668.9 | 13.4 | 10.9 | 37.9 | 22.6 | 29.4 | 353.0 | 2128.3 | 1830 | 31.8 | 8262.1 |
| Nepal | 2.7 | 13.2 | 21.7 | 29.6 | 25.9 | 120.7 | 296.4 | 707.4 | 3.7 | 6.5 | 52.9 | 12.4 | 34.0 | 453.5 | 2484.0 | 1730 | 4.1 | 1958.6 |
| Netherlands | 18.0 | 61.2 | 25.3 | 226.4 | 59.6 | 351.0 | 255.9 | 478.1 | 31.5 | 41.9 | 35.5 | 77.9 | 55.0 | 408.1 | 3264.3 | 2000 | 15.5 | 44743.0 |
| New Zealand | 26.5 | 66.2 | 27.6 | 307.9 | 64.3 | 313.9 | 359.3 | 425.7 | 36.6 | 20.9 | 35.9 | 81.9 | 37.5 | 432.9 | 3179.3 | 1940 | 39.8 | 31028.1 |
| Nicaragua | 15.7 | 41.9 | 26.6 | 67.0 | 28.7 | 119.1 | 81.2 | 415.8 | 8.2 | 10.5 | 45.0 | 20.0 | 39.4 | 419.8 | 2469.0 | 1760 | | 3885.6 |
| Niger | 3.7 | 12.4 | 21.5 | 61.9 | 16.7 | 74.8 | 223.1 | 604.3 | 9.0 | 8.3 | 61.5 | 16.7 | 39.0 | 421.4 | 2501.3 | 1700 | 25.1 | 824.2 |
| Nigeria | 8.9 | 23.0 | 23.1 | 24.9 | 37.8 | 164.1 | 210.0 | 1007.9 | 3.4 | 6.7 | 54.0 | 6.2 | 56.7 | 471.7 | 2708.7 | 1730 | 22.3 | 5010.3 |
| Norway | 21.6 | 63.2 | 25.9 | 182.9 | 93.3 | 383.4 | 222.1 | 521.8 | 22.7 | 42.3 | 43.8 | 88.4 | 59.2 | 426.6 | 3470.7 | 1970 | 25.8 | 58772.4 |
| Pakistan | 4.7 | 19.2 | 23.4 | 39.7 | 47.3 | 93.3 | 102.3 | 425.1 | 5.2 | 20.0 | 37.7 | 34.6 | 38.8 | 378.0 | 2423.7 | 1740 | 26.0 | 4134.2 |
| Panama | 23.7 | 55.4 | 26.6 | 162.2 | 35.2 | 241.9 | 94.3 | 399.0 | 19.5 | 19.2 | 34.1 | 34.9 | 31.6 | 403.1 | 2502.7 | 1790 | | 14619.5 |
| Paraguay | 14.5 | 45.5 | 25.4 | 115.3 | 48.8 | 209.0 | 167.7 | 640.5 | 14.7 | 14.6 | 36.7 | 35.7 | 54.2 | 368.0 | 2545.0 | 1820 | 24.6 | 6865.8 |
| Peru | 18.5 | 50.6 | 25.9 | 56.4 | 19.0 | 267.4 | 190.5 | 662.6 | 10.1 | 15.1 | 45.0 | 16.0 | 28.6 | 458.5 | 2516.0 | 1780 | | 9714.5 |
| Philippines | 4.1 | 18.9 | 22.9 | 91.4 | 20.6 | 295.5 | 181.6 | 518.1 | 10.9 | 14.3 | 34.8 | 30.3 | 20.4 | 471.8 | 2582.7 | 1760 | 15.8 | 5500.2 |
| Poland | 23.1 | 63.4 | 26.1 | 205.4 | 78.7 | 140.7 | 339.2 | 736.9 | 26.6 | 23.7 | 48.9 | 72.3 | 41.2 | 488.1 | 3370.3 | 1970 | 18.7 | 20757.4 |

| Country | BMI≥30,18+,2010, WHO | BMI≥25, 18+, 2010, WHO | BMI mean,18+, 2010 WHO | Meat, total 2007-09 mean (g/capita/day), FAO | Fat (plant oil+ animal fat)(g/capita/day) 07-09, FAO | Fruits (g/capita/day) 2007-09 mean, FAO | Fiber (Vgetables + Pulses) (g/capita/day), 2007-09 FAO | Starch (Cereals Starchy root) (g/capita/day 2007-09 mean,FAO | Meat +protein(g/capita/day), 2007-09 mean, FAO | Animal protein, excl meat protein (g/capita/day) 2007-09 mean,FAO | Plant protein (g/capita/day) 2007-09 mean, FAO | Animal Fat Total (g/capita/day), 2007-09 mean, FAO | Plant oil total (g/capita/day),2007-09 mean, FAO | Carbohydrate (g/capita/day) 2007-09 mean, self-calculated | Calories Total (kcal/capita/day) 2007-09 mean, FAO | Minimum Dietary Energy Requirement (kcal/person/day) 2006-08, FAO | Physicak Inactivity 2010 %, 18+,WHO | GDP ppp (current US$), the World Bank, 2010 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Portugal | 18.4 | 59.3 | 26.1 | 256.2 | 90.3 | 317.7 | 496.6 | 563.9 | 32.0 | 39.8 | 44.6 | 75.8 | 68.2 | 452.4 | 3571.0 | 1950 | 34.9 | 26924.4 |
| Republic of Korea | 4.2 | 32.7 | 23.6 | 154.0 | 53.5 | 207.2 | 592.2 | 440.8 | 17.3 | 25.3 | 49.1 | 35.4 | 58.0 | 492.0 | 3175.7 | 1900 | 33.4 | 30465.3 |
| Republic of Moldova | 13.6 | 43.4 | 26.5 | 89.3 | 47.0 | 120.4 | 285.2 | 591.0 | 10.9 | 21.4 | 38.8 | 38.3 | 44.1 | 435.4 | 2767.7 | 1930 | 12.3 | 3845.6 |
| Romania | 20.2 | 60.3 | 25.2 | 172.6 | 45.9 | 165.2 | 480.3 | 749.3 | 21.2 | 34.3 | 54.2 | 60.7 | 46.6 | 501.5 | 3410.0 | 1980 | 25.3 | 16252.2 |
| Russian Federation | 22.2 | 57.2 | 26.2 | 172.9 | 52.4 | 181.3 | 293.6 | 715.8 | 21.3 | 30.9 | 46.8 | 53.4 | 41.3 | 511.8 | 3295.7 | 1950 | 9.5 | 20498.0 |
| Rwanda | 3.1 | 10.6 | 21.7 | 17.2 | 9.8 | 425.4 | 217.5 | 865.4 | 2.6 | 2.6 | 46.8 | 4.6 | 20.4 | 404.3 | 2049.3 | 1710 | 15.3 | 1236.5 |
| Saint Kitts and Nevis | 25.7 | 50.8 | 29.2 | 142.9 |  | 233.2 | 74.6 | 250.4 | 26.1 | 20.2 | 28.6 | 37.7 | 42.8 | 361.4 | 2469.7 | 1850 | 32.4 | 20064.4 |
| Saint Lucia | 23.6 | 48.1 | 28.8 | 253.9 | 15.9 | 280.5 | 109.9 | 349.0 | 30.9 | 22.8 | 32.0 | 55.4 | 24.7 | 390.4 | 2624.7 | 1870 | 41.2 | 10302.6 |
| Saint Vincent & Grenadines | 21.8 | 47.6 | 26.8 | 251.4 |  | 527.8 | 266.4 | 435.7 | 31.7 | 17.3 | 36.8 | 40.0 | 39.0 | 452.2 | 2862.7 | 1870 |  | 9734.0 |
| Samoa | 41.8 | 68.0 | 31.4 | 224.1 | 34.9 | 533.0 | 35.6 | 552.9 | 28.1 | 17.5 | 31.6 | 58.7 | 73.5 | 340.0 | 2859.0 | 1810 | 16.2 | 5295.7 |
| Sao Tome and Principe | 10.4 | 23.7 | 24.4 | 38.1 |  | 489.2 | 196.3 | 717.7 | 4.7 | 11.9 | 43.7 | 9.9 | 65.2 | 429.4 | 2634.3 | 1700 | 15.6 | 2687.9 |
| Saudi Arabia | 32.0 | 65.3 | 28.1 | 142.6 | 48.5 | 218.1 | 270.5 | 491.5 | 18.9 | 11.3 | 50.7 | 25.0 | 56.8 | 481.6 | 2987.0 | 1860 | 61.0 | 44675.4 |
| Senegal | 8.5 | 21.0 | 22.7 | 35.3 | 44.5 | 46.4 | 178.4 | 526.0 | 4.9 | 11.7 | 43.3 | 10.9 | 58.1 | 386.4 | 2406.3 | 1760 | 25.0 | 2137.6 |
| Serbia | 17.9 | 57.6 | 25.7 | 124.6 | 31.1 | 316.6 | 341.2 | 446.7 | 15.7 | 20.8 | 42.8 | 45.0 | 37.6 | 414.6 | 2719.0 | 1960 | 38.7 | 11805.3 |
| Sierra Leone | 6.2 | 16.2 | 22.5 | 19.5 | 36.3 | 102.6 | 174.6 | 609.9 | 2.7 | 10.7 | 39.3 | 4.8 | 46.6 | 384.7 | 2212.3 | 1750 | 14.2 | 1319.9 |
| Slovakia | 23.4 | 63.6 | 26.3 | 165.6 | 68.6 | 179.3 | 261.7 | 503.2 | 17.9 | 16.9 | 37.6 | 59.9 | 42.3 | 410.7 | 2852.3 | 1990 | 17.8 | 24432.2 |
| Slovenia | 23.2 | 64.1 | 26.7 | 237.7 | 80.4 | 351.9 | 227.3 | 526.7 | 28.2 | 30.5 | 42.3 | 67.6 | 55.1 | 420.5 | 3190.0 | 1960 | 21.3 | 27563.7 |

| Country | BMI≥30, 18+, 2010, WHO | BMI≥25, 18+, 2010, WHO | BMI mean, 18+, 2010 WHO | Meat, total 2007-09 mean (g/capita/day), FAO | Fat (plant oil+ animal fat)(g/capita/day) 07-09, FAO | Fruits (g/capita/day) 2007-09 mean, FAO | Fiber (Vgetables + Pulses) (g/capita/day), 2007-09 FAO | Starch (Cereals Starchy root) (g/capita/day 2007-09 mean,FAO | Meat +protein(g/capita/day), 2007-09 mean, FAO | Animal protein, excl meat protein (g/capita/day) 2007-09 mean,FAO | Plant protein (g/capita/day) 2007-09 mean, FAO | Animal Fat Total (g/capita/day), 2007-09 mean, FAO | Plant oil total (g/capita/day),2007-09 mean, FAO | Carbohydrate (g/capita/day) 2007-09 mean, self-calculated | Calories Total (kcal/capita/day) 2007-09 mean, FAO | Minimum Dietary Energy Requirement (kcal/person/day) 2006-08, FAO | Physicak Inactivity 2010 %, 18+,WHO | GDP ppp (current US$), the World Bank, 2010 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Solomon Islands | 26.1 | 52.0 | 25.4 | 30.8 | 7.2 | 163.2 | 62.9 | 1169.3 | 4.2 | 12.6 | 37.7 | 12.9 | 35.0 | 450.5 | 2451.0 | 1730 | 35.1 | 1751.2 |
| Somalia | 3.9 | 14.2 | 21.7 | | | | | | | | | | | | | | | |
| South Africa | 24.1 | 39.3 | 26.9 | 152.4 | 38.9 | 94.3 | 127.0 | 570.2 | 21.3 | 11.2 | 48.3 | 32.4 | 50.2 | 465.0 | 2926.7 | 1900 | 46.9 | 11842.4 |
| Spain | 22.1 | 64.6 | 26.6 | 280.1 | 91.7 | 236.4 | 421.9 | 440.7 | 33.7 | 34.7 | 41.3 | 59.5 | 91.4 | 356.3 | 3221.7 | 1960 | 30.5 | 32350.5 |
| Sri Lanka | 4.8 | 16.0 | 22.5 | 17.7 | 8.6 | 88.3 | 131.2 | 458.5 | 2.3 | 11.6 | 41.8 | 8.2 | 37.2 | 448.2 | 2424.7 | 1800 | 23.8 | 7418.8 |
| Sudan | 6.3 | 20.0 | 24.9 | 66.6 | 19.0 | 147.6 | 202.9 | 376.8 | 9.6 | 20.9 | 44.4 | 37.7 | 32.5 | 353.4 | 2344.7 | 1780 | | 3215.6 |
| Suriname | 23.4 | 50.1 | 26.8 | 134.8 | 42.6 | 227.0 | 138.1 | 438.0 | 15.0 | 10.6 | 34.2 | 22.8 | 52.0 | 444.0 | 2688.0 | 1870 | | 14034.8 |
| Swaziland | 15.9 | 25.8 | 26.2 | 68.8 | 12.7 | 207.1 | 57.0 | 515.0 | 10.1 | 6.5 | 40.9 | 15.0 | 31.2 | 409.8 | 2285.0 | 1790 | 36.8 | 5743.1 |
| Sweden | 18.8 | 60.8 | 25.6 | 216.9 | 90.2 | 337.2 | 251.3 | 440.7 | 28.8 | 42.6 | 36.4 | 70.3 | 55.2 | 389.6 | 3119.3 | 1980 | 28.7 | 41727.2 |
| Switzerland | 17.8 | 61.5 | 25.1 | 202.2 | 93.7 | 238.8 | 277.6 | 402.0 | 24.0 | 35.9 | 34.1 | 91.8 | 65.3 | 414.2 | 3446.3 | 1990 | | 51321.5 |
| Syrian Arab Republic | 21.3 | 52.2 | 27.7 | | | | | | | | | | | | | 1800 | | |
| Tajikistan | 12.1 | 38.6 | 24.9 | 37.1 | 27.4 | 57.5 | 368.2 | 527.0 | 4.8 | 6.4 | 40.3 | 12.8 | 44.9 | 336.5 | 2071.0 | 1840 | | 2067.6 |
| Thailand | 6.7 | 22.7 | 23.8 | 78.6 | 22.0 | 324.1 | 142.6 | 432.5 | 9.3 | 13.8 | 35.3 | 24.1 | 35.6 | 495.4 | 2752.3 | 1850 | 14.8 | 12562.3 |
| FYR Macedonia | 18.2 | 57.5 | 25.7 | 115.7 | 67.9 | 317.6 | 494.5 | 518.6 | 14.3 | 19.4 | 45.0 | 46.7 | 64.7 | 421.4 | 3002.7 | 1960 | | 11449.5 |
| Timor-Leste | 1.7 | 9.7 | 20.9 | 93.1 | 13.0 | 40.9 | 102.8 | 657.7 | 13.4 | 2.4 | 38.9 | 15.1 | 25.5 | 366.1 | 2049.0 | 1700 | | 1741.2 |
| Togo | 6.5 | 16.1 | 22.9 | 29.8 | 22.9 | 23.9 | 106.8 | 916.9 | 4.0 | 3.0 | 47.0 | 5.3 | 41.7 | 419.9 | 2318.7 | 1770 | 10.4 | 1220.8 |
| Trinidad and Tobago | 27.0 | 51.5 | 28.0 | 169.3 | 39.6 | 199.1 | 122.3 | 432.2 | 19.5 | 16.7 | 36.9 | 38.2 | 43.3 | 452.7 | 2836.7 | 1910 | 41.5 | 28727.9 |
| Tunisia | 24.6 | 56.0 | 26.3 | 71.2 | 57.5 | 240.3 | 640.5 | 657.3 | 9.8 | 15.9 | 67.8 | 21.3 | 67.6 | 531.0 | 3298.0 | 1860 | 23.5 | 10200.3 |
| Turkey | 27.0 | 61.1 | 27.5 | 69.1 | 69.5 | 335.8 | 680.3 | 731.4 | 9.2 | 19.2 | 73.1 | 27.3 | 84.8 | 550.4 | 3616.7 | 1930 | 32.8 | 16193.4 |

| Country | BMI≥30,18+,2010, WHO | BMI≥25, 18+, 2010, WHO | BMI mean,18+, 2010 WHO | Meat, total 2007-09 mean (g/capita/day), FAO | Fat (plant+ animal fat)(g/capita/day) 07-09, FAO | Fruits (g/capita/day) 2007-09 mean, FAO | Fiber (Vgetables + Pulses) (g/capita/day), 2007-09 FAO | Starch (Cereals Starchy root) (g/capita/day 2007-09 mean,FAO | Meat +protein(g/capita/day), 2007-09 mean, FAO | Animal protein, excl meat protein (g/capita/day) 2007-09 mean,FAO | Plant protein (g/capita/day) 2007-09 mean, FAO | Animal Fat Total (g/capita/day), 2007-09 mean, FAO | Plant oil total (g/capita/day),2007-09 mean, FAO | Carbohydrate (g/capita/day) 2007-09 mean, self-calculated | Calories Total (kcal/capita/day) 2007-09 mean, FAO | Minimum Dietary Energy Requirement (kcal/person/day) 2006-08, FAO | Physicak Inactivity 2010 %, 18+,WHO | GDP ppp (current US$), the World Bank, 2010 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Turkmenistan | 17.6 | 50.7 | 25.8 | 158.9 | 28.0 | 165.8 | 419.7 | 657.0 | 22.2 | 17.1 | 53.8 | 51.6 | 29.2 | 444.6 | 2877.7 | 1880 | | 9828.7 |
| Uganda | 3.7 | 12.0 | 21.7 | 34.3 | 22.4 | 431.9 | 112.8 | 664.2 | 4.6 | 7.7 | 37.9 | 12.1 | 35.2 | 408.2 | 2259.7 | 1710 | | 1484.6 |
| Ukraine | 18.7 | 53.3 | 25.9 | 131.9 | 51.9 | 124.3 | 386.5 | 761.7 | 16.4 | 25.8 | 45.7 | 46.7 | 47.7 | 504.6 | 3219.3 | 1950 | 12.2 | 7685.6 |
| United Arab Emirates | 34.5 | 71.0 | 28.6 | 205.9 | 41.0 | 255.3 | 297.8 | 482.1 | 26.7 | 22.3 | 55.9 | 43.3 | 45.6 | 484.7 | 3159.7 | 1970 | 38.4 | 55498.1 |
| United Kingdom | 25.5 | 65.9 | 27.0 | 226.0 | 66.0 | 352.7 | 259.9 | 602.4 | 29.0 | 29.6 | 44.9 | 75.1 | 65.2 | 435.7 | 3419.0 | 1940 | 37.3 | 35920.0 |
| Tanzania | 5.6 | 15.3 | 22.8 | 23.6 | 22.3 | 215.8 | 155.7 | 705.5 | 3.3 | 6.3 | 45.5 | 8.4 | 34.2 | 389.1 | 2160.0 | 1730 | 6.9 | 2099.5 |
| United States | 31.2 | 70.3 | 28.5 | 337.3 | 93.0 | 305.5 | 338.9 | 451.6 | 40.8 | 32.1 | 40.4 | 69.7 | 92.3 | 449.6 | 3709.7 | 1980 | 32.4 | 48374.1 |
| Uruguay | 23.8 | 58.9 | 26.4 | 147.0 | 33.8 | 213.0 | 158.7 | 607.0 | 17.2 | 25.5 | 41.8 | 46.7 | 35.6 | 450.4 | 2879.7 | 1880 | 31.7 | 16748.9 |
| Uzbekistan | 13.6 | 43.8 | 25.6 | 73.9 | 32.8 | 125.6 | 567.8 | 616.8 | 10.6 | 15.6 | 50.0 | 30.6 | 36.5 | 417.9 | 2579.7 | 1890 | 19.2 | 4100.5 |
| Vanuatu | 33.3 | 61.2 | 26.0 | 101.5 | 12.2 | 259.3 | 142.4 | 842.9 | 12.9 | 15.7 | 39.9 | 34.3 | 66.7 | 419.3 | 2861.0 | 1730 | 8.4 | 2888.8 |
| Venezuela | 23.3 | 58.6 | 27.0 | 231.5 | 51.2 | 204.5 | 167.3 | 489.3 | 29.4 | 15.3 | 37.6 | 34.6 | 51.4 | 433.2 | 2836.3 | 1840 | | 16201.8 |
| Viet Nam | 2.6 | 14.2 | 21.1 | 126.6 | 14.2 | 180.1 | 233.5 | 500.2 | 14.4 | 12.3 | 44.4 | 41.8 | 23.1 | 432.9 | 2600.3 | 1820 | 23.9 | 4395.5 |
| Yemen | 15.7 | 38.0 | 25.5 | 48.6 | 23.4 | 150.2 | 111.7 | 462.0 | 6.6 | 6.0 | 44.4 | 11.0 | 34.4 | 378.2 | 2150.0 | 1700 | | 4442.5 |
| Zambia | 7.1 | 17.8 | 22.3 | 35.6 | 13.7 | 28.7 | 75.6 | 615.2 | 5.1 | 4.2 | 38.5 | 7.2 | 29.2 | 325.4 | 1820.0 | 1720 | 20.5 | 3381.0 |
| Zimbabwe | 9.7 | 15.5 | 23.3 | 53.2 | 33.5 | 39.0 | 53.3 | 470.0 | 7.4 | 4.1 | 40.3 | 13.9 | 43.6 | 348.1 | 2116.7 | 1790 | 22.4 | 1454.2 |

**Table S3 Data descriptive and summary**

| | N | Minimum | Maximum | Mean | Std. Deviation | Skewness | Std. Error | Data source |
|---|---|---|---|---|---|---|---|---|
| BMI≥30,18+,2010 | 170 | 1.7 | 41.8 | 16.670 | 8.9937 | 0.06 | 0.19 | WHO |
| BMI≥25,18+, 2010 | 170 | 9.7 | 72.8 | 43.021 | 19.4304 | -0.37 | 0.19 | WHO |
| BMI mean,18+, 2010 | 170 | 20.3 | 31.4 | 25.258 | 2.1433 | -0.26 | 0.19 | WHO |
| Meat, total 2007-09 mean (g/capita/day) | 167 | 10.73 | 337.26 | 129.57 | 82.53 | 0.42 | 0.19 | FAO |
| Fat (plant oil+ animal fat) (g/capita/day) 2007-09 | 161 | 4.63 | 137.87 | 42.55 | 26.83 | 1.10 | 0.19 | FAO |
| Fruits (g/capita/day) 2007-09 mean | 167 | 15.26 | 964.69 | 229.70 | 142.71 | 1.39 | 0.19 | FAO |
| Fiber (Vgetables + Pulses) (g/capita/day), 2007-09 mean | 167 | 35.59 | 873.12 | 251.97 | 164.61 | 1.32 | 0.19 | FAO |
| Starch (Cereals + Starchy root) (g/capita/day) 2007-09 mean | 167 | 219.53 | 1345.32 | 573.11 | 176.81 | 1.23 | 0.19 | FAO |
| Meat protein(g/capita/day), 2007-09 mean | 167 | 1.53 | 40.83 | 16.37 | 10.21 | 0.49 | 0.19 | FAO |
| Animal protein, excl meat protein (g/capita/day) 2007-09 mean | 167 | 1.87 | 65.30 | 18.94 | 11.71 | 0.90 | 0.19 | FAO |
| Plant protein (g/capita/day) 2007-09 mean | 167 | 23.88 | 77.56 | 43.50 | 9.08 | 0.85 | 0.19 | FAO |
| Animal Fat Total (g/capita/day), 2007-09 mean | 167 | 4.58 | 105.35 | 37.45 | 24.94 | 0.62 | 0.19 | FAO |
| Plant oil total (g/capita/day),2007-09 mean | 167 | 15.39 | 98.75 | 45.99 | 16.68 | 0.69 | 0.19 | FAO |
| Carbohydrate (g/capita/day) 2007-09 mean, self-calculated | 167 | 303.47 | 627.64 | 431.23 | 55.46 | 0.29 | 0.19 | FAO |
| Calories Total (kcal/capita/day) 2007-09 mean | 167 | 1820.00 | 3791.67 | 2793.59 | 462.09 | 0.13 | 0.19 | FAO |
| Minimum Dietary Energy Requirement (kcal/person/day) 2006-08 | 167 | 1690 | 2000 | 1852.40 | 87.655 | -0.05 | 0.19 | FAO |
| Prevalence of physicak inactivity %, 18+, 2010 | 131 | 4.1 | 63.6 | 25.302 | 11.4495 | 0.73 | 0.21 | WHO |
| GDP PPP (current US$), 2010 | 165 | 619.47 | 84200.57 | 15224.06 | 15714.26 | 1.70 | 0.19 | World Bank |
| Valid N (listwise) | 123 | | | | | | | |

**Table S4: Countries included in Latin America in our study.**

Based on region primarily speaking romance languages, Latin America contains the following countries.

Argentina, Bolivia (Plurinational State of), Brazil, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Haiti, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Uruguay, Venezuela (Bolivarian Republic of)

Article 5/10: Meat in modern diet, just as bad as sugar, correlates to worldwide obesity: an ecological analysis (Published at Journal of Nutrition & Food Sciences 2016)

Wenpeng You[1]*, Maciej Henneberg[1,2]

[1] Biological Anthropology and Comparative Anatomy Unit, School of Medicine, the University of Adelaide, Adelaide, SA, Australia 5000

[2] Institute of Evolutionary Medicine, University of Zürich, Switzerland

**Published:** Wenpeng You, Maciej Henneberg. *Meat in modern diet, just as bad as sugar, correlates to worldwide obesity: an ecological analysis*, Journal of Nutrition & Food Sciences, 2016, 6 (4), 1-10.

✉ **Correspondence:** Wenpeng You Wenpeng.you@adelaide.edu.au

## Contexture Statement

Sugar is well-established predictor of obesity worldwide because it may cause metabolic syndrome and it may provide energy surplus to human body. Our previous study suggested that meat protein may have the same mechanism to contribute to body weight increase, thus obesity.

In this study, we collected the empirical nutrient data at the country level to assess and compare the correlation levels of obesity prevalence to sugar and meat consumption respectively.

## Statement of Authorship

# Statement of Authorship

| | |
|---|---|
| Title of Paper | Meat in modern diet, just as bad as sugar, correlates to worldwide obesity: an ecological analysis |
| Publication Status | ☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Publication Style |
| Publication Details | Wenpeng You, Maciej Henneberg. Meat in modern diet, just as bad as sugar, correlates to worldwide obesity: an ecological analysis, Journal of Nutrition & Food Sciences, 6 (4), 1-10. |

### Principal Author

| | | | |
|---|---|---|---|
| Name of Principal Author (Candidate) | Wenpeng You | | |
| Contribution to the Paper | Conceived the hypothesis, collected data, performed analysis on all samples, interpreted data, wrote manuscript and acted as corresponding author. | | |
| Overall percentage (%) | 60 | | |
| Signature | | Date | 22/12/2017 |

### Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

| | | | |
|---|---|---|---|
| Name of Co-Author | Maciej Henneberg | | |
| Contribution to the Paper | Supervised development of work, formalized and interpreted data analysis, and evaluated and edited manuscript | | |
| Signature | | Date | 22/12/2017 |

| | | | |
|---|---|---|---|
| Name of Co-Author | NA | | |
| Contribution to the Paper | | | |
| Signature | | Date | |

Please cut and paste additional co-author panels here as required.

## Abstract

**Background:** The public have been educated that sugar intake should be minimized to avoid obesity, but no such recommendation regarding meat exists. We used FAO published comparable sugar and meat availability data to examine if they both contribute to obesity prevalence to the same extent.

**Methods:** Country-specific Body Mass Index (BMI) estimates of obesity and overweight were obtained. These were matched with country-specific per capita per day availability of major food groups (meat, sugar, starch crops, fibers, fats and fruits), total calories, per capita Gross Domestic Product (GDP PPP), urbanization and physical inactivity prevalence. Fisher's r-to-z transformation and Beta (B) range (B ± 2 Standard Error) overlapping were used to test for potential differences between correlations and regressions results respectively. SPSS 22.0 was used for log-transformed data analysis.

**Results:** Pearson correlation showed that sugar and meat availability significantly correlated with obesity prevalence to the same extent ($r=0.715$, $p<0.001$ and $r=0.685$, $p<0.001$ respectively). These relationships remained in partial correlation analysis ($r=0.359$, $p<0.001$ and $r=0.354$, $p<0.001$ respectively) when controlling for calories availability, physical inactivity, urbanization and GDP PPP. Fisher's r-to-z transformation revealed no significant difference in Pearson correlation coefficients ($z=-0.53$, $p=0.60$), partial correlation coefficients ($z=-0.04$, $p=0.97$) between sugar and meat availability with obesity prevalence.

Multiple linear regression analysis indicated that sugar and meat availability were the two most significant predictors of obesity prevalence in both Enter ($B=0.455$, $SE=0.113$, $p<0.001$ and $B=0.381$, $SE=0.096$, $p<0.001$, respectively) and Stepwise ($B=0.464$, $SE=0.093$, $p<0.001$ and $B=0.433$, $SE=0.072$, $p<0.001$, respectively) models. B ranges overlapping found in the Enter (0.289-0.573) and Stepwise (0.294-0.582) models showed sugar and meat availability correlated to obesity with no statistically significant difference.

**Conclusion:** Sugar and meat availability comparably contribute to global obesity prevalence. Dietary guidelines should also advocate to minimize meat consumption to avoid obesity.

**Keywords:** Obesity, Sugar, Fructose, Meat, Meat protein, Fats, Insulin resistance, Energy surplus

## Introduction

Obesity has been considered a major epidemic of the 21st century, and it has become a prelude to adverse health and premature death (1). The World Health Organisation (WHO) estimates that obesity contributes significantly to the disease burdens of, among others, top causes of diseases, such as diabetes (44%), ischaemic heart disease (23%) and carcinogenesis (7-41%) (2). Moreover, those considered overweight or obese have been subject to discrimination and prejudice (3).

Obesity and overweight are defined as abnormal or excessive fat accumulation that presents a risk to health. A crude population measure of obesity is the body mass index (BMI). A person with a BMI of 30 kg/m$^2$ or more is generally considered obese. A person with a BMI equal to or more than 25 kg/m$^2$ is considered overweight (http://who.int/topics/obesity/en/).

Until the invention of food production in the Holocene for several million years, human diet had relied on foods that could be found in natural environments. Since humans are unable to extract nutrition from cellulose, our food sources were limited to animals, fruits, nuts and tubers. This Palaeolithic diet contained large quantities of meat obtained through hunting (4) while it had less carbohydrates, especially simple carbohydrates. Besides large game that was hunted or scavenged, small vertebrates and invertebrates were gathered, and, where possible, fish were caught. Game meat does not contain much fats, so our metabolic system evolved to be efficient in using animal protein as a source of energy (5). Deriving Acetyl CoA for use in the citric acid cycle from proteins is a complex process using a number of enzymes to obtain peptides, break them into separate aminoacids and then deaminate those amino acids to obtain carbon skeletons – a source of pyruvates. Any pyruvates not used in the citric acid cycle to obtain energy can be converted via de novo lipogenesis into fats and stored. Obtaining pyruvates from carbohydrates is a simpler metabolic process, especially when simple carbohydrates that are easily breakable into glucose are consumed. Therefore, when simple carbohydrates are available in the diet they, soon after their ingestion and absorption, can be used to provide energy in the citric acid cycle while additional pyruvates coming later from protein digestion may be surplus to direct need for energy, and therefore converted into fat (6). Sucrose, being a compound of glucose and fructose provides easily accessible energy from glucose while fructose is not easily digested. Since the introduction of agriculture, and especially from the time of industrialised food production, sucrose became readily available in large quantities. In traditional agricultural economies meat was expensive to produce and thus was consumed in small quantities, rather rarely. Mass animal husbandry lowered cost of meat production and now meat is readily available and regularly

consumed in significant quantities in developed economics.Diet patterns have been extensively considered as the contributing factor to obesity. Sugar and meat are now two major food groups in our daily diet. The prevailing dogma is that we should limit or avoid sugar intake, and eat a moderate amount of meat, preferably lean meat since it is a source of essential aminoacids. This dogma is supported by various dietary or nutrition guidelines published by the authorities. Numerous studies have reported that meat (7) and added sugar (sugar in short hereafter) (8) food groups were in significant correlations to obesity and/or body weight increase. However, the majority of the studies could not single out total sugar or all meat consumption in our diet for the correlation analysis. One of the concrete evidences that sugar consumption was correlated to obesity is that sugar-sweetened beverages (SSBs) intake is associated with obesity prevalence (9). Despite the correlations between sugar food and beverage products and obesity are controversial, SSBs has been mostly consistently correlated to obesity prevalence (8). However, sugar consumed via beverage is only part of the dietary sugar intake, and other sugar products, such as confectionery and bakery products were not included in the study designs. Similarly, meat containing food groups rather than pure meat are considered for example processed meat (10) instead of total meat intake (11) have been linked to obesity prevalence. Another issue with meat food group is that data used for study may not be able to exclude bias from other food components, which may have been linked to body weight increase. For instance, wheat consumption has been correlated to obesity (12, 13), and meat food groups containing wheat products (frankfurter and sausage) could be associated with obesity and central obesity (11) because of their wheat content. Likewise, the correlation between SSBs and obesity prevalence may be biased with other obesity associated additives in SSBs, such as preservatives. Therefore, these research results may not present the whole picture of the correlation between obesity prevalence and sugar or meat consumption. Using these data may not allow us to explore and compare the correlations between obesity prevalence and total intake of sugar and meat accurately.

The Food and Agriculture Organization (FAO) Food Balance Sheet presents the comparative per capita availability of major food items during the reference period by country after combining sources of supply and its utilization in terms of nutrient value. This study aimed to use empirical, macro-level nutrient availability data at the country level to evaluate and compare, from a global perspective, the correlation levels of obesity prevalence to sugar and meat availability.

## Materials and Methods

### Data

The country specific data were collected for this study:

1) The WHO Global Health Observatory (GHO) data on estimated prevalence rates of obesity and overweight (percent of population aged 18+ with BMI ≥ 30 and 25 kg/m$^2$ respectively) of the population aged 18+ by country were obtained for the year 2010 (http://www.who.int/gho/database/en/). We did not use the most recent version of body weight status in 2014 because of other key variables of interest (described below).

From GHO, we also captured the estimated prevalence rate of physical inactivity for each country for the people aged 18+. The estimated prevalence rate of physical inactivity is defined as percent of a given population attaining less than 150 minutes of moderate-intensity physical activity per week, or less than 75 minutes of vigorous-intensity physical activity per week, or equivalent.

2) The FAOSTAT Food Balance Sheet (FBS) data on major food group availability per capita per day of: i) sugar (total sugar & sweeteners); ii) total meat; iii) starch crops (mixed cereals and starchy root); iv) fibers (vegetables, treenuts and pulses); v) fats (plant oils and animal fats) and vi) fruits. We also extracted the per capita per day availability of grand total calories (calories in shorted hereafter) as one of the potential confounders of our data analysis. Unfortunately, FAOSTAT does not contain data allowing separation of processed meats from “pure meat”.

Because obesity develops after cumulative exposure to dietary risks (i.e. high intake of risk food groups today does not lead to immediate obesity, but a prolonged exposure to high intake of risk food type(s) is required (14-16).), we calculated the mean food availability per person per day over a 3-year period (2007-2009) in each of food categories to represent typical long-term exposure to each of these dietary components. The rationale for this decision is that studies have shown that three years is a practical period to develop metabolic syndrome leading to obesity after exposure to dietary risks. For instance, high intake of meat today does not lead to immediate obesity. Using the mean of three years of nutrients and food groups may also reduce the random errors during the data collection and calculation by FAO.

The food items in each food groups were listed the Additional file 1: Food items in each food group.

3) The World Bank data on per capita GDP PPP (expressed in gross domestic product converted to international dollars using purchasing power parity rates) and country specific urbanization (the percent of population living in urban areas. Urbanization has been closely linked to human lifestyle change due to its process of modernization and industrialization.

How the above variables, such as food types (nutrients) and BMI were data collected and how they lead to their robustness and to the subsequent validity of the current analysis have been described in details elsewhere (17, 18).

WHO, FAO and the World Bank are intergovernmental organizations using specialized information relevant to their respective fields. Their professional personnel should have evaluated these data in consideration of their possible use, e.g. for scientific research and decision making, before they were published. Therefore, the data reporting is as free of bias and error as it can be with government statistics. This means that errors are reduced but some inaccuracies related to reporting quality may still be present in the data. Similar data from the same sources were recently used to analyse the relationships between nutrients and obesity (18, 19) and diabetes (20-22) in a number of publications.

We obtained data for 170 countries after we matched the prevalence estimates of obesity and overweight to the year-and country-specific food groups and other variables. Each country was treated individually as the subject and all their availability for other variables information was analysed. Data sources and summary statistics were further described in Additional file 2 Data descriptive summary and source.

For particular analyses, the number of countries included for variables may have differed somewhat because all information on other variables was not uniformly available for all countries due to unavailability from relevant UN agencies. All the data were extracted and saved in Microsoft Excel® for analysis.

**Statistical analyses**

It has been commonly believed that obesity is an affluence related medical condition (23), which is generally caused by eating too much (dietary) (24) and moving too little (lifestyle) (25). Urbanization is a population shift from rural to urban areas. It causes changes in diet and exercise patterns of the population (26). Therefore, in addition to the seven dietary predictors (availability of sugar, meat, fats, fruits, fibers, starch and calories), we also incorporated GDP PPP, urbanization and prevalence of physical inactivity for data analysis.

To assess the difference between relationships between obesity prevalence and availability of sugar and meat, the analysis proceeded in four steps.

1. Pearson correlation was used to evaluate the strength and direction of the associations between all variables.

2. Partial correlation of Pearson moment-product approach was used to find the relationship between obesity prevalence and each food group respectively while keeping calories availability, GDP PPP, physical inactivity and urbanization statistically constant. In order to show that meat and sugar availability contributed to obesity prevalence independent of each other, we controlled for availability of the other food groups (starch crops, fibers, fats and fruits) in addition to GDP PPP, urbanization, total calories availability, physical inactivity.

Fisher's r-to-z transformation was performed to test significance of differences between correlation coefficients. The significance was reported when P-value was <0.01.

We kept sugar and meat availability statistically constant respectively together with all the other variables to test if they were correlated to obesity prevalence significantly independent of each other in addition to all other variables.

3. Standard multiple linear regression (Enter) was conducted to describe the relationships between obesity prevalence and all independent variables, which include all the dietary, lifestyle and socioeconomic predictors.

Standard multiple linear regression (Stepwise) was also performed to regress multiple variables while simultaneously retaining sugar and meat availability as the important predictors of obesity prevalence.

Analysis results of multiple linear regression (Enter and Stepwise) model included both the indicative value of beta coefficient (B) and its standard error (SE). The actual B may fall into a range determined with its standard error. Therefore, we added twice the standard error (SE) to their respective B to obtain the upper bound of the range and subtracted two SEs from B to obtain the lower bound of the range. We compared the ranges of B's of obesity prevalence to sugar and meat availability to determine if the relationships were significantly different. If two B ranges have overlap, the difference between the B's would not be considered as significant. If there is no overlap, the difference would be considered significant.

4. We used scatter plots to explore and visualize the correlations between obesity and availability of sugar and meat. To compare the two relationships, we reversed x and y axes to allow the two correlations in one figure (chart).

### Additional variables

We reassessed our models using overweight (BMI ≥ 25 kg/m$^2$ instead of obesity (BMI ≥ 30 kg/m$^2$) in case of sugar and meat availability as a late-stage predictor of obesity. The results were reported in tables aligning with those relationships between obesity prevalence and sugar and meat availability. To incorporate overweight data for analysis may allow us to reassure the quality of data which were used for this study.

SPSS v. 22 (SPSS Inc., Chicago Il USA) was used for data analysis. Prior to analysis data were log-transformed (natural logarithms) to bring their distributions close to normal.

## Results

Pearson correlation analysis showed that both sugar and meat availability were significantly correlated with prevalence of obesity ($r=0.715$, $p<0.001$ and $r=0.685$, $p<0.001$, respectively) (Table 1). Spearman rho values were $r= 0.664$ ($p<0.001$) and $r=0.664$ ($p<0.001$) respectively. Fisher's r-to-z transformation revealed no significant difference in Pearson correlations between sugar and meat availability with obesity ($z=0.53$, $p=0.5961$). The difference between two coefficients' values was negligible, indicating that both meat and sugar were related to obesity to the same extent.

When we controlled for availability of total calories, prevalence of physical inactivity, urbanization and GDP PPP in partial correlation analysis, sugar and meat availability were still in significant correlation with prevalence of obesity ($r=0.359$, $p<0.001$ and $r=0.354$, $p<0.001$, respectively) (Table 2). This indicates that it is not just the contribution of sugar and meat to the total caloric intake that relates to obesity, but specific contents of these two food groups that influence metabolic processes. Fisher's r-to-z transformation revealed no significant difference in partial correlations between sugar and meat availability with obesity prevalence based on the comparison of two correlations ($z=0.04$, $p=0.9681$). This means that both sugar and meat availability contributes to obesity to the same extent.

Table 1 Pearson correlation matrix for all variables

| | BMI 30 | BMI 25 | Sugar | Meat | Fats | Fruits | Fibers | Starch crop | Calories | GDP | Urbanization | Physical Inactivity |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BMI 30 | 1.000 | 0.931*** | 0.715*** | 0.685*** | 0.523*** | 0.477*** | 0.678*** | -0.220** | 0.619*** | 0.678*** | 0.497*** | 0.448** |
| BMI 25 | | 1.000 | 0.776*** | 0.792*** | 0.644*** | 0.546*** | 0.806*** | -0.290** | 0.748*** | 0.798*** | 0.632*** | 0.458*** |
| Sugar | | | 1.000 | 0.718*** | 0.571*** | 0.470*** | 0.714*** | -0.492*** | 0.650*** | 0.727*** | 0.529*** | 0.437*** |
| Meat | | | | 1.000 | 0.614*** | 0.520*** | 0.826*** | -0.431*** | 0.695*** | 0.831*** | 0.565*** | 0.406*** |
| Fats | | | | | 1.000 | 0.373*** | 0.696*** | -0.223** | 0.701*** | 0.684*** | 0.651*** | 0.300** |
| Fruits | | | | | | 1.000 | 0.565*** | -0.215** | 0.499*** | 0.560*** | 0.353*** | 0.230** |
| Fibers | | | | | | | 1.000 | -0.370*** | 0.779*** | 0.994*** | 0.625*** | 0.439*** |
| Starch crop | | | | | | | | 1.000 | -0.029 | -0.394*** | -0.150* | -0.425*** |
| Calories | | | | | | | | | 1.000 | 0.763** | 0.643** | 0.243** |
| GDP PPP | | | | | | | | | | 1.000 | 0.620*** | 0.437*** |
| Urbanization | | | | | | | | | | | 1.000 | 0.385*** |
| Physical Inactivity | | | | | | | | | | | | 1.000 |

Number of countries included in the analysis range from 126 to 170. * p<0.05; ** p< 0.01, ***P< 0.001

BMI≥30 and BMI ≥ 25 are percentages of defined population with a body mass index (BMI) of no less than 30 kg/m2 and 25 kg/m2 respectively.

Data sources: Dietary data from the FAO's FAOSTAT. BMI (≥30 and ≥25) and Physical Inactivity data from the WHO Global Health Observatory. GDP PPP and urbanization data from the World Bank.

Table 2 also presented that fats availability was in strong and significant correlation (Pearson) with obesity prevalence ($r=0.517$, $p<0.001$), but the level of correlation was not retained in partial correlation analysis ($r=0.057$, $p=0.537$). Starch crops availability was in relative strong correlation with obesity prevalence, but this correlation almost disappeared in partial correlation analysis.

When we controlled for availability of fats, fruits, fibers and starch, prevalence of physical inactivity, total calories, urbanization and GDP PPP in partial correlation analysis, both sugar and meat availability were still in significant correlation with prevalence of obesity ($r=0.431$, $p< 0.001$ and $r=0.339$, $p< 0.001$, respectively) (Table 2). Fisher's r-to-z transformation did not show a significant difference in the correlations between obesity and sugar and meat availability ($z=0.81$, $p=0.4179$). Therefore, sugar and meat contributions to obesity are independent of the availability of other food groups.

Interestingly, meat and sugar availability significantly correlated with each other in Pearson correlation ($r=0.718$, $p<0.001$) analysis (Table 1) but this correlation disappeared in partial correlation analysis when we controlled for availability of fats, fruits, fibers and starch crops, calories, GDP, urbanization and physical inactivity prevalence. Partial correlation coefficient became very weak and insignificant ($r=0.144$, $p=0.124$, not indicated in Table 2). This means that sugar and meat availability may contribute to obesity prevalence independent of each other.

The further investigation on this independence showed that both sugar ($r=0.375$, $p<0.001$) and meat ($r=0.308$, $p<0.001$) availability were still significantly correlated to obesity prevalence when we respectively controlled for sugar and meat availability together with all the other variables (fats, fruits, fibers and starch crops, calories, GDP, urbanization and physical inactivity prevalence) for testing each other’s relationship with obesity prevalence (Table 2). Fisher's r-to-z transformation did not show significant difference between these two independent relationships ($z=0.57$, $p=0.2843$).

Table 2 Pearson and partial correlation analysis of different food groups to prevalence estimates of obesity and overweight

| | Pearson correlation | | | Partial Correlation | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | n | BMI≥ 30 | BMI≥ 25 | df | BMI≥30 | BMI≥25 | df | BMI≥ 30 | BMI≥ 25 | df | BMI 30 | BMI 25 | df | BMI≥30 | BMI≥25 |
| Sugar | 167 | 0.715*** | 0.776*** | 118 | 0.359*** | 0.372*** | 114 | 0.431*** | 0.399*** | 114 | 0.375*** | 0.363*** | - | - | - |
| Meat | 167 | 0.685*** | 0.792*** | 118 | 0.354*** | 0.418*** | 114 | 0.339*** | 0.370*** | - | - | - | 114 | 0.308*** | 0.341*** |
| Fats | 161 | 0.523*** | 0.644*** | 118 | 0.057 | 0.110 | - | - | - | - | - | - | - | - | - |
| Fruits | 167 | 0.477*** | 0.546*** | 118 | 0.112 | 0.159 | - | - | - | - | - | - | - | - | - |
| Fibers | 169 | 0.678*** | 0.806*** | 118 | 0.248** | 0.269** | - | - | - | - | - | - | - | - | - |
| Starch crops | 167 | -0.220** | -0.290** | 118 | 0.070 | -0.036 | - | - | - | - | - | - | - | - | - |
| Calories | 167 | 0.619*** | 0.748*** | - | - | - | - | - | - | - | - | - | - | - | - |
| GDP PPP | 165 | 0.678*** | 0.798** | - | - | - | - | - | - | - | - | - | - | - | - |
| Urbanization | 169 | 0.497*** | 0.632*** | - | - | - | - | - | - | - | - | - | - | - | - |
| Physical Inactivity | 131 | 0.448*** | 0.458*** | - | - | - | - | - | - | - | - | - | - | - | - |

* p<0.05, ** P< 0.01, ***P< 0.001. -, controlled variable.

BMI≥30 and BMI ≥ 25 are percentages of defined population with a body mass index (BMI) of no less than 30 kg/m2 and 25 kg/m2 respectively.

Data sources: Dietary data from the FAO's FAOSTAT. BMI (≥30 and ≥25) and Physical Inactivity data from the WHO Global Health Observatory. GDP and urbanization data from the World Bank.

Table 3 presented the results of multiple linear regression analyses to identify dietary, lifestyle and socioeconomic predictors of prevalence estimates of obesity and overweight. We found that both sugar and meat were the significant predicators of estimates of obesity (B=0.455, SE=0.113 and B=0.381, SE=0.096, respectively) at the same significance level of $p<0.001$. The B ranges between obesity prevalence and availability of sugar (0.229-0.681) and meat (0.189-0.573) overlapped each other greatly (0.229-0.573). This meant that meat availability was no different from sugar availability to predict the estimates of prevalence of obesity.

Table 3 Results of enter multiple linear regression analyses to identify dietary, lifestyle and socioeconomic predictors of prevalence estimates of overweight and obesity

| | Obesity prevalence (%) | | | | Overweight prevalence (%) | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 1 (Enter), R2= 0.656 | | | | Model 1 (Enter), R2=0.823 | | | |
| Predictors | B | SE | p | B range | B | SE | p | B range |
| Sugar | 0.455 | 0.113 | <0.001 | 0.229-0.681 | 0.315 | 0.066 | <0.001 | 0.183-0.447 |
| Meat | 0.381 | 0.096 | 0.001 | 0.189-0.573 | 0.307 | 0.056 | <0.001 | 0.195-0.419 |
| Fats | 0.053 | 0.095 | 0.565 | - | 0.056 | 0.055 | 0.391 | - |
| Fruits | 0.034 | 0.070 | 0.633 | - | 0.054 | 0.041 | 0.290 | - |
| Fibers | -0.170 | 0.314 | 0.777 | - | 0.214 | 0.182 | 0.618 | - |
| Starch crops | 0.349 | 0.215 | <0.001 | - | 0.164 | 0.124 | 0.008 | - |
| GDP PPP | 0.370 | 0.272 | 0.525 | - | 0.002 | 0.157 | 0.997 | - |
| Urbanization | -0.040 | 0.119 | 0.635 | - | 0.054 | 0.069 | 0.368 | - |
| Physical Inactivity | 0.163 | 0.098 | 0.015 | - | 0.081 | 0.057 | 0.090 | - |
| Calories | -0.147 | 0.544 | 0.233 | - | 0.069 | 0.315 | 0.434 | - |

B, Beta; SE, Std. Error; p, Sig.

BMI≥30 and BMI ≥ 25 are percentages of defined population with a body mass index (BMI) of no less than 30 kg/m2 and 25 kg/m2 respectively.

Data sources: Dietary data from the FAO's FAOSTAT. BMI (≥30 and≥25) and Physical Inactivity data from the WHO Global Health Observatory. GDP and urbanization data from the World Bank.

Table 4 indicated that sugar (B=0.464, SE=0.093, p<0.001) and meat (B=0.433, SE=0.072, p<0.001) availability stood out as the significant predictor of obesity prevalence simultaneously in stepwise multiple linear regression analyses. Two overlapping B ranges (0.294-0.582) indicated that there was no difference between sugar and meat availability to predict obesity prevalence.

One of the highlights in our data analysis with linear regression was that fats availability was a minor predictor of obesity prevalence in both Enter method (B=0.053, SE=0.095, p=0.565) (Table 3) and Stepwise method (fats availability was a removed variable) (Table 4).

Table 4 Results of stepwise multiple linear regression analyses to identify dietary, lifestyle and socioeconomic predictors of prevalence estimates of overweight and obesity

| | Obesity prevalence (%) | | | | Overweight prevalence (%) | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 4 (Stepwise), Adjusted $R^2$= 0.630 | | | | Model 4(Stepwise), Adjusted $R^2$=0.802 | | | |
| Predictors | B | SE | p | B Range | B | SE | p | B Range |
| Sugar | 0.464 | 0.093 | <0.001 | 0.278-0.650 | 0.363 | 0.059 | <0.001 | 0.245-0.481 |
| Meat | 0.438 | 0.072 | <0.001 | 0.294-0.582 | 0.340 | 0.055 | <0.001 | 0.230-0.450 |
| Fats | - | - | - | | - | - | - | - |
| Fruits | - | - | - | | - | - | - | - |
| Fibers | - | - | - | | 0.359 | 0.034 | <0.001 | - |
| Starch crops | 0.464 | 0.171 | <0.001 | | 0.187 | 0.097 | <0.001 | - |
| GDP PPP | - | - | - | | - | - | - | - |
| Urbanization | - | - | - | | - | - | - | - |
| Physical Inactivity | 0.171 | 0.094 | 0.008 | | - | - | - | - |
| Calories | - | - | - | | - | - | - | - |

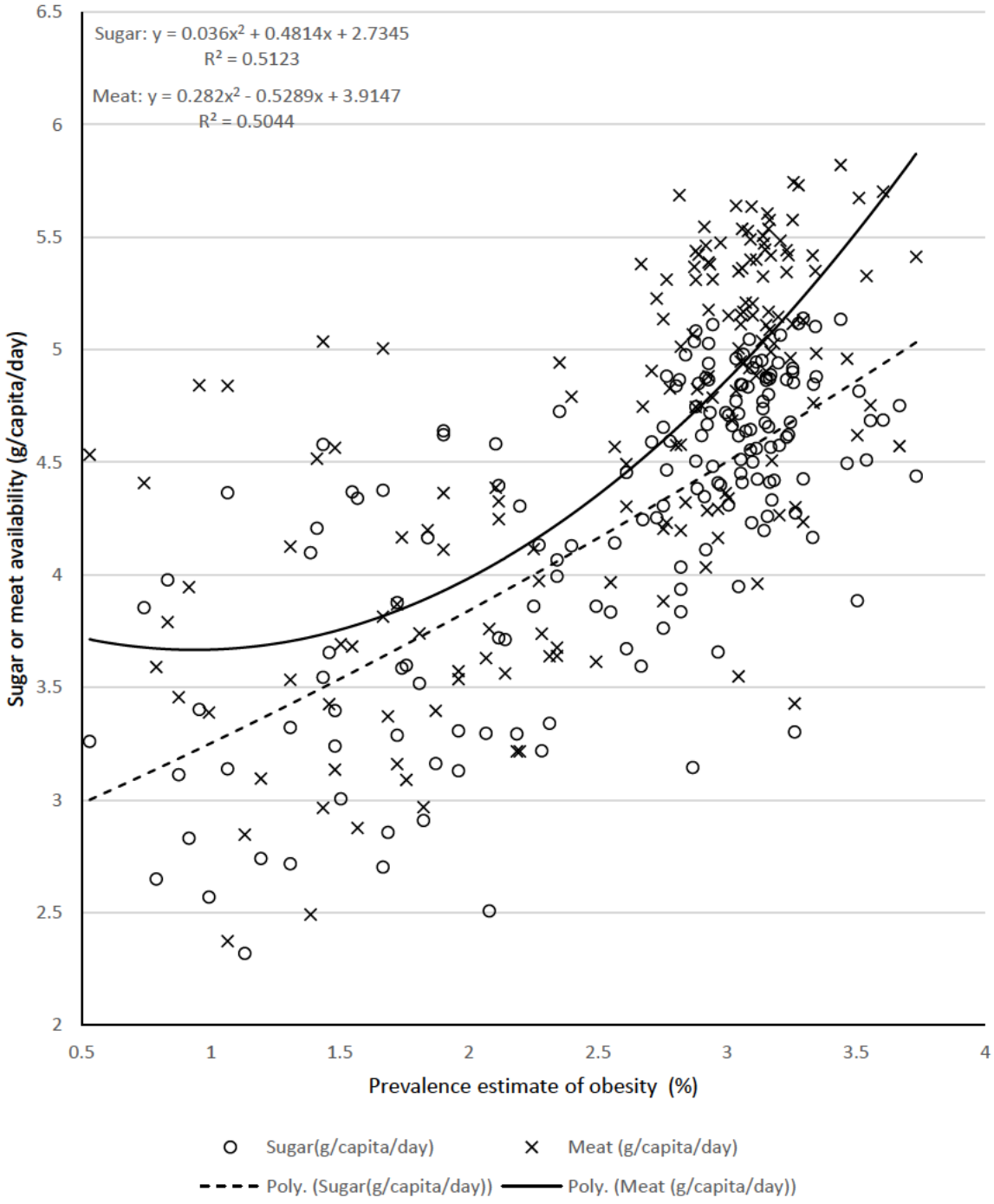
B, Beta; SE, Std. Error; p, Sig.; -, removed variable

BMI≥30 and BMI ≥ 25 are percentages of defined population with a body mass index (BMI) of no less than 30 kg/m2 and 25 kg/m2 respectively.

Data sources: Dietary data from the FAO's FAOSTAT. BMI (≥30 and ≥25 mean) and Physical Inactivity data from the WHO Global Health Observatory. GDP and urbanization data from the World Bank.

Figure 1 showed the unadjusted correlation between prevalence estimate of obesity and sugar and meat availability. The scatterplots are very similar. The relationships were noted to be best described by polynomial regression equations with strong correlation at very similar levels.

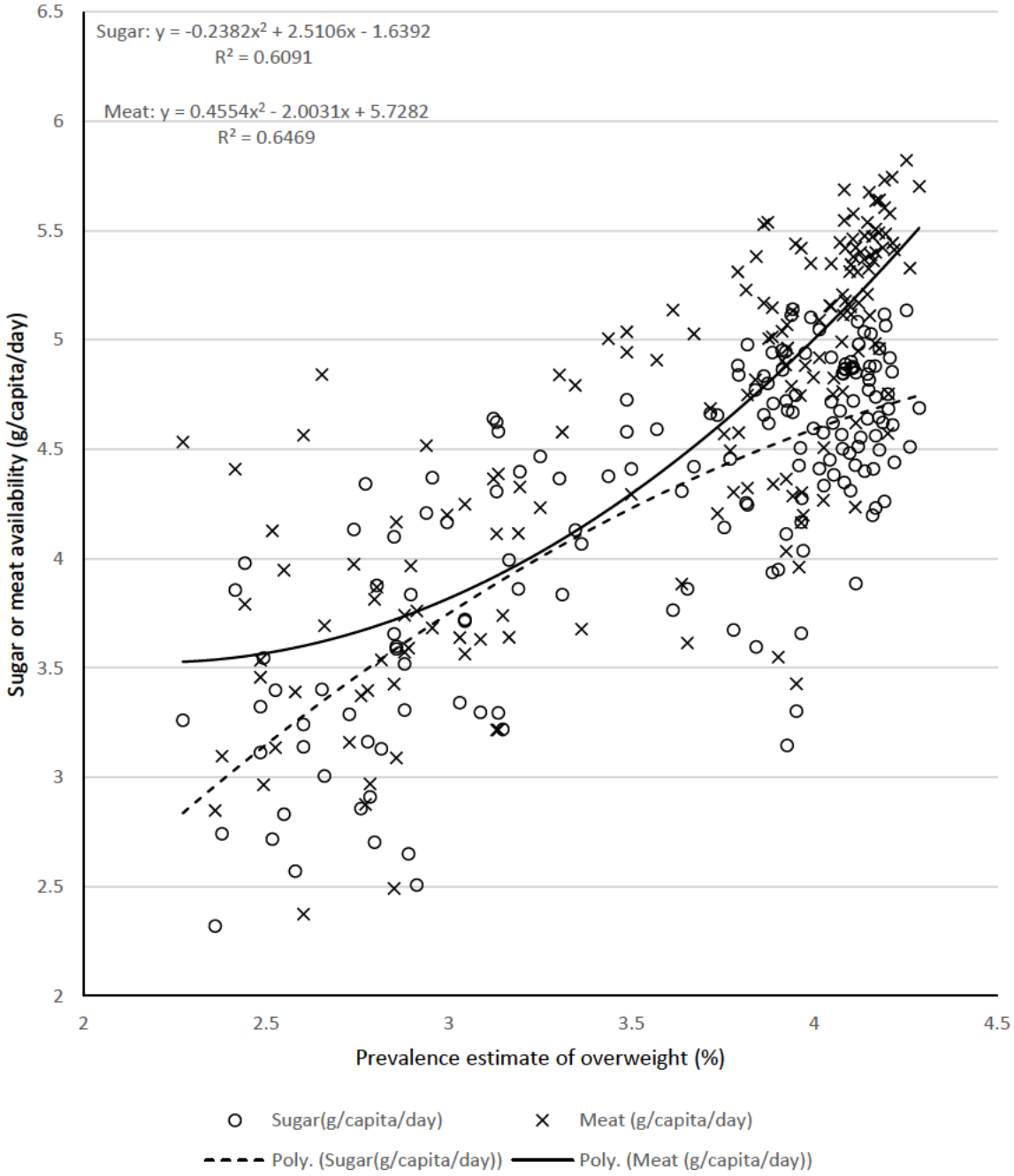
Figure 1 Polynomial correlation plot of obesity prevalence and sugar and meat availability




Our data analysis in different models showed the similar relationships between overweight prevalence and respective availability of sugar and meat when we substituted overweight prevalence for obesity prevalence. We did not describe the results in

narrative form, but they were shown in Figure 2 and Tables (1-4) aligning with those with obesity prevalence.

Figure 2 Polynomial correlation plot of overweight prevalence and sugar and meat availability




## Discussion

By examining the comparable per capita availability data of the sugar and meat for 170 countries we have shown that:

1. Sugar and meat consumptions may be two significant determinants of obesity prevalence.
2. The consumption of sugar and meat have statistically significant relations to obesity independent of the effect of other major food groups, socioeconomic and lifestyle factors.
3. Availability of sugar and meat availability are correlated to obesity prevalence independent of each other.
4. Statistically, there was no significant difference between sugar and meat relationship to global obesity prevalence at a population level.

Values of Pearson correlation coefficients may be influenced by non-homoscedasticity of distribution of correlated variables. We have tried to minimize such possibility by using logarithmically transformed data. Comparison of the values of Pearson correlation coefficients with Spearman rho values shows that effects of distributions are negligible. Thus, our partial correlation analysis produced acceptable results.

There is ample research on foods and diet patterns that contribute to body weight increase. Using the similar source of data, Siervo *et al.* reported that both meat and sugar availability were correlated to global obesity prevalence (18). However, their study did not conduct in-depth investigation to compare the correlation levels of meat and sugar availability to obesity prevalence.

There are two similar mechanisms that may explain why sugar and meat availability contribute to obesity comparably.

1. Fructose and meat protein may produce energy surplus due to their slower digestion process.

Sucrose and high-fructose corn syrup (HFCS; 42% or 55% fructose) are two primarily consumed sugars, and they are very similar in their composition. HFCS is only consumed in the U.S., Canada, Japan, and some parts of Europe, while the rest of the world primarily consumes sucrose (50% fructose). Sucrose contains 50% fructose and 50% glucose. HFCS in common usage within the food industry comprises similar percentages (40-55%) of glucose and fructose, water and other carbohydrates which are readily hydrolysable polymers of glucose. Fructose, as the major component of sugar, is slow to absorb (27) and hard to assimilate and it can only be metabolized by the liver to have glycogen most of which may be converted into fat for storage (28-30).

Meat is mainly composed of protein, fat and water. The absolute energy value of meat is determined by the protein and fat content (31). Because meat of domesticated animals

contains relatively large amount of fat, a number of studies have considered meat consumption as a higher risk of obesity and waist circumference (WC) (11, 32). Studies have already shown that dietary fat may not be a major determinant of obesity (33, 34). Animal breeding and butchering techniques in modern agriculture have significantly reduced meat fat content and increased protein content in the past few decades, so dietary meat is much leaner than ever (35, 36). The macronutrient energy values are 9.0 kcal/g for fat, 4.0 kcal/g for protein and carbohydrates (37). Therefore, meat is not high in “energy and fat content” because it contains less fat and higher meat protein due to modern agriculture techniques.

Despite energy value of protein is not high (4.0 kcal/g), it has been postulated to contribute to the development of obesity because it may only be digested later than fats and carbohydrates (17, 38, 39). Modern agriculture has been bringing the cost of availability of carbohydrates rich crops, such as cereals and starchy roots, and fat (oil), such as rape and soy significantly down. Cheap carbohydrates and fats in a meal can easily supply enough energy to meet human needs. This may make the energy from slow digested protein a surplus and stored as fat (17, 38, 39). This postulation was supported by our data analysis result (Table 2) with the changes of correlations between food groups (fats and starch crops correlated to obesity and prevalence in Pearson r correlation, but not in partial correlation) and obesity and overweight in Pearson and partial correlations. In modern diet, foods rich in carbohydrates and fats have been able to provide enough energy to meet human daily energy requirements, so meat protein has been postulated to produce energy surplus, thus contributing to obesity may support our hypothesis in this study (17).

Plant protein is always mixed with fiber which makes it difficult to digest. Therefore, meat protein as the major source of digestible protein may contribute to the “energy surplus” significantly.

2. Sugar and meat consumption may cause insulin resistance, a metabolic syndrome contributing to obesity.

Insulin is an anabolic hormone in human body. It encourages the synthesis of carbohydrate, fat and protein, inhibits the production of glucose by the liver (40). Insulin also increases the storage of fat in fat cells and prevents fat cells from releasing fat for energy (41-44). The cells in insulin resistance patient become “resistant” to insulin, and sugars in blood cannot enter cells for calories production (45-50), but are metabolized into glycogen in their liver, which may be forced by insulin to metabolise into fat (28, 51) and accelerates body weight gain (52) independent of excessive energy intake (53). Insulin resistance is a major underlying cause of excess weight and obesity (54, 55).

Many studies showed that sugar consumption is linked to insulin resistance in both children and adults, especially when it is consumed in large amounts (45, 46). High-dose fructose feeding can also cause insulin resistance in normal healthy human in as little as a week (56) and it can exacerbate insulin resistance in overweight and obese people (57).

Likewise, a number of studies have associated meat consumption as a risk factor for insulin resistance because: 1) Meat fat enhances intracellular lipid storage and impairs insulin metabolism (58, 59); 2) Heme iron from meat may damage pancreas cells (60, 61) and 3) Meat sourced nitrites and sodium may impair the function of the pancreatic beta cells (62).

Leptin is a hormone made by adipose cells that helps to regulate energy balance by inhibiting hunger. A number of studies reported that sugar consumption was correlated to production of leptin, but the results were controversial (63, 64).

The role of sugar consumption in the development of overweight and obesity has overly received scientific and policy attention. There are literally thousands of postings on the internet related to putative healthy diet guideline links between sugar and body obesity as well as insulin resistance. For instance, Te Morenga *et al.* concluded that intake of sugar is a determinant of body weight after assessment of 6,557 relevant academic publications (8). Some authorities have taken action to limit young students' access to sugar products. Furthermore, taxation of sugar has been advocated or implemented in some countries/areas as a potential public health strategy to curb the obesity epidemic.

Meat, by contrast, has not been singled out as one of the worst dietary offenders. In terms of balanced diet components, what the public have been told overwhelmingly is that a moderate amount of meat (lean meat preferred) should be included in our daily diet as it contains essential protein and minerals. Although meat protein is nutritious and integral to our health, protein halo should not be prevailing. While recognizing that the evidence of harm to health against meat is statistically as strong as sugar, we should avoid the trap of waiting for absolute proof before allowing public health action to be taken. A survey of approximately 100,000 North American members of the Seventh Day Adventist Church (65) indicated that vegan members had the lowest BMI values, while mean BMI increased gradually with increasing amounts of animal protein consumed by lacto-ovo vegetarians, pesco-vegetarians and semi-vegetarians reaching the highest value in non-vegetarians.

A strength of this study is that we used comparable per capita availability data from 170 countries which enabled us to examine and compare relationships between obesity prevalence and different food groups (sugar and meat) at population level. However,

there are several limitations in this study. Firstly, although we attempted to remove the potential confounding effects of variables such as GDP, caloric *etc.* by means of partial correlation analysis, some confounding factors may still influence correlations we found. Secondly, there may be some variables not included in our analysis that influence the correlation found in this study. It is however difficult to see what such variables may be. Thirdly, we could only use an international food group database that tracks the general market availability of different food types, not the actual human consumption. There are no direct measures of actual human consumption that can account for food wastage and provide precise measures of food consumption internationally. The database did not contain sufficient information to separate effects of "pure meat" from meat products that may contain other nutrients. Finally, the data analysed are calculated for per capita in each country, so we can only demonstrate a relationship between food group availability and obesity at a country level, which does not necessarily correspond to the same relationships holding true at the individual level. Prospective cohort studies are proposed to explore these associations further.

## Conclusion

Both sugar and meat availability are correlated to obesity prevalence worldwide, and there is no significant difference between the levels of two correlations. Similar to the public campaign against excessive sugar consumption, considering the findings of adverse effects of meat on obesity and the environmental impact of meat production, the country authorities should also advise the public not to adopt a high-meat diet for long-term healthy weight management.

## Declarations

### Ethics approval and Availability of data

All the data used in this study were freely downloaded from the United Nations (UN) agencies' websites. No ethical approval or written informed consent for participation was required.

### Competing interests

The authors (WPY & MH) declare that there are no conflicts of interest, financial or otherwise, related to this paper.

### Authors' contributions

MH conceived the idea for this study. WPY extracted the data. MH and WPY analysed and interpreted the data. WPY reviewed the literature and drafted the manuscript. WPY and MH edited and approved the manuscript for submission to the journal.

**Acknowledgments**

The authors express appreciation to the Statistics Department of FAO for the assistance in locating and defining the data.

## References

1. Kannel WB, Lebauer EJ, Dawber TR, McNamara PM. Relation of body weight to development of coronary heart disease. The Framingham study. Circulation. 1967;35(4):734-43.

2. WHO. Global Health Risks Mortality and Burden of Disease Attributable to Selected Major Risks. Geneva: Geneva : World Health Organization; 2009.

3. Puhl R, Brownell KD. Bias, discrimination, and obesity. Obesity research. 2001;9(12):788-104.

4. Thieme H. The Lower Paleolithic art of hunting: The case of Schöningen 13 II–4, Lower Saxony, Germany. In: Gamble C, Parr M, editors. The hominid individual in context: archaeological investigations of Lower and Middle Paleolithic landscapes, locales and artefacts. 115–13. London: Routeledge; 2005.

5. Henneberg M, Grantham J. Obesity - a natural consequence of human evolution. Anthropological Review. 2014;77(1):1-10.

6. Grantham JP, Staub K, Rühli FJ, Henneberg M. Modern diet and metabolic variance--a recipe for disaster? Nutrition journal. 2014;13:15:01-10.

7. Rouhani M, Salehi-Abargouei A, Surkan P, Azadbakht L. Is there a relationship between red or processed meat intake and obesity? A systematic review and meta-analysis of observational studies. Obes Rev2014. p. 740-8.

8. Te Morenga L, Mallard S, Mann J. Dietary sugars and body weight: systematic review and meta-analyses of randomised controlled trials and cohort studies. BMJ. 2013;346:01-25.

9. Hu FB. Resolved: there is sufficient scientific evidence that decreasing sugar-sweetened beverage consumption will reduce the prevalence of obesity and obesity-related diseases. Obesity reviews : an official journal of the International Association for the Study of Obesity. 2013;14(8):606-19.

10. Schulze MB, Fung TT, Manson JE, Willett WC, Hu FB. Dietary patterns and changes in body weight in women. Obesity (Silver Spring, Md). 2006;14(8):1444.

11. Wang Y, Beydoun MA. Meat consumption is associated with obesity and central obesity among US adults. International journal of obesity. 2009;33(6):621.

12. You W, Henneberg M. Cereal Crops Are not Created Equal: Wheat Consumption Associated with Obesity Prevalence Globally and Regionally. AIMS Public Health. 2016;3(2):313-28.

13. Davis WR. Wheat Belly: Lose the Wheat, Lose the Weight, and Find Your Path Back to Health. Rodale Books. 2011.

14. Davis B, Wansink B. Fifty years of fat: news coverage of trends that predate obesity prevalence. BMC public health. 2015;15:629:1-6.

15. den Engelsen C, Gorter KJ, Salome PL, Rutten GE. Development of metabolic syndrome components in adults with a healthy obese phenotype: a 3-year follow-up. Obesity. 2013;21(5):1025-30.

16. Trøseid M, Seljeflot I, Weiss TW, Klemsdal TO, Hjerkinn EM, Arnesen H. Arterial stiffness is independently associated with interleukin-18 and components of the metabolic syndrome. Atherosclerosis. 2010;209(2):337-9.

17. You W, Henneberg M. Meat consumption providing a surplus energy in modern diet contributes to obesity prevalence: an ecological analysis. BMC Nutrition. 2016;2(1).

18. Siervo M, Montagnese C, Mathers JC, Soroka KR, Stephan BCM, Wells JCK. Sugar consumption and global prevalence of obesity and hypertension: an ecological analysis. 2014;17(3):587-96.

19. Roccisano D, Henneberg M. Soy Consumption and Obesity. Food and Nutrition Sciences. 2012;03(02):260-6.

20. Basu S, Yoffe P, Hills N, Lustig RH. The relationship of sugar to population-level diabetes prevalence: an econometric analysis of repeated cross-sectional data. PloS one. 2013;8(2):e57873: 1-8.

21. Basu S, Stuckler D, McKee M, Galea G. Nutritional determinants of worldwide diabetes: an econometric study of food markets and diabetes prevalence in 173 countries. Public health nutrition. 2013;16(1):1-8.

22. Weeratunga P, Jayasinghe S, Perera Y, Jayasena G, Jayasinghe S. Per capita sugar consumption and prevalence of diabetes mellitus – global and regional associations. BMC public health. 2014;14:186-91.

23. Giskes K, van Lenthe FJ, Turrell G, Kamphuis CB, Brug J, Mackenbach JP. Socioeconomic position at different stages of the life course and its influence on body weight and weight gain in adulthood: a longitudinal study with 13-year follow-up. Obesity. 2008;16(6):1377-81.

24. Nestle M. Increasing portion sizes in American diets: More calories, more obesity. Journal of the American Dietetic Association. 2003;103(1):39-40.

25. Berentzen T, Sorensen TI. Physical inactivity, obesity and health. Scandinavian journal of medicine & science in sports. 2007;17(4):301-2.

26. Allender S, Foster C, Hutchinson L, Arambepola C. Quantification of urbanization in relation to chronic diseases in developing countries: a systematic review. Journal of urban health : bulletin of the New York Academy of Medicine. 2008;85(6):938-51.

27. Sun SZ, Empie MW. Fructose metabolism in humans – what isotopic tracer studies tell us. Nutrition & metabolism. 2012;9(89):01-15.

28. Faeh D, Minehira K, Tappy L, Schwarz J-M, Seongsu P, Periasami R. Effect of fructose overfeeding and fish oil administration on hepatic de novo lipogenesis and insulin sensitivity in healthy men. Diabetes. 2005;54(7):1907-13.

29. Livesey G. Fructose, Obesity, and Related Epidemiology. Critical Reviews in Food Science and Nutrition. 2010;50:26-8.

30. RK C, RM L, PE R. Effects of glucose or fructose feeding on glycogen repletion in muscle and liver after exercise or fasting. 1987:126-32.

31. Latham MC. Human nutrition in the developing world Rome: Food and Agriculture Organization of the United Nations; 1997.

32. Rouhani MH, Salehi-Abargouei A, Surkan PJ, Azadbakht L. Is there a relationship between red or processed meat intake and obesity? A systematic review and meta-analysis of observational studies. Obesity reviews : an official journal of the International Association for the Study of Obesity. 2014;15(9):740-8.

33. Willett W, Leibel R. Dietary fat is not a major determinant of body fat. American Journal Of Medicine. 2002;113:47-59.

34. Melanson EL, Astrup A, Donahoo WT. The relationship between dietary fat and fatty acid intake and body weight, diabetes, and the metabolic syndrome. Annals of nutrition & metabolism. 2009;55(1-3):229-43.

35. Pearce KL, Norman HC, Hopkins DL. The role of saltbush-based pasture systems for the production of high quality sheep and goat meat. Small Ruminant Research. 2010;91(1):29-38.

36. Lawrie RA, Ledward DA. Lawrie's meat science (7th ed.). Cambridge: Woodhead Publishing Limited; 2006.

37. FAO. Food energy - methods of analysis and conversion factors Rome: Food and Agriculture Organization, 2003 92-5-105014-7.

38. Grantham JP, Staub K, Rühli FJ, Henneberg M. Modern diet and metabolic variance--a recipe for disaster? Nutrition journal. 2014;13:15.

39. Henneberg M, Rühli FJ, Gruber P, Woitek U. Alanine Transaminase Individual Variation Is a Better Marker than Socio-Cultural Factors for Body Mass Increase in Healthy Males. Food and Nutrition Sciences. 2011;02(10):1054-62.

40. Sonksen P, Sonksen J. Insulin: understanding its action in health and disease. BJA: International Journal of Anaesthesia. 2000;85(1):69-79.

41. Singh M, Shin Y-K, Yang X, Zehr B, Chakrabarti P, Kandror KV. 4E-BPs control fat storage by regulating the expression of Egr1 and ATGL. J Biol Chem. 2015:01-21

42. Musselman LP, Fink JL, Ramachandran PV, Patterson BW, Okunade AL, Maier E, et al. Role of fat body lipogenesis in protection against the effects of caloric overload in Drosophila. The Journal of biological chemistry. 2013;288(12):8028.

43. DiAngelo JR, Birnbaum MJ. Regulation of Fat Cell Mass by Insulin in Drosophila melanogaster. Molecular and Cellular Biology. 2009;29(24):63416352.

44. Lee G, Park J. Hemolymph sugar homeostasis and starvation-induced hyperactivity affected by genetic manipulations of the adipokinetic hormone-encoding gene in Drosophila melanogaster. Genetics. 2004;167(1):311-23.

45. Elliott SS, Keim NL, Stern JS, Teff K, Havel PJ. Fructose, weight gain, and the insulin resistance syndrome.(Abstract). American Journal of Clinical Nutrition. 2002;76(5):911-22.

46. Basciano H, Federico L, Adeli K. Fructose, insulin resistance, and metabolic dyslipidemia. Nutrition & metabolism. 2005;2:01-14.

47. Dandona P, Aljada A, Bandyopadhyay A. Inflammation: the link between insulin resistance, obesity and diabetes. Trends in Immunology. 2004;25(1):4-7.

48. Shoelson SE, Herrero L, Naaz A. Obesity, Inflammation, and Insulin Resistance. Gastroenterology. 2007;132(6):2169-80.

49. Hallfrisch J, Lazar F, Jorgensen C, Reiser S. Insulin and glucose responses in rats fed sucrose or starch. Insulin and glucose responses in rats fed sucrose or starch. 1979;32(4):787-93.

50. Hwang IS, Ho H, Hoffman BB, Reaven GM. Fructose-induced insulin resistance and hypertension in rats. Hypertension. 1987;10(5):512-6.

51. Stanhope KL, Schwarz J-m, Havel PJ. Adverse metabolic effects of dietary fructose: results from the recent epidemiological, clinical, and mechanistic studies. Current Opinion in Lipidology. 2013;24(3):198-206.

52. Isganaitis E, Lustig RH. Fast food, central nervous system insulin resistance, and obesity. Arteriosclerosis, thrombosis, and vascular biology. 2005;25(12):2451-62.

53. Johnson RJ, Nakagawa T, Sanchez-Lozada LG, Shafiu M, Sundaram S, Le M, et al. Sugar, uric acid, and the etiology of diabetes and obesity. Diabetes. 2013;62(10):3307.

54. Grundy SM. Hypertriglyceridemia, insulin resistance, and the metabolic syndrome. The American Journal of Cardiology. 1999;83(9):25-9.

55. Steven EK, Rebecca LH, Kristina MU. Mechanisms linking obesity to insulin resistance and type 2 diabetes. Nature. 2006;444(7121):840-6.

56. Beck-Nielsen H, Pedersen O, Lindskov HO. Impaired cellular insulin binding and insulin sensitivity induced by high-fructose feeding in normal subjects. Impaired cellular insulin binding and insulin sensitivity induced by high-fructose feeding in normal subjects. 1980;33(2):273-8.

57. Stanhope KL, Schwarz JM, Keim NL, Griffen SC, Bremer AA, Graham JL, et al. Consuming fructose-sweetened, not glucose-sweetened, beverages increases visceral adiposity and lipids and decreases insulin sensitivity in overweight/obese humans. The Journal of clinical investigation. 2009;119(5):1322-34.

58. Sparks LM, Xie H, Koza RA, Mynatt R, Hulver MW, Bray GA, et al. A high- fat diet coordinately downregulates genes required for mitochondrial oxidative phosphorylation in skeletal muscle. Diabetes. 2005;54(7):1926-33.

59. Petersen KF, Dufour S, Befroy D, Garcia R, Shulman GI. Impaired Mitochondrial Activity in the Insulin- Resistant Offspring of Patients with Type 2 Diabetes. The New England Journal of Medicine. 2004;350(7):664-71.

60. Rajpathak SN, Crandall JP, Wylie-Rosett J, Kabat GC, Rohan TE, Hu FB. The role of iron in type 2 diabetes in humans. BBA - General Subjects. 2009;1790(7):671-81.

61. Hua NW, Stoohs RA, Facchini FS. Low iron status and enhanced insulin sensitivity in lacto-ovo vegetarians. British Journal of Nutrition. 2007;86(04):515-9.

62. Pereira EC, Ferderbar S, Bertolami MC, Faludi AA, Monte O, Xavier HT, et al. Biomarkers of oxidative stress and endothelial dysfunction in glucose intolerance and diabetes mellitus. Clinical biochemistry. 2008;41(18):1454-60.

63. Lana A, Rodriguez-Artalejo F, Lopez-Garcia E. Consumption of sugar-sweetened beverages is positively related to insulin resistance and higher plasma leptin concentrations in men and nonoverweight women. J Nutr. 2014;144(7):1099-105.

64. Teff KL, Grudziak J, Townsend RR, Dunn TN, Grant RW, Adams SH, et al. Endocrine and metabolic effects of consuming fructose- and glucose-sweetened beverages with meals in obese men and women: influence of insulin resistance on plasma triglyceride responses. The Journal of clinical endocrinology and metabolism. 2009;94(5):1562-9.

65. Tonstad S, Butler T, Yan R, Fraser GE. Type of vegetarian diet, body weight, and prevalence of type 2 diabetes.(ORIGINAL ARTICLE: Clinical Care/Education/Nutrition/Psychosocial Research). Diabetes Care. 2009;32(5):791.

## Additional materials

Additional file 1: Food items in each food group

| Food group | | Items included |
|---|---|---|
| This study | FAO definition | |
| Meat | Meat, total | Beef and veal, Buffalo meat, Pig meat, Mutton and lamb, Goat meat, Horse meat, Chicken meat, Goose meat, Duck meat, Turkey meat, Rabbit meat, Game meat and Offal |
| Fruits | Fruits | Melons, Watermelons, Apples, Apricots, Avocados, Cherries, Figs, Grapes, Mangoes, Papaya, Peaches, Pears, Persimmons, Pineapples, Plums, Quinces, Blueberries, Gooseberries, Raspberries, Strawberries, Kiwi, Dates, Figs (dried), Prunes, Currants, Raisins and Other fresh and dried fruits |
| Fibers | Vegetables | Beets, Carrots, Turnips, Rutabagas or swedes, Onions (green), Onions (dry), Artichokes, Tomatoes, Asparagus, Cabbage, Cauliflower, Celery, Kale, Lettuce, Spinach, Beans (green), Broad beans (green), Chilli peppers, Garlic, Cucumbers, Mushrooms, Eggplant, Peas (green), Pumpkins, Squash, Gourds, Okra, Radishes and Other vegetables |
| | Pulses | Beans (dry), Broad beans (dry), Peas (dry), Chick peas, Cow peas, Pigeon peas, Lentils, Vetches, Lupins and Other pulses |
| Starch | Starchy root | Potatoes, Sweet potatoes, Cassava, Taro, Yams and Other roots and tubers |
| | Cereals | Wheat, Rye, Barley, Oats, Maize, Rice, Mixed grains, Buckwheat, Sorghum, Millet, Quinoa and Other cereals |
| Fats | Vegetable oil | Sunflower seed oil, Cottonseed oil, Linseed oil, Hempseed oil, Sesame seed oil, Copra and coconut oil, Palm kernel oil, Palm oil, Soybean oil, Olive oil and Maize oil |
| | Animal fats | Butter, Ghee, Fish liver oil, Whale oil and Other animal fats |

*Source: http://www.fao.org/docrep/003/x9892e/x9892e02.htm*

Additional file 2: Data descriptive summary and source

| | N Statistic | Minimum Statistic | Maximum Statistic | Mean Statistic | Std. Deviation Statistic | Skewness Statistic | Skewness Std. Error | Data source |
|---|---|---|---|---|---|---|---|---|
| BMI≥30,18+,2010 | 170 | 1.7 | 41.8 | 16.670 | 8.9937 | 0.06 | 0.19 | WHO |
| BMI≥25,18+, 2010 | 170 | 9.7 | 72.8 | 43.021 | 19.4304 | -0.37 | 0.19 | WHO |
| BMI mean,18+, 2010 | 170 | 20.3 | 31.4 | 25.258 | 2.1433 | -0.26 | 0.19 | WHO |
| Meat, total 2007-09 mean (g/capita/day) | 167 | 10.73 | 337.26 | 129.57 | 82.53 | 0.42 | 0.19 | FAO |
| Fat (plant oil+ animal fat) (g/capita/day) 07-09 | 161 | 4.63 | 137.87 | 42.55 | 26.83 | 1.10 | 0.19 | FAO |
| Fruits (g/capita/day) 2007-09 mean | 167 | 15.26 | 964.69 | 229.70 | 142.71 | 1.39 | 0.19 | FAO |
| Fiber (Vegetables + Pulses) (g/capita/day), 2007-09 mean | 167 | 35.59 | 873.12 | 251.97 | 164.61 | 1.32 | 0.19 | FAO |
| Starch (Cereals + Starchy root) (g/capita/day) 07-09 mean | 167 | 219.53 | 1345.32 | 573.11 | 176.81 | 1.23 | 0.19 | FAO |
| Meat protein (g/capita/day), 2007-09 mean | 167 | 1.53 | 40.83 | 16.37 | 10.21 | 0.49 | 0.19 | FAO |
| Animal protein, excl meat protein (g/capita/day) 2007-09 mean | 167 | 1.87 | 65.30 | 18.94 | 11.71 | 0.90 | 0.19 | FAO |
| Plant protein (g/capita/day) 2007-09 mean | 167 | 23.88 | 77.56 | 43.50 | 9.08 | 0.85 | 0.19 | FAO |
| Animal Fat Total (g/capita/day), 2007-09 mean | 167 | 4.58 | 105.35 | 37.45 | 24.94 | 0.62 | 0.19 | FAO |
| Plant oil total (g/capita/day),2007-09 mean | 167 | 15.39 | 98.75 | 45.99 | 16.68 | 0.69 | 0.19 | FAO |
| Carbohydrate (g/capita/day) 2007-09 mean, self-calculated | 167 | 303.47 | 627.64 | 431.23 | 55.46 | 0.29 | 0.19 | FAO |
| Calories Total (kcal/capita/day) 07-09 mean | 167 | 1820.00 | 3791.67 | 2793.59 | 462.09 | 0.13 | 0.19 | FAO |
| Minimum Dietary Energy Requirement (kcal/person/day)06-08 | 167 | 1690 | 2000 | 1852.40 | 87.655 | -0.05 | 0.19 | FAO |
| Prevalence of physicak inactivity %, 18+, 2010 | 131 | 4.1 | 63.6 | 25.302 | 11.4495 | 0.73 | 0.21 | WHO |
| GDP PPP (current US$), 2010 | 165 | 619.47 | 84200.57 | 15224.06 | 15714.26 | 1.70 | 0.19 | World Bank |
| Valid N (listwise) | 123 | | | | | | | |

Article 6/10: Prostate cancer incidence is correlated to total meat (flesh) intake– A cross-national analysis of 172 countries

Wenpeng You[1], Maciej Henneberg[1,2]

[1] Biological Anthropology and Comparative Anatomy Unit, Adelaide Medical School, the University of Adelaide, Adelaide, SA, Australia 5000.

[2] Institute of Evolutionary Medicine, University of Zürich, Switzerland.
maciej.henneberg@adelaide.edu.au

**Under review:** Wenpeng You, Maciej Henneberg. "Meat consumption and prostate cancer incidence-global and regional associations (Abstract)" *BJU INTERNATIONAL*. Vol. 118. 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY-BLACKWELL, 2016. The manuscript, titled Prostate cancer incidence is correlated to total meat (flesh) intake– A cross-national analysis of 172 countries, is under review by BJU International.

✉ **Correspondence:** Wenpeng You Wenpeng.you@adelaide.edu.au

## Contexture Statement

Meat types (red meat, instead of white meat) and levels of doneness have been constantly associated with prostate cancer (PC61). However, there has been no substantial research into the real difference between red meat and white meat in terms of their detrimental health effects. Contrarily, studies have suggested that white meat and red meat contribute to body weight at the similar level. Red meat with high level of doneness may have been well-known to PC61 patients to contribute to their cancer initiation. Therefore, it is easy for them to exaggerate how much red meat and level of doneness of red meat they had before they were diagnosed with PC61.

Total fresh meat, regardless of meat types and levels of doneness, may be an objective measure for meat consumption. We postulated and assessed the correlation between total meat (fresh) consumption and PC61 incidence at population level.

## Statement of Authorship

# Statement of Authorship

| | |
|---|---|
| Title of Paper | Prostate cancer incidence is correlated to total meat (flesh) intake– A cross-national analysis of 172 countries |
| Publication Status | ☐ Published ☐ Accepted for Publication<br>☑ Submitted for Publication ☐ Publication Style |
| Publication Details | Under review by BJU International since 02-Feb-2018, IDBJU-2018-0171 |

### Principal Author

| | | | |
|---|---|---|---|
| Name of Principal Author (Candidate) | Wenpeng You | | |
| Contribution to the Paper | Conceived the hypothesis, collected data, performed analysis on all samples, interpreted data, wrote manuscript and acted as corresponding author | | |
| Overall percentage (%) | 65 | | |
| Signature | | Date | 22/01/2017 |

### Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

| | | | |
|---|---|---|---|
| Name of Co-Author | Maciej Henneberg | | |
| Contribution to the Paper | Supervised development of work, formalized and interpreted data analysis, and evaluated and edited manuscript | | |
| Signature | | Date | 02/02/2018 |

| | | | |
|---|---|---|---|
| Name of Co-Author | | | |
| Contribution to the Paper | | | |
| Signature | | Date | |

| | | | |
|---|---|---|---|
| Name of Co-Author | | | |
| Contribution to the Paper | | | |
| Signature | | Date | |

Please cut and paste additional co-author panels here as required.

## Abstract

**Objective:** To examine the association of total meat (flesh of animals) consumption to Prostate Cancer incidence (PC61) at population level.

**Subjects and Methods:** Data from 172 countries were extracted for analysis. Associations between country specific per capita total meat intake and PC61 incidence at country level were examined using Pearson's r and Spearman rho, partial correlation, stepwise multiple linear regression analyses with ageing, GDP, $I_s$ (index of magnitude of prostate cancer gene accumulation at population level), obesity prevalence and urbanization included as the confounding factors. Countries were also grouped for regional association analysis. The data were log-transformed for analysis in SPSS. Microsoft Excel, and ANOVA Post hoc Scheffe tests were applied to calculate and compare mean differences between country groupings.

**Results:** Worldwide, total meat intake was strongly and positively associated with PC61 incidence in Pearson's r ($r= 0.595$, $p<0.001$) and Spearman rho ($r= 0.637$, $p<0.001$) analyses. This relationship remained significant in partial correlation ($r= 0.295$, $p<0.001$) when ageing, GDP, Is, obesity prevalence and urbanization were kept statistically constant. GDP was weakly and insignificantly associated with PC61 when total meat intake was kept statistically constant. Stepwise multiple regression identified that total meat was a significant predictor of PC61 with total meat intake and all the five confounders included as the independent variables ($R^2=0.417$). GDP was not identified as the statistically significant predictor of PC61 in either of the models including or excluding total meat as the independent variable. Post hoc Scheffe tests revealed nine significant mean differences of PC61 between the six WHO regions, but all disappeared when the contributing effect of total meat on PC61 incidence rate was removed.

**Conclusions:** Total meat intake is an independent predictor of PC61 worldwide, and the determinant of regional variation of PC61. The longitudinal cohort studies are proposed to explore the association further.

**Keywords**: Total meat (flesh of animals), Prostate cancer, Carcinogen, Regional variation

## Introduction

Prostate cancer (PC61) is the second most common cancer among men in the world [1]. It has become an enormous public health concern in most developed countries and an emerging public health problem in developing countries [2, 3]. Globally, an estimated 0.9 million men in 2008 were diagnosed with prostate cancer worldwide [4], but, in 2012,

the number increased to 1.1 million [5] and the majority of cases (almost 70%) occur in developed countries [5].

A complete understanding of the aetiology of PC61 remains elusive to the public and professionals [6, 7]. Genetic background is the well-established risk factor through studies of PC61 in family histories [8, 9], and this background may be have been accumulated in human population due to the reduced natural selection [10-12]. Researches into the relationship between ageing and PC61 have revealed that, essentially, ageing process leads to the acquisition of mutations and the formation of a molecular and cellular environment which favours carcinogenesis [13-15]. Recent studies have shown that people who are obese may have more exposure to PC61 risk because they have the increased blood levels of insulin and insulin-like growth factor-1 (IGF-1) [16, 17]. Urbanization has been closely linked to human lifestyle change, such as more meat intake [18, 19] and less physical exercise [20], due to its process of modernization and industrialization. Therefore, it has been postulated as the risk factor of PC61 [21].

PC61 epidemiology has revealed that its incidence varies more than 25-fold worldwide [22]. In the past years, researchers from the International Agency for Research on Cancer (IARC) as the specialized cancer agency of the World Health Organization (WHO) published several articles/reports associating the regional variation of PC61 incidence with regional socioeconomic levels [1, 5].

Diet pattern plays a very important role in causing a large percentage of cancers [23, 24]. Plant sourced food products, such as vegetables [25], fruits [25] and grains [26], have been reported as not associated with prostate cancer.

In the last decades, a number of large cohort and case-control studies have controversially and circumstantially linked red meat intake to the development of PC61 [27-30]. It has been suggested that there was no substantial difference between “red meat” and “white meat” in terms of the nutrient components [19, 31]. Therefore, both red meat and white meat might contribute to PC61 together when people had diets which usually include the combination of red meat and white meat. Researches, which simply correlated red meat intake and PC61 risk, may have a defect in the study designs because the contributing effects of white meat intake to PC61 was not removed. Statistically, we may say that white meat intake was not kept constant when the correlation of red meat intake to PC61 was analysed [32, 33].

It is proposed to use ecological study for ascertaining a new association between total meat (flesh of animals) intake and PC61 risk at population level. We examined this

relationship with the country specific data on total meat intake and PC61 incidence rate published by the United Nations (UN) agencies.

## Materials and Methods

### Data Collection and Selection

The country specific data were collected for this study:

The most recent IARC data on estimated PC61 incidence rate in 2012 for the adult (aged 15+ years old) part of each population were extracted as the dependent variable [1].

Total meat intake (expressed in kg/capita/year) in 2011 from the FAOSTAT Food Balance Sheet (FBS) [34] was obtained as the independent predictor of PC61. FAO defined total meat as “flesh of animals used for food”, which includes beef and veal, buffalo meat, pig meat, mutton and lamb, goat meat, horse meat, chicken meat, goose meat, duck meat, turkey meat, rabbit meat, game meat and offal [34]. For the interest of discussing the relationships between PC61 and white meat intake and red meat intake, we extracted poultry meat (flesh) as white meat (expressed in kg/capita/year). We calculated red meat intake by subtracting white meat intake from the total meat intake.

We extracted the following data as the confounding variables as they have been postulated as the risk factors of PC61.

Ageing, expressed with the percentage of males age 65 and above in each country in 2011 was extracted from the World Bank [35].

The World Bank data on per capita GDP PPP (gross domestic product converted to international dollars using purchasing power parity rates) in 2011 [35]. Ferlay *et al.* indicated that the PC61 incidence rate varies significantly largely because of how widespread the prostate specific antigen (PSA) testing and subsequent biopsy are in practice in those countries and regions [1]. The testing depends on the GDP because of funding for medical services. GDP PPP was incorporated as the confounding factor to reduce/remove the bias on PC61 incidence in addition to other socioeconomic level related factors which may affect the association between meat intake and PC61 incidence.

Country-specific index of the total opportunity for natural selection in modern populations ($I_s$) was extracted from previous studies [10-12]. An $I_s$ value signifies here the magnitude of the country to accumulate the PC61 genes [10-12]. The calculation methods and significance of $I_s$ recently used by You and Henneberg [10, 36] is based on the Biological State Index as described in Henneberg [37] and Henneberg and Piontek [38]. PC61 has strong genetic background which is heritable [3, 39]. Therefore, $I_s$ was chosen as the

confounding factor to remove the confounding effect of country-specific PC61 genetic background on the association between meat intake and PC61 incidence [10-12, 40].

The WHO Global Health Observatory (GHO) data on the estimated prevalence rate of obesity (percent of population aged 18+ with BMI ≥ 30 kg/m2) of the male population in 2010 [41].

The World Bank data on urbanization (the percent of males living in urban areas in each country in 2011 [35].

We simply extracted the country-specific meat intake data from the FAO Food Balance Sheet for 172 countries, that is all countries of the world for which these data were available. And then, we matched the other variables with the meat intake data. All the independent variables were backdated 1-2 years to reflect the exposure with delayed presentation of PC61.

Each country was treated as the individual subject for data analysis in this study. For particular analyses, the number of countries included for variables may have differed somewhat because all information on other variables was not uniformly available for all countries due to unavailability from relevant UN agencies. All the data were extracted and saved in Microsoft Excel® for performing the data analysis.

**Statistical analyses**

To assess the relationship between PC61 incidence rate and total meat intake, the analysis proceeded in six steps.

1. Scatter plots was produced with the original data in Microsoft Excel® to explore and visualize the strength, shape and direction of association between meat intake and PC61 incidence at the global level.

We also calculated and compared the means of PC61 of the 10 countries with highest and lowest meat intake in the Excel to show how meat consumption changes average incidence rates of PC61

For the data analysis in SPSS (Steps 2 -5), the original data were log-transformed (natural logarithms) to bring their distributions closer to normal, which may increase homoscedasticity of data distributions.

2. Bivariate correlation (Pearson's and Nonparametric) was used to evaluate the strength and direction of the associations between both dependent variable (PC61 incidence) and all independent variables (Meat intake, Ageing, GDP PPP, Obesity and Urbanization).

3. Partial correlation of Pearson moment-product approach was used to find the relationship between PC61 incidence and meat intake while keeping ageing, GDP PPP, obesity and urbanization statistically constant.

The independent relationships between PC61 and each of the five variables were explored with partial correlation of Pearson's moment-product approach while we kept the meat intake statically constant. This allows us to identify how strongly the meat intake affects the association between PC61 and each of the five variables.

A number of previous ecological studies [18, 19, 42, 43] revealed that meat intake was in significant and strong correlation to GDP. We alternated GDP and meat intake as the predictor and confounding factor for the partial correlation analysis.

4. Stepwise multiple linear regression modelling was performed to identify and rank predictors (independent variables) of PC61. We included and excluded meat intake as the one of the predictors in the two analyses to observe how strongly the meat intake affected the predictor ranking in Stepwise linear analysis.

5. Pearson's r was calculated to investigate the regional correlation between meat intake and PC61 incidence. Fisher's r-to-z transformation was performed to test significance of differences between correlation coefficients. We did this analysis because meat intake varies in human diet patterns due to the availability and affordability in different regions, and also because the WHO and its agent the IARC reported that PC61 incidence varies in different regions [1, 3]. The 173 countries were grouped as per WHO region division [44] and the World Bank income classifications [45] for correlation analyses.

6. Post hoc Scheffe (Oneway ANOVA) testing was performed to compare the mean difference of meat intake (original data), PC61 incidence (original data), and residual of PC61 incidence standardized on meat intake (original data) between six WHO regions. This may allow us to investigate the importance of meat intake in determining the regional variation of PC61.

The equation ($y = 0.7643x + 1.1864$) of the best fitting trendline obtained in the scatter plots analysis of correlation between meat intake and PC61 incidence was used to calculate and remove the contributing effect of total meat intake on PC61 incidence rate. Thus, we created a new dependent variable, "PC61 incidence standardized on meat intake" and subsequently "Residual of PC61 incidence standardised on meat intake" after subtracting the "PC61 incidence standardized on meat intake" from the PC61 incidence rate.

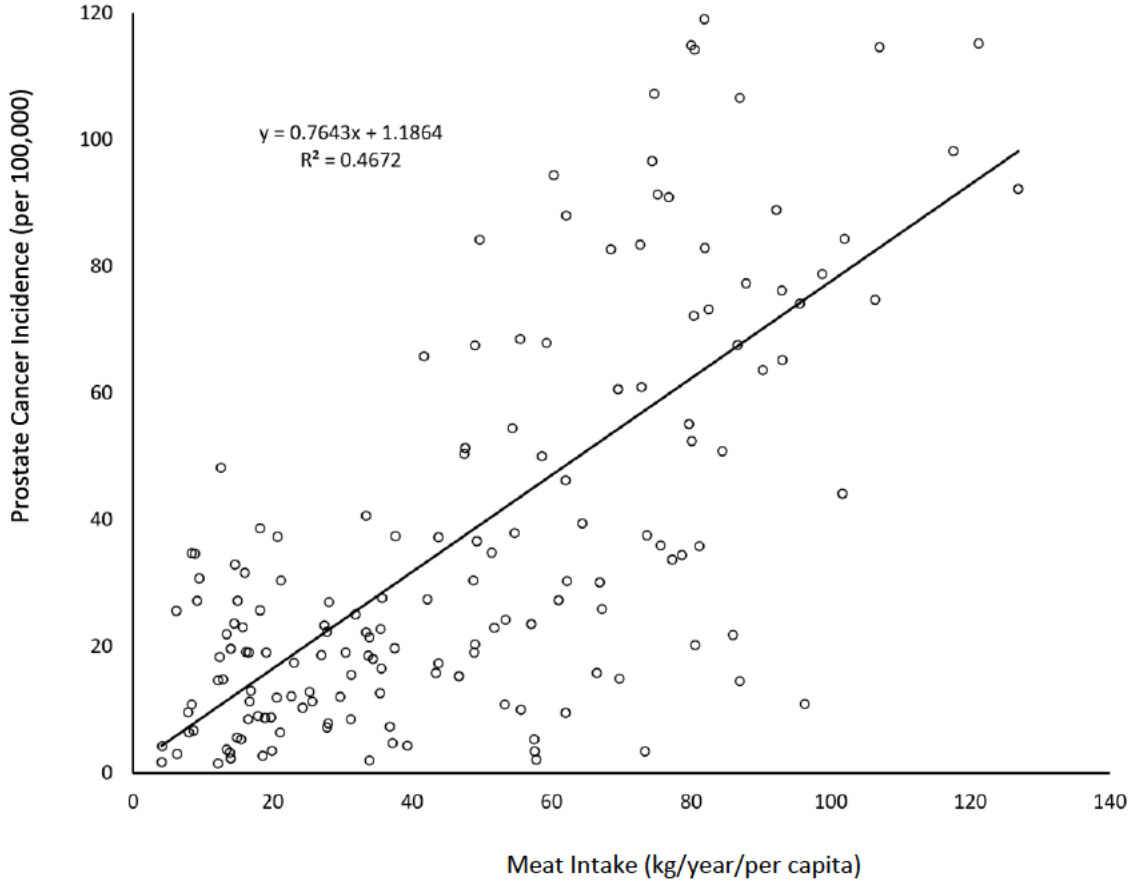
The means of PC61 and meat intake of the six WHO regions were compared and the association between meat intake and PC61 incidence in each was obtained in the Excel.

SPSS v. 22 (SPSS Inc., Chicago Il USA) and Microsoft Excel® were used for data analysis. The significance was kept at the 0.05 level, but 0.01 and 0.001 levels are also reported. Stepwise multiple linear regression analysis criteria were set at probability of F to enter ≤ 0.05 and probability of F to remove ≥ 0.10.

## Results

Figure 1 showed the unadjusted correlation between meat intake and PC61 incidence. The relationship was noted to be best described by linear equation (y = 0.7643x + 1.1864) with strong correlation (r=0.684, p<0.001).

Figure 1 Linear correlation plot of meat intake and prostate incidence




The average PC61 incidence rate of the 10 countries with highest meat intakes (79.22 per 100,000) was 5.77 times greater than the average of the 10 countries with lowest meat intakes (13.73 per 100,000).

Pearson correlation and nonparametric analyses showed that meat intake was in significantly strong correlation to PC61 incidence (r=0.595, p<0.001 and r=0.637, p<0.001 respectively) (Table 1). Pearson r correlation coefficient of meat consumption to PC61 became lower in scatter plots (Figure 1) because the variables were log-transformed. The strong and significant correlations were also observed between PC61 and ageing, GDP PPP, obesity and urbanization respectively. This warranted our

selection to include them as the confounding factors in exploring the correlation between meat intake and PC61 incidence.

Table 1 Pearson's r and nonparametric correlation matrix between all variables involved in this study

| | PC61 | Meat | Ageing | GDP PPP | $I_s$ | Obesity % | Urbanization |
|---|---|---|---|---|---|---|---|
| PC61 | 1 | 0.595*** | 0.555*** | 0.529*** | -0.480*** | 0.489*** | 0.470*** |
| Meat | 0.637*** | 1 | 0.648*** | 0.810*** | 0.674*** | 0.761*** | 0.588*** |
| Ageing | 0.587*** | 0.699*** | 1 | 0.706*** | 0.686*** | 0.596*** | 0.498*** |
| GDP | 0.573*** | 0.833*** | 0.750*** | 1 | 0.738*** | 0.717*** | 0.664*** |
| $I_s$ | -0.565*** | 0.794*** | 0.864*** | 0.871*** | 1 | 0.708*** | 0.505*** |
| Obesity % | 0.501*** | 0.737*** | 0.630*** | 0.729*** | 0.745*** | 1 | 0.671*** |
| URBAN | 0.516*** | 0.635*** | 0.563*** | 0.737*** | 0.665*** | 0.735*** | 1 |

Pearson r (above diagonal) and nonparametric (below diagonal) correlations were reported. Significance levels: * P <0.05, ** P< 0.01, *** P< 0.001. Number of country range, 157-172.

Meat intake (kg/capita/year) sourced from the Food and Agriculture Organization; Ageing (percent of males ages 65 and above) and GDP PPP (gross domestic product converted to international dollars using purchasing power parity rates) and urbanization (the percent of males living in urban areas) were sourced from the World Bank. Male obesity prevalence (percent of males aged 18+ with BMI ≥ 30 kg/m2); Is was extracted from previous publications.

These bivariate correlations were also reflected in the WHO regions showing increased correlation of meat intake with PC61 (Table 2). AFRO region was the exception. In general, the bivariate correlations were also true in country groupings based on economy status as defined by the GDP.

Table 2 Correlation of meat availability to each level of in different country groupings

| Country groupings | Pearson r | p | nonparametric | p |
|---|---|---|---|---|
| Worldwide (n=163) | 0.595 | <0.001 | 0.637 | P<0.001 |
| World Bank income classifications | | | | |
| High Income, n=47 | 0.528 | <0.001 | 0.346 | <0.05 |
| Low Income, n=26 | 0.429 | <0.05 | 0.372 | 0.061 |
| Low Middle Income, n=43 | 0.305 | <0.05 | 0.216 | 0.164 |
| Upper Middle, n=47 | 0.402 | <0.01 | 0.419 | P<0.003 |
| WHO regions | | | | |
| AFRO, n=38 | 0.180 | 0.280 | 0.049 | 0.771 |
| AMRO, n=29 | 0.570 | <0.001 | 0.555 | <0.01 |
| EMRO, n=18 | 0.524 | <0.05 | 0.556 | <0.05 |
| EURO, n=50 | 0.723 | <0.001 | 0.654 | <0.001 |
| SEARO, n=10 | 0.549 | 0.101 | 0.661 | <0.05 |
| WPRO, n=18 | 0.591 | <0.01 | 0.513 | <0.05 |

Pearson r and nonparametric correlations within country groupings were reported.

Meat intake (kg/capita/year) sourced from the Food and Agriculture Organization;

Partial correlation analysis revealed that meat intake was a strong and significant predictor of PC61 independent of ageing, GDP PPP, obesity and urbanization (r=0.295, p<0.001, Table 3). When meat intake was stabilised as a confounding factor in partial correlation analysis, it was revealed that: 1) ageing was identified as a significant independent predictor (r=0.277, p<0.001) of PC61 incidence; 2) urbanization showed weak and significant correlation to PC61 incidence (r=0.185, p<0.05); and 3) GDP, Is and Obesity showed barely a correlation to PC61 incidence (Table 3). This suggested that meat intake had great confounding effects on the correlation between PC61 incidence and GDP PPP, Is, obesity and urbanization respectively.

Table 3 Partial correlations between prostate cancer incidence and independent variable when meat was included as the independent and confounder respectively

| | Partial Correlation to PC61 | | | Partial Correlation to PC61 | | |
|---|---|---|---|---|---|---|
| Variables | r | p | df | r | p | df |
| Meat | 0.295 | <0.001 | 150 | - | - | - |
| Ageing | - | - | - | 0.277 | <0.001 | 160 |
| GDP | - | - | - | 0.100 | 0.209 | 160 |
| Is | - | - | - | -0.041 | 0.608 | 158 |
| Obesity | - | - | - | 0.070 | 0.382 | 158 |
| Urbanization | - | - | - | 0.185 | P<0.05 | 160 |

Partial correlations were reported.

Meat intake (kg/capita/year) sourced from the Food and Agriculture Organization; Ageing (percent of males ages 65 and above) and GDP PPP (gross domestic product converted to international dollars using purchasing power parity rates) and urbanization (the percent of males living in urban areas) were sourced from the World Bank. Male obesity prevalence (percent of males aged 18+ with BMI ≥ 30 kg/m2); Is was extracted from previous publications.

- Included as the confounding factor

When meat intake was excluded as the PC61 predictor, ageing and urbanization were selected as the significant predictors of PC61 with $R^2$ = 0.354 in the standard multiple linear regression (Stepwise) analysis. When meat intake was incorporated as an independent variable, it was placed first as the major predictor of PC61 with increasing $R^2$ to 0.417. GDP was not selected as the major predictor of PC61 in Stepwise linear regression. Additionally, it was not in strong or significant correlation to PC61 incidence in partial correlation (Table 2). This may suggest that GDP PPP may not be the strong predictor of PC61, but meat intake is.

Table 4 Results of Stepwise multiple linear regression analyses to sort significant predictors of Prostate Cancer incidence

| | Excluding meats | | | Including meat | |
|---|---|---|---|---|---|
| Rank | Variables Entered | Adjusted R Square | Rank | Variables Entered | Adjusted R Square |
| 1 | Ageing | 0.310 | 1 | Meat | 0.332 |
| 2 | Urbanization | 0.354 | 2 | Ageing | 0.386 |
| 3 | $I_{bs}$ | Not a major predictor | 3 | $I_s$ | 0.404 |
| 4 | GDP PPP | Not a major predictor | 4 | Urbanization | 0.417 |
| 5 | Obesity % | Not a major predictor | 5 | GDP PPP | Not a major predictor |
| | | | 6 | Obesity | Not a major predictor |

Stepwise multiple linear regression modelling is reported. Contribution of variables is listed in order of how much they contribute to prostate cancer incidence.

Meat intake (kg/capita/year) sourced from the Food and Agriculture Organization; Ageing (percent of males ages 65 and above) and GDP PPP (gross domestic product converted to international dollars using purchasing power parity rates) and urbanization (the percent of males living in urban areas) were sourced from the World Bank. Male obesity prevalence (percent of males aged 18+ with BMI ≥ 30 kg/m2); Is was extracted from previous publications.

Table 5 showed the calculated means of meat intake and PC61 incidence rates in all the six WHO regions. In general, at country grouping level, meat intake was in strong correlation to PC61 incidence based on the best fit trendline ($r=0.832$, $p<0.05$). This is consistent with the correlation between meat intake and PC61 incidence at the individual country level ($r=0.684$, $p<0.001$) (Table 5).

A post hoc Scheffe analysis conducted on the multiple mean comparisons revealed that there were numerous significant mean differences in PC61 incidence rates between different WHO regions (Table 5). Mean of PC61 incidence in Africa was significantly lower than that in Americas and Europe. Mean of PC61 incidence in the Eastern Mediterranean was significantly lower than that in Americas and Europe. The mean PC61 incidence in South-Eastern Asia was significantly lower than that in Americas and Europe.

A subsequent ANOVA with post hoc Scheffe procedure performed on the means of "Residual of PC61 standardised on meat intake" in different WHO regions showed no significant differences among and between regions (Table 5). The results from post hoc Scheffe tests conducted on mean comparison between the WHO regions suggested that regional variations of PC61 incidence may only reach statistically significant levels if the contribution of their respective meat intake was included. This result was supported by the findings identified in our previous bivariate and partial correlation (Table 3) and multiple linear regression (Table 4) that meat intake is the major risk factor of PC61 incidence.

Table 5 Mean difference between WHO regions, and between UN developed and developing regions

| | Meat | | PC61 incidence rate | | | Residual of PC61 incidence standardised on meat | | |
|---|---|---|---|---|---|---|---|---|
| I (Region) | J (Region) | Mean difference (I-J) | I (Region) | J (Region) | Mean difference (I-J) | I (Region) | J (Region) | Mean difference (I-J) |
| AF | AM | -43.07*** | AF | AM | -33.75*** | AF | AM | -1.97 |
| n=39 | EM | -11.98 | n=38 | EM | 10.35 | n=38 | EM | 18.92 |
| mean= 21.14 | EU | -45.51*** | mean=22.70 | EU | -32.19*** | mean= 5.37 | EU | 2.01 |
| | SEA | 3.40 | | SEA | 16.42 | | SEA | 13.24 |
| | WP | -40.83*** | | WP | -11.39 | | WP | 20.15 |
| AM | AF | 43.07*** | AM | AF | 33.75*** | AM | AF | 1.97 |
| n=36 | EM | 31.09*** | n=29 | EM | 44.10*** | n=29 | EM | 20.89 |
| mean= 33.12 | EU | -2.44 | mean=12.35 | EU | 1.56 | mean= 9.44 | EU | 3.98 |
| | SEA | 46.47*** | | SEA | 50.17*** | | SEA | 15.21 |
| | WP | 2.23 | | WP | 22.36 | | WP | 22.13 |
| EM | AF | 11.98 | EM | AF | -10.35 | EM | AF | -18.92 |
| n=18 | AM | -31.09*** | n=18 | AM | -44.10*** | n=18 | AM | -20.89 |
| mean= 40.77 | EU | -33.53*** | mean=40.77 | EU | -42.54*** | mean= -14.15 | EU | -16.91 |
| | SEA | 15.38 | | SEA | 6.07 | | SEA | -5.68 |
| | WP | -28.86* | | WP | -21.74 | | WP | 1.23 |
| EU | AF | 45.51*** | EU | AF | 32.19*** | EU | AF | -2.01 |
| n=50 | AM | 2.44 | n=50 | AM | -1.56 | n=50 | AM | -3.98 |
| mean=66.95 | EM | 33.53*** | mean=54.89 | EM | 42.54*** | mean=2.77 | EM | 16.91 |
| | SEA | 48.91*** | | SEA | 48.61*** | | SEA | 11.23 |
| | WP | 4.68 | | WP | 20.80 | | WP | 18.15 |
| SEA | AF | -3.40 | SEA | AF | -16.42 | SEA | AF | -13.24 |
| | AM | -46.47*** | | AM | -50.17*** | n=10 | AM | -15.21 |

| | Meat | | PC61 incidence rate | | | Residual of PC61 incidence standardised on meat | | |
|---|---|---|---|---|---|---|---|---|
| I (Region) | J (Region) | Mean difference (I-J) | I (Region) | J (Region) | Mean difference (I-J) | I (Region) | J (Region) | Mean difference (I-J) |
| n=10 | EM | -15.38 | n=10 | EM | -6.07 | mean=-8.47 | EM | 5.68 |
| mean= 17.74 | EU | -48.91*** | mean= 6.28 | EU | -48.61*** | | EU | -11.23 |
| | WP | -44.23*** | | WP | -27.81 | | WP | 6.91 |
| WP | AF | 40.83*** | WP | AF | 11.39 | WP | AF | -20.15 |
| n= 19 | AM | -2.23 | n= 18 | AM | -22.36 | n= 18 | AM | -22.13 |
| mean =61.97 | EM | 28.86* | mean=34.09 | EM | 21.74 | mean= -15.38 | EM | -1.23 |
| | EU | -4.68 | | EU | -20.80 | | EU | -18.15 |
| | SEA | 44.23*** | | SEA | 27.81 | | SEA | -6.91 |

Mean comparisons between WHO regions (One-way ANOVA, Post hoc Scheffe) were reported.

Meat intake (kg/capita/year) sourced from the Food and Agriculture Organization; Ageing (percent of males ages 65 and above) and GDP PPP (gross domestic product converted to international dollars using purchasing power parity rates) and urbanization (the percent of males living in urban areas) were sourced from the World Bank.

## Discussion

The results from our study suggested, at population level, total meat (flesh) intake was strongly and significantly associated with incidence rate of PC61 globally and regionally. Worldwide, total meat intake may be a major predictor of PC61 regardless of the influence from other risk factors, such as ageing, GDP, Is, obesity and urbanization. Our results also suggested that meat consumption, instead of GDP, may be a determinant of the regional variation of PC61.

Red and processed meat increasing risk of PC61 has been a central dogma reported in the majority of the studies into relationship between meat intake and PC61. The dogma, which is supported by the IARC [46], stipulates multiple etiologies through which red and processed meat intake contributes to PC61 risk [47]:

1) Carcinogens such as heterocyclic amines (PhIP), 2-amino-3,8-dimethylimidazo-[4,5-b] quinoxaline (MeIQx), and 2-amino-3, 4,8-trimethylimidazo-[4,5-f] quinoxaline (DiMeIQx), and polycyclic aromatic hydrocarbons may be formed when meat is cooked at high-temperature [48-53].

2) N-nitroso compounds (NOCs) may be produced endogenously from meat itself or preservatives added to processed meats [54-56].

3) Heme iron has catalytic effects on (i) the endogenous formation of carcinogenic *N*-nitroso compounds and (ii) the formation of cytotoxic and genotoxic aldehydes by lipoperoxidation [54, 57-61].

4) Meat may cause metabolic syndrome (MetS) [62], which play a role in the development of PC61 [63].

Recent studies reported that meat protein from both red meat and white meat may be digested slowly and later than other maco-nutrients, such as carbohydrates and fat [18, 19]. This may highlight the role of meat in contribute to PC61. However, the results from these studies may not be rigorous as they only focused on the relationship between red meat intake, instead of total meat intake, and PC61. It may not be wise to exclude white meat from the studies because: 1) The contents of red meat and white meat are quite similar although the quantities of the specific compounds are different. 2) Both red and white meat can produce the same mutagens or carcinogens when they are cooked at high temperature [64-66]. 3) Fat [33] and heme iron [58] [59-61] in red meat have been

postulated as the carcinogen. However, red meat has been leaner than ever over the past few decades due to leaner animals being bred and improved butchery and feeding techniques that make fat content fall significantly [67, 68]. Blood, which contains lots of heme iron, has been extensively consumed in Asian cuisines for thousands of years, but the PC61 incidence in Asia (9.4 per 100,000) is much lower than in other continents, such as Africa (23.2 per 100,000), Americas (75.0 per 100,000), Europe (61.3 per 100,000) and Oceania (101.9 per 100,000) [69]. Additionally, The National Pork Board of the United States used to classify pork, a major "red meat", as "the other white meat" [70]. Therefore, the contribution of white meat to PC61 may not be ignored in those studies. However, those studies into the relationship between red meat and PC61 did not remove the influence of white meat on PC61. In other words, statistically, there may be a defect in these studies as they did not establish the relationship independent of white meat consumption.

Some studies do not support that red meat should be the only meat category to be associated with PC61. Globally, the overall consumption of white meat (poultry in per capita per year) between 1990 and 2009 has increased by 76.6% [71]. Accompanying this process, the PC61 incidence keeps increasing [2-5] worldwide. At the specific country level, for instance, in Australia, between 1982 and 2009, poultry meat has increased by 105%, but red meat has decreased by 22%. However, during this period, the PC61 incidence rate increased from 79.4 (per 100,000) in 1982 to 193.9 (per 100,000) in 2009 [72].

Our data showed that both white meat intake and red meat intake were in strong and significant correlation to PC61 in Pearson r ($r=0.515$, $p<0.001$ and $r=0.531$, $p<0.001$ respectively) and non-parametric correlations ($r=0.560$, $p<0.001$ and $r=0.551$, $p<0.001$ respectively) (Table 6). However, only white meat intake, instead of red meat was significantly correlated to PC61 when ageing, GDP, Is, obesity and urbanization were statistically kept constant ($r=0.337$, $p<0.001$) (Table 6). Interestingly, when we incorporated red meat as the confounding factor, white meat intake was still significantly correlated to PC61 ($r=0.384$, $p<0.001$) (Table 6). This suggested that, if we consume both white and red meat, white meat may be able to contribute to PC61 when we remove the influence of red meat intake on PC61. To the best of our knowledge, statistically, this finding has not been reported by other studies.

Table 6 Pearson r, nonparametric and partial correlations of prostate cancer incidence to white and red meat respectively

| | Pearson r | | Spearman rho | | Partial | | Partial | |
|---|---|---|---|---|---|---|---|---|
| | r | n | r | n | r | n | r | n |
| White meat | 0.515*** | 163 | 0.560*** | 163 | 0.337*** | n=150 | 0.3484*** | n=149 |
| Red meat | 0.531*** | 163 | 0.551*** | 163 | 0.092 | n=150 | - | - |
| Ageing | 0.555*** | 163 | 0.587*** | 163 | - | - | - | - |
| GDP | 0.529*** | 157 | 0.573*** | 157 | - | - | - | - |
| $I_s$ | 0.274*** | 161 | 0.565*** | 161 | - | - | - | - |
| Obesity % | 0.489*** | 161 | 0.501*** | 161 | - | - | - | - |
| URBAN | 0.470*** | 163 | 0.516*** | 163 | - | - | - | - |

Pearson r, nonparametric and partial correlations were reported. Significance levels: * P <0.05, ** P< 0.01, *** P< 0.001.

White meat (poultry) intake (kg/capita/year) sourced from the Food and Agriculture Organization, and red meat intake (kg/capita/year) was calculated through subtracting white meat from total meat intake; Ageing (percent of males ages 65 and above) and GDP PPP (gross domestic product converted to international dollars using purchasing power parity rates) and urbanization (the percent of males living in urban areas) were sourced from the World Bank. Male obesity prevalence (percent of males aged 18+ with BMI ≥ 30 kg/m2); $I_s$ was extracted from previous publications.

White meat intake was placed second increasing $R^2$ to 0.363 from 0.315 with ageing selected as the variable having the greatest influence on PC61 ($R^2$ = 0.315) in Stepwise multiple regression analysis. When we replaced white meat intake with red meat intake as the independent variable, red meat was not selected as the most influential predictor of PC61.

Although, statistically, we found that white meat intake may be a major predictor of PC61, it may not be proper to conclude that white meat intake is a major predictor of PC61, while red meat intake is not, considering the similarities between white meat and red meat (see above for details) and the controversial and circumstantial findings in previous studies.

A cohort study based on the dietary habits of 917 subjects with PC61 concluded that there were no association between chicken intake and the risk of aggressive prostate cancer [73]. This result may not conflict with our finding that white meat was a major and independent predictor of PC61 because of a couple differences in study designs: 1) Only chicken which is main component, but not all, of white (poultry) meat. Our study included all the meat from poultry. 2) The research subjects in this study were PC61 patients, but

our study chose all the males. 3) Cooked chicken was used as the independent variable in the previous study, but poultry flesh was included as the independent variable in our study.

The association between processed meat intake and PC61 has been tentative [32, 74]. Processed meat is usually composed of both red meat and white meat [75]. Therefore, this may support that white meat also contributes to PC61. There may be several issues with those studies. Firstly, the cariogenic effects of processing aids, such as sodium nitrite (E250) on PC61 were (or could) not be removed from the association between processed meat intake and PC61. Secondly, the total processed meat, such as hotdogs and sausages, instead of pure meat were included for study. Therefore, the quality of the data may be questionable. Similarly, statistically, the influence of unprocessed meat intake on PC61 was not removed from the association between processed meat and PC61. This may be the defect in these studies as well.

A recent study conducted by Murphy *et al*. concluded that, due to similarities between pork, beef and chicken diets, people on these three diets for three months did not have different changes of the Body Mass Index (BMI) or any other marker of adiposity [31]. Similarly, another study did not deem that it was necessary to differentiate meat into different categories for investigating the relationship between meat intake and obesity [19].

Categorizing meats and associating some meat types, such as red meat and processed meat, with detrimental health effects in the different circumstances is not supported by the health eating guideline published by the authorities from different country governments, such as Australia [76, 77], Canada [78], Europe [77] and United States [79]. One of the reasons may be that the conclusions from these studies are still controversial and not convincing enough.

We have to point out a strong advantage of this study. This study does not list any circumstance for the existing relationship between total meat intake and PC61. The majority of the previous studies categorized meats for investigating the association between specific meat groups, such as red meat, and PC61 in the specific circumstances [80]. However, generally, people do not eat individual meats but rather meats in combination in broad circumstances [19, 81]. We used the total meat intake, defined as the “flesh of animals used for food”, as the independent variable in this study [68, 82]. The cooking methods, processing methods or nutritional function were not used to differentiate meat types. However, previous studies always listed one or more circumstances (categorizing meat) when the relationship between meat intake and PC61 was investigated. The circumstances may include, but limited to, level of doneness [53,

64, 66, 80], myoglobin content (red and white meat) [46, 83, 84], modification methods (processed meat) [83, 85], ethnicity [85-87] and stage of the PC61 [80, 84, 85]. The definitions of these circumstances varied greatly and were not crystal clear. These ambiguous circumstances may have produced the controversial relationships between specific meat intake and PC61 [87]. Without any circumstance, the relationship between total meat intake and PC61 identified in our study may offer the new insight into the study of the adverse health effects of meat intake [81].

There have been a couple of investigations into the relationship between total meat intake and PC61 risk. John *et al.* concluded that total meat intake was not associated with the risk of advanced prostate cancer [85]. Compared with our findings, the relationship in this study is very circumstantial because the results were based on the specific ethnicity (Non-Hispanic White and African-American men) and specified cooking methods and degree of doneness and stage of PC61 (advanced). In addition to the multiple circumstances included in this study, the data on total meat intake may be biased because newly diagnosed PC61 patients were included as the main research subjects in this study. These patients may more easily recall the negative life events (total meat intake) which have been considered as PC61 risks [88-92]. Our results were in agreement with the findings reported by Koutros *et al.* that total meat was in weak association with the increased risk of incident PC61 and increased risk of advanced PC61 although in this study “well or very well done total meat” was indicated as the independent variable [80].

Several limitations in this study need to be declared. Firstly, the total meat intake data analysed were calculated for per capita in each country. Therefore, the relationship between meat intake and PC61 may only be demonstrated at a country level, which does not necessarily correspond to the same relationship holding true at the individual level. Furthermore, the general market availability of total meat, not the actual human consumption, were tracked for this study. We could not be able to access the direct measures of actual meat consumed by humans as we did not have the data to measure food wastage and provide actual meat intake at country level. Secondly, we included ageing, GDP, magnitude of PC61 accumulation, obesity and urbanization as the potential confounding variables in partial correlation analysis, but other confounding factors may still have influenced the associations reported in this study. For instance, meat intake varies worldwide due to availability, cultural beliefs or religious preferences. However, we could not locate and include other variables as the confounding factor in this study. Thirdly, the PC61 incidence rate was extracted from the GLOBOCAN database. It is probable that datasets from developing countries are less complete than those from developed countries due to issues of underdiagnoses. We attempted to

remove the different levels of PC61 diagnoses through controlling for GDP and urbanization, but this removal might not be sufficient. Fourthly, total meat ("flesh of animals") was used as the independent predictor of PC61 in this study. However, it is constantly reported that specific types, cooking methods, doneness levels and processing methods of meat may be the factors which make meat contribute to PC61.

## Conclusion

Per capita total meat (flesh of animals) consumption may be an independent predictor of PC61 incidence at a global level. Major shifts in dietary habits featured with more meat intake should be investigated globally to determine its adverse health effects. It is novel to include total meat as the predictor of the worldwide non-communicable disease epidemic. This study creates avenues for further study into the subject with exposure based longitudinal cohort studies.

### Declarations

### Ethics approval and Availability of data

All the data used in this study were freely downloaded from the United Nations (UN) agencies' websites. No ethical approval or written informed consent for participation was applicable.

### Competing interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Authors' contributions

WY and MH conceived the idea for this study. WY extracted the data, and MH and WY analysed and interpreted the data. WY reviewed the literature and drafted the manuscript. WY and MH edited and approved the manuscript for submission to the journal.

### Acknowledgments

This research was supported by the Mäxi Foundation, Zürich, Switzerland.

## References

1. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* Int J Cancer, 2015. **136**(5): p. E359-86.
2. Jemal, A., et al., *Global cancer statistics.* CA Cancer J Clin, 2011. **61**(2): p. 69-90.
3. Stewart, B.W., *World Cancer Report 2014*. 2014, Lyon: Lyon, FRA: International Agency for Research on Cancer.
4. Ferlay, J., et al., *Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008.* Int J Cancer, 2010. **127**(12): p. 2893-917.

5. International Agency for Research on Cancer of WHO. *GLOBOCAN Cancer Fact Sheets: Prostate Cancer*. 2016 01.05.2016]; Available from: http://globocan.iarc.fr.

6. Grönberg, H., *Prostate cancer epidemiology.* The Lancet, 2003. **361**(9360): p. 859-864.

7. Hsing AW and C. AP, *Prostate cancer epidemiology.* 2006. **11**: p. 1388-413.

8. Steinberg, G.D., et al., *Family history and the risk of prostate cancer.* The prostate, 1990. **17**(4): p. 337-347.

9. Lichtenstein, P., et al., *Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland.* New England journal of medicine, 2000. **343**(2): p. 78-85.

10. You, W. and H. M, *Cancer incidence increasing globally: The role of relaxed natural selection.* Evol Appl., 2017. **00:1–13**.

11. You, W.-P. and M. Henneberg, *Type 1 diabetes prevalence increasing globally and regionally: the role of natural selection and life expectancy at birth.* BMJ Open Diabetes Research & Care, 2016. **4**(1): p. e000161.

12. Budnik, A. and M. Henneberg, *Worldwide increase of obesity is related to the reduced opportunity for natural selection.* PloS one, 2017. **12**(1): p. e0170098.

13. Shavers, V.L., W. Underwood, and R.P. Moser, *Race/ethnicity and the perception of the risk of developing prostate cancer.* American journal of preventive medicine, 2009. **37**(1): p. 64-67.

14. Campisi, J., *Cancer and ageing: rival demons?* Nature Reviews Cancer, 2003. **3**(5): p. 339-349.

15. Majeed, A., et al., *Trends in prostate cancer incidence, mortality and survival in England and Wales 1971–1998.* BJU International, 2000. **85**(9): p. 1058-1062.

16. Calle, E.E., et al., *Overweight, obesity, and mortality from cancer in a prospectively studied cohort of US adults.* N Engl j Med, 2003. **2003**(348): p. 1625-1638.

17. Schuurman, A.G., et al., *Anthropometry in relation to prostate cancer risk in the Netherlands Cohort Study.* American journal of epidemiology, 2000. **151**(6): p. 541-549.

18. You, W. and M. Henneberg, *Meat in Modern Diet, Just as Bad as Sugar, Correlates with Worldwide Obesity: An Ecological Analysis.* J Nutr Food Sci **6: 517**(4).

19. You, W. and M. Henneberg, *Meat consumption providing a surplus energy in modern diet contributes to obesity prevalence: an ecological analysis.* BMC Nutrition, 2016. **2**(1).

20. Allender, S., et al., *Quantification of urbanization in relation to chronic diseases in developing countries: a systematic review.* J Urban Health, 2008. **85**(6): p. 938-951.

21. Baade, P.D., et al., *Urban- rural differences in prostate cancer outcomes in Australia: what has changed?* The Medical journal of Australia, 2011. **194**(6): p. 293.

22. The IARC. *GLOBOCAN Cancer Fact Sheets: prostate cancer*. 2018; Available from: http://globocan.iarc.fr/old/FactSheets/cancers/prostate-new.asp.

23. Püssa, T., *Toxicological issues associated with production and processing of meat.* Meat science, 2013. **95**(4): p. 844-853.

24. Doll, R. and R. Peto, *The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today.* Journal of the National Cancer Institute, 1981. **66**(6): p. 1191.

25. Key, T.J., *Fruit and vegetables and cancer risk.* British Journal of Cancer, 2010. **104**(1): p. 6.

26. Wang, R.-J., et al., *Dietary fiber, whole grains, carbohydrate, glycemic index, and glycemic load in relation to risk of prostate cancer.(ORIGINAL RESEARCH)(Report).* 2015. **8**: p. 2415.

27. Vasundara, V. and H.K. Laurence, *Diet and prostate cancer: mechanisms of action and implications for chemoprevention.* Nature Reviews Urology, 2010. **7**(8): p. 442.

28. Ma, R.L. and K. Chapman, *A systematic review of the effect of diet in prostate cancer prevention and treatment.* Journal of human nutrition and dietetics, 2009. **22**(3): p. 187-199.

29. Gathirua-Mwangi, W.G. and J. Zhang, *Dietary Factors and Risk of Advanced Prostate Cancer.* European journal of cancer prevention: the official journal of the European Cancer Prevention Organisation (ECP), 2014. **23**(2): p. 96.

30. National Cancer Institute. *Chemicals in Meat Cooked at High Temperatures and Cancer Risk - National Cancer Institute*. 2018; Available from: https://www.cancer.gov/about-cancer/causes-prevention/risk/diet/cooked-meats-fact-sheet.

31. Murphy, K.J., et al., *A comparison of regular consumption of fresh lean pork, beef and chicken on body composition: a randomized cross-over trial.* Nutrients, 2014. **6**(2): p. 682-96.

32. Alexander, D.D., et al., *A review and meta-analysis of prospective studies of red and processed meat intake and prostate cancer.* Nutrition journal, 2010. **9**(1): p. 50.

33. Mandair, D., et al., *Prostate cancer and the influence of dietary factors and supplements: a systematic review.* Nutr Metab (Lond), 2014. **11**: p. 30.

34. FAO. *FAOSTAT-Food Balance Sheet*. 2015 [11.26.2015]; Available from: http://faostat3.fao.org/.

35. The World Bank. *Indicators | Data*. 2018; Available from: https://data.worldbank.org/indicator.

36. Saniotis, A. and M. Henneberg, *Evolutionary Medicine and Future of Humanity: Will Evolution Have the Final Word?* Humanities, 2013. **2**(2): p. 278-291.

37. Henneberg, M., *Reproductive possibilities and estimations of the biological dynamics of earlier human populations.* Journal of Human Evolution, 1976. **5**: p. 41-8.

38. Henneberg, M. and J. Piontek, *Biological state index of human groups.* Przeglad Anthropologiczny, 1975. **XLI**: p. 191-201.

39. Zeegers, M., A. Jellema, and H. Ostrer, *Empiric risk of prostate carcinoma for relatives of patients with prostate carcinoma.* Cancer, 2003. **97**(8): p. 1894-1903.

40. Stephan, C.N. and M. Henneberg, *Medicine may be reducing the human capacity to survive.* Medical Hypotheses, 2001. **57**(5): p. 633-37.

41. WHO. *Global Health Observatory, the data repository*. WHO 2015 [11.26.2015]; Available from: http://www.who.int/gho/database/en/.

42. You, W. and M. Henneberg, *Cereal Crops Are not Created Equal: Wheat Consumption Associated with Obesity Prevalence Globally and Regionally.* AIMS Public Health, 2016. **3(2)**: p. 313-328.

43. Siervo, M., et al., *Sugar consumption and global prevalence of obesity and hypertension: an ecological analysis.* 2014. **17**(3): p. 587-596.

44. WHO. *WHO regional offices*. [11.26.2015]; Available from: http://www.who.int.

45. The World Bank. *Country and Lending Groups | Data*. 2015 11.26.2015]; Available from: http://data.worldbank.org/about/country-and-lending-groups.

46. International Agency for Research on Cancer, *IARC Monographs evaluate consumption of red meat and processed meat*. 2015: Lyon, France.

47. Sinha, R., et al., *Meat and meat- related compounds and risk of prostate cancer in a large prospective cohort study in the United States.* American Journal of Epidemiology, 2009. **170**(9): p. 1165.

48. Kazerouni, N., et al., *Analysis of 200 food items for benzo [a] pyrene and estimation of its intake in an epidemiologic study.* Food and chemical toxicology, 2001. **39**(5): p. 423-436.

49. Knize, M., et al., *Heterocyclic amine content in restaurant-cooked hamburgers, steaks, ribs, and chicken.* Journal of Agricultural and Food Chemistry, 1998. **46**(11): p. 4648-4651.

50. Knize, M., et al., *Heterocyclic amine content in fast-food meat products.* Food and Chemical Toxicology, 1995. **33**(7): p. 545-551.

51. Knize, M., et al., *Formation of heterocyclic amine mutagens/carcinogens during home and commercial cooking of muscle foods.* Journal of Muscle Foods, 1996. **7**(3): p. 271-279.

52. Sinha, R., et al., *2-Amino-1-methyl-6-phenylimidazo [4, 5-b] pyridine, a carcinogen in high-temperature-cooked meat, and breast cancer risk.* Journal of the National Cancer Institute, 2000. **92**(16): p. 1352-1354.

53. Sinha, R., et al., *Heterocyclic amine content of pork products cooked by different methods and to varying degrees of doneness.* Food and Chemical Toxicology, 1998. **36**(4): p. 289-297.

54. Cross, A., J. Pollock, and S. Bingham. *Increased endogenous N-nitrosation in the human colon: a response to red and white meat?* in *BRITISH JOURNAL OF CANCER*. 2000. CHURCHILL LIVINGSTONE JOURNAL PRODUCTION DEPT, ROBERT STEVENSON HOUSE, 1-3 BAXTERS PLACE, LEITH WALK, EDINBURGH EH1 3AF, MIDLOTHIAN, SCOTLAND.

55. Cross, A.J. and R. Sinha, *Meat-related mutagens/carcinogens in the etiology of colorectal cancer.* Environmental and molecular mutagenesis, 2004. **44**(1): p. 44-55.

56. Hughes, R., et al., *Dose-dependent effect of dietary meat on endogenous colonic N-nitrosation.* Carcinogenesis, 2001. **22**(1): p. 199-202.

57. Cross, A.J., et al., *Iron and colorectal cancer risk in the α-tocopherol, β-carotene cancer prevention study.* International journal of cancer, 2006. **118**(12): p. 3147-3152.

58. Lewin, M.H., et al., *Red meat enhances the colonic formation of the DNA adduct O6-carboxymethyl guanine: implications for colorectal cancer risk.* Cancer research, 2006. **66**(3): p. 1859-1865.

59. Cross, A.J., J.R. Pollock, and S.A. Bingham, *Haem, not protein or inorganic iron, is responsible for endogenous intestinal N-nitrosation arising from red meat.* Cancer Research, 2003. **63**(10): p. 2358-2360.

60. Cross, A., J. Pollock, and S. Bingham, *Red meat and colorectal cancer risk: the effect of dietary iron and haem on endogenous N-nitrosation.* IARC scientific publications, 2002. **156**: p. 205.

61. Grant, W.B., *An ecological study of cancer mortality rates including indices for dietary iron and zinc.* Anticancer research, 2008. **28**(3B): p. 1955-1963.

62. Babio, N., et al., *Association between red meat consumption and metabolic syndrome in a Mediterranean population at high cardiovascular risk: cross-sectional and 1-year follow-up assessment.* Nutrition, Metabolism and Cardiovascular Diseases, 2012. **22**(3): p. 200-207.

63. De Nunzio, C., et al., *The Correlation Between Metabolic Syndrome and Prostatic Diseases.* European Urology, 2011.

64. Gu, D., et al., *A comprehensive approach to the profiling of the cooked meat carcinogens 2-amino-3, 8-dimethylimidazo [4, 5-f] quinoxaline, 2-amino-1-methyl-6-phenylimidazo [4, 5-b] pyridine, and their metabolites in human urine.* Chemical research in toxicology, 2010. **23**(4): p. 788-801.

65. Sinha, R., et al., *Heterocyclic amine content in beef cooked by different methods to varying degrees of doneness and gravy made from meat drippings.* Food and Chemical Toxicology, 1998. **36**(4): p. 279-287.

66. Sugimura, T., et al., *Heterocyclic amines: Mutagens/carcinogens produced during cooking of meat and fish.* Cancer Science, 2004. **95**(4): p. 290-299.

67. Pearce, K.L., H.C. Norman, and D.L. Hopkins, *The role of saltbush-based pasture systems for the production of high quality sheep and goat meat.* Small Ruminant Research, 2010. **91**(1): p. 29-38.

68. Lawrie, R.A. and D.A. Ledward, *Lawrie's meat science (7th ed.)*. 2006, Cambridge: Woodhead Publishing Limited.

69. Ferlay, J., et al. *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]*. 2013 28.05.2016]; Available from: http://globocan.iarc.fr.

70. Levere, J., *Advertising*. 2005: New York, N.Y. p. C5.

71. Henchion, M., et al., *Meat consumption: Trends and quality matters.* Meat Science, 2014. **98**(3): p. 561-568.

72. Food and Agriculture Organization. *FAOSTAT: Food Supply - Livestock and Fish Primary Equivalent*. 2018; Available from: http://www.fao.org/faostat/en/#data/CL.

73. Amin, M., et al., *Dietary habits and prostate cancer detection: a case–control study.* Canadian Urological Association Journal, 2008. **2**(5): p. 510.

74. World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR). *Food, nutrition, physical activity, and the prevention of cancer: A global perspective*. 2007 12.05.2015; Available from: http://www.dietandcancerreport.org/cancer_resource_center/downloads/Second_Expert_Report_full.pdf

75. Pearson, A.M. and T.A. Gillett, *Processed meats*. 2012: Springer.

76. National Health and Medical Research Council and Department of Health(Australian Government). *The Five Food Groups | Eat For Health*. 2012 2012-09-24T11:25+10:00 14 January 2018]; Available from: https://www.eatforhealth.gov.au/food-essentials/five-food-groups.

77. The European Food Information Council. *Food-Based Dietary Guidelines in Europe: (EUFIC)*. 2018 14 January 2018]; Available from: http://www.eufic.org/en/healthy-living/article/food-based-dietary-guidelines-in-europe.

78. Government of Canada-Health-Food and Nutrition-Healthy Eating. *Build a healthy meal: use the Eat Well Plate*. 2015 2015-06-12 14 January 2018]; Available from: http://www.healthycanadians.gc.ca/eating-nutrition/healthy-eating-saine-alimentation/tips-conseils/interactive-tools-outils-interactifs/eat-well-bien-manger-eng.php.

79. USDA. *All about the Protein Foods Group*. 2015 2015-02-23 14 January 2018]; Available from: https://www.choosemyplate.gov/protein-foods.

80. Koutros, S., et al., *Meat and meat mutagens and risk of prostate cancer in the Agricultural Health Study.* Cancer Epidemiology and Prevention Biomarkers, 2008. **17**(1): p. 80-87.

81. Tantamango-Bartley, Y., et al., *Are strict vegetarians protected against prostate cancer?* The American journal of clinical nutrition, 2015. **103**(1): p. 153-160.

82. The FAO. *Products from Slaughtered Animals*. 2018 22 January 2018].

83. Wolk, A., *Potential health hazards of eating red meat.* Journal of internal medicine, 2017. **281**(2): p. 106-122.

84. Joshi, A.D., et al., *Red meat and poultry, cooking practices, genetic susceptibility and risk of prostate cancer: results from a multiethnic case–control study.* Carcinogenesis, 2012. **33**(11): p. 2108-2118.

85. John, E.M., et al., *Meat consumption, cooking practices, meat mutagens, and risk of prostate cancer.* Nutrition and cancer, 2011. **63**(4): p. 525-537.

86. Sarwar, M.H., et al., *The importance of cereals (Poaceae: Gramineae) nutrition in human health: A review.*

87. Richman, E.L., et al., *Egg, red meat, and poultry intake and risk of lethal prostate cancer in the prostate-specific antigen-era: incidence and survival.* Cancer Prevention Research, 2011.

88. Bai, A., et al., *A survey of overall life satisfaction and its association with breast diseases in Chinese women.* Cancer Med, 2016. **5**(1): p. 111-9.

89. Courtney, J.G., et al., *Stressful Life Events and the Risk of Colorectal Cancer: Results from a case-control study.* Epidemiology, 1993. **138**(8): p. 628-628.

90. Cohen, L.H., L.C. Towbes, and R. Flocco, *Effects of Induced Mood on Self-Reported Life Events and Perceived and Received Social Support.* Journal of Personality and Social Psychology, 1988. **55**(4): p. 669-674.

91. Brett, J.F., et al., *Negative Affectivity and the Reporting of Stressful Life Events.* Health Psychology, 1990. **9**(1): p. 57-68.

92. Blaney, P.H., *Affect and Memory: A Review.* Psychological Bulletin, 1986. **99**(2): p. 229-246.

Article 7/10: Cereal Crops Are not Created Equal: Wheat Consumption Associated with Obesity Prevalence Globally and Regionally (Published at AIMS Public Health, 2016)

Wenpeng You[1], Maciej Henneberg[1,2]

1. Biological Anthropology and Comparative Anatomy Unit, School of Medicine, the University of Adelaide, Adelaide, SA, Australia

2. Institute of Evolutionary Medicine, University of Zürich, Zürich, Switzerland

**Published:** Wenpeng You, Maciej Henneberg. *Cereal Crops Are not Created Equal: Wheat Consumption Associated with Obesity Prevalence Globally and Regionally*; AIMS Public Health, 2016, 3(2): 313-328. doi: 10.3934/publichealth.2016.2.313

✉ **Correspondence**: Wenpeng You wenpeng.you@adelaide.edu.au

## Contextual Statement

Wheat protein (gluten) has been associated with metabolic syndrome, which may increase body weight. Wheat, rice and maize are the three major cereal crops. Rice and maize have already been advocated as the body weight control dietary components. However, cereals, including wheat, have also been suggested as the healthy food group in terms of body weight management.

We postulated and tested that wheat consumption may have been contributing to obesity worldwide in a hidden way because wheat's contributing effects may have been balanced by the other two cereal crops, rice and maize, which are beneficial for body weight loss.

## Statement of Authorship

# Statement of Authorship

| | |
|---|---|
| Title of Paper | Cereal Crops Are not Created Equal: Wheat Consumption Associated with Obesity Prevalence Globally and Regionally |
| Publication Status | ☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Publication Style |
| Publication Details | Wenpeng You, Maciej Henneberg. *Cereal Crops Are not Created Equal: Wheat Consumption Associated with Obesity Prevalence Globally and Regionally*; AIMS Public Health, 2016, 3(2): 313-328. doi: 10.3934/publichealth.2016.2.313 |

### Principal Author

| | | | |
|---|---|---|---|
| Name of Principal Author (Candidate) | Wenpeng You | | |
| Contribution to the Paper | Conceived the hypothesis, collected data, performed analysis on all samples, interpreted data, wrote manuscript and acted as corresponding author. | | |
| Overall percentage (%) | 60 | | |
| Signature | | Date | 22/12/2017 |

### Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

| | | | |
|---|---|---|---|
| Name of Co-Author | Maciej Henneberg | | |
| Contribution to the Paper | Supervised development of work, formalized and interpreted data analysis, and evaluated and edited manuscript | | |
| Signature | | Date | 22/12/2017 |

| | | | |
|---|---|---|---|
| Name of Co-Author | NA | | |
| Contribution to the Paper | | | |
| Signature | | Date | |

Please cut and paste additional co-author panels here as required.

## Abstract

**Background:** Cereals have been extensively advocated as the beneficial food group in terms of body weight management, but each staple cereal crop may contribute in different ways. Studies of the association between wheat availability and risk of obesity are controversial. This study aimed to test the global and regional association between wheat availability as reported by FAO and obesity prevalence at a population level. FAO does not distinguish between whole grain wheat and refined wheat.

**Methods:** Population-specific data from 170 countries on prevalence of obesity, availabilities of mixed cereals, wheat, rice, maize, meat, sugar, fat, soy and calories and GDP are obtained from the UN agencies. All variables were measured as per capita per day (or per year). Each country is treated as an individual subject. SPSS v. 22 is used to analyse these data for all the 170 countries and official country groupings (regions) using non-parametric and parametric correlations, including partial correlation analysis.

**Results:** Pearson's correlation coefficient analysis showed that obesity prevalence was positively associated with wheat availability ($r=0.500$, $p<0.001$), but is inversely associated with availabilities of total cereals ($r=-0.132$, $p=0.087$), rice ($r=-0.405$, $p<0.001$) and maize ($r= -0.227$, $p=0.004$). These associations remain in partial correlation model when we keep availabilities of meat, fat, sugar, soy and caloric intake and GDP statistically constant. Overall, positive association between wheat availability and obesity prevalence remain in different regions. Maize and mixed cereal availabilities do not show independent associations with the obesity prevalence.

**Conclusions:** Our study suggests that wheat availability is an independent predictor of the obesity prevalence both worldwide and with special regard to the regions of Africa, Americas and Asia.

**Key Words:** Rice, Maize, Correlation, Cereals, Ecological Study

## Background

Obesity is a serious global public health problem that needs to be urgently addressed among all populations (1-3). Obesity increases mortality and morbidity risk from various chronic diseases, such as cardiovascular diseases, diabetes, certain types of cancers and musculoskeletal disorders (4). Despite progress in knowledge of reasons of obesity, some causes for the obesity epidemic and the disparities between population groups are still unclear.

Obesity is caused by a complex interaction between the environment, genetic predisposition and human behaviour (5). Diet habits have been implicated in the

development of obesity as they may bring environmental factor exposures to people, and this relationship between environmental factor exposure and obesity is complex and not completely understood (2, 6). Food components such as soy products and sugar in our diets were postulated to contribute to obesity (7) and diabetes (8, 9) respectively in additioon to a number of factors, such as physical activities, diet composition and genetics. Mixed cereals are important sources of many nutrients including dietary fiber, resistant starch, oligosaccharides, trace elements, vitamins, and other compounds of interest in disease prevention, including phytoestrogens and antioxidants (10). Dietary guidelines recommend the consumption of mixed cereals to prevent chronic diseases and/or their risk factors. For instance, whole-grain? mixed cereals, instead of the individual cereal crop, have been extensively advocated as the major food group for healthy body weight management (11-17), and have been shown their protective role of mixed cereals in reducing the risk of chronic diseases (18) including cancer (19, 20), type 2 diabetes (16, 17) and cardiovascular disease (16, 17).

Wheat makes up a substantial part of the human diet and is the most important food cereal source for humans (21). Due to the adoption of western-style diets, its demand for human consumption is increasing globally, including countries which are climatically unsuited for wheat production (22). In the recent years, the association between wheat intake and body weight management has been debated (17, 22-31). Wheat is provided for human consumption in different forms, principally unrefined or refined. The different forms of wheat products may have different health effects. It has been argued that whole wheat is beneficial for human health (17, 18). A number of studies suggested that wheat consumption contributes to obesity prevalence in several ways including its use in energy dense and refined products. It has been suggested that wheat protein (gluten) may develop metabolic syndrome which may lead to body weight increase (25, 27, 32). In this study, we tested the association between the prevalence of obesity (expressed in percentage of defined population with a body mass index (BMI) of 30 kg/m$^2$ or higher) and wheat availability at the population level on the basis of most recent complete data published by the United Nations (UN) agencies.

## Methods and Materials

### Data sources:

The country specific data were collected for this ecological study:

1) The WHO Global Health Observatory (GHO) data on the estimates of prevalence of adult obesity (percentage of BMI ≥ 30 kg/m2 in country population, 2014).

The GHO is an initiative of the WHO to share data on global health, including statistics by country and information about specific diseases and health measures. The GHO specifically assembles prevalence data of the biological risk factors, including obesity for WHO Member States using standardized protocols (http://www.who.int/gho/ncd/methods/en/).

2) The FAOSTAT data on food availability per capita per day in 2011 of mixed cereals (excluding beer), wheat and products, rice (paddy equivalent), maize and products, total meat, sugar and sweeteners, grand total fat, soy products and grand total calories. These data were abbreviated as cereals, wheat, rice, maize, meat, sugar, fat, soy and total calories respectively in this paper.

The FAOSTAT database disseminates statistical data collected and maintained by the FAO. FAOSTAT data are provided as a time-series from 1961 in most domains through the Food Balance Sheet (FBS, http://faostat3.fao.org/home/E). The FBS presents a comprehensive picture of the pattern of a country's food supply during a specified reference period. The FBS shows for each food item i.e. each primary commodity availability for human consumption which corresponds to the sources of supply and its utilisation. The total quantity of foodstuffs produced in a country added to the total quantity imported and adjusted to any change in stocks that may have occurred since the beginning of the reference period gives the supply available during that period. On the utilisation side a distinction is made between the quantities exported, fed to livestock + used for seed, losses during storage and transportation, and food supplies available for human consumption. The per capita supply of each such food item available for human consumption is then obtained by dividing the respective quantity by the related data on the population actually partaking in it (33). Unfortunately, no separate data on consumption of refinded and unrefined of wheat products are available.

3) The World Bank data on GDP per capita (USD per year, 2011)

The World Bank dataset measures progress on aggregate outcomes for member countries for selected indicators. GDP per capita is gross domestic product divided by midyear population (http://data.worldbank.org/indicator/NY.GDP.PCAP.CD). GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources.

WHO, FAO and the World Bank are intergovernmental organizations using specialized information relevant to their respective fields. Their professional personnel should have evaluated these data in consideration of their possible use, e.g. for scientific research

and decision making, before they were published. Therefore, the data reporting is as free of bias and error as it can be with government statistics. This means that errors are reduced but some inaccuracies related to reporting quality may still be present in the data. Similar data from the same sources were recently used to analyse the relationships between nutrients and obesity (7, 34) and diabetes (8, 9, 35) in a number of publications.

**Criteria for data inclusion**

The data were selected in consideration of their fulfilment of 1) completeness of data across all analysed variables, 2) the most updated and recent datasets available, 3) major food types that were indicated in the literature to have relationships with obesity, specifically: wheat, rice, maize, meat, fat and sugar. For instance, barley and rye also contain gluten like wheat (36), but we did not include them in our study due to their extremely low availabilities in limited areas in the world. Following these conditions, country-level data on obesity prevalence in 2014, cereal availability in 2011 (mixed cereals, wheat, rice, and maize), and potential confounders (meat, sugar, fat, soy, total calories and GDP) in 2011 were matched. We backdated variables and potential confounders to 2011 to reflect exposure with delayed obesity presentation in 2014. The rationale for this decision is that studies have shown that three years is a practical period to develop obesity and metabolic syndrome after exposure to dietary risks (i.e., high intake of wheat today does not lead to immediate obesity) (37-39).

In order to contrast the association between wheat availability and obesity prevalence and availability of other cereal crops, we used the availability data of mixed cereals and the other two staple food cereal crops, rice and maize for comparative analysis.

All the aforementioned data were freely downloaded from the UN agencies websites. No ethical approval or written informed consent for participation was required.

**Data analysis**

We obtained data for 170 countries that had information required for both obesity and wheat availability in a uniform format. Each country was treated individually and all their availability for other variables information was analysed. In this paper the variables and

confounders were only referred to with their abbreviations instead of full names followed by their units. For particular analyses, the number of countries included may have differed somewhat because all information on other variables was not uniformly available for all countries due to unavailability from relevant UN agencies. The minimum sample size is 148 for correlation with soy availability. All the data were extracted and saved in Microsoft Excel® for analysis.

Human diet patterns varying in different food components may be affected by the types of food availability in a particular region, socio-economic status and cultural beliefs. In order to demonstrate that association between obesity prevalence and wheat availability is universal regardless of these factors, countries were grouped for correlation analyses. The criteria for grouping countries are UN macro geographical regions, the World Bank income classifications, WHO regions, countries sharing specific characteristics like geography, culture, development role or socio-economic status, like Latin America and the Caribbean (LAC), Organisation for Economic Co-operation and Development (OECD), Asia-Pacific Economic Cooperation (APEC), Southern African Development Community (SADC), the Arab World, Latin America (LA), European Union (EU) and Asia Cooperation Dialogue (ACD). All the country listings are sourced from their official websites for matching except LA which is self-classified based on region primarily speaking romance languages. Countries included in LA are Argentina, Bolivia (Plurinational State of), Brazil, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Haiti, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Uruguay, and Venezuela (Bolivarian Republic of).

Table 1 Sources of country grouping criteria and correlation between obesity prevalence and wheat availability in different country groupings

| Country groupings | Pearson's | | Nonparametric | | Source of country grouping criteria |
|---|---|---|---|---|---|
| | r | Sig. | rho | Sig. | |
| Worldwide (n=170) | 0.500 | <0.001 | 0.555 | <0.001 | http://who.int/en & http://faostat3.fao.org |
| UN macro geographical regions | | | | | |
| Africa (n=47) | 0.790 | <0.001 | 0.767 | <0.001 | http://unstats.un.org |
| Americas (n=35) | 0.518 | 0.001 | 0.604 | <0.001 | http://unstats.un.org |
| Asia (n=42) | 0.616 | <0.001 | 0.639 | <0.001 | http://unstats.un.org |
| Europe (n=39) | 0.222 | 0.174 | 0.222 | 0.173 | http://unstats.un.org |
| Oceania (n=7) | -0.053 | 0.911 | 0.214 | 0.645 | http://unstats.un.org |
| Sub-continents within UN macro geographic regions | | | | | |
| Sub-Africa: SADC (n=14) | 0.633 | 0.015 | 0.770 | 0.001 | http://www.sadc.int |
| Sub-Asia: ACD (n=26) | 0.592 | 0.001 | 0.729 | <0.001 | http://www.acddialogue.com |
| Sub-Americas: LA (n=20) | 0.567 | 0.009 | 0.502 | 0.024 | Self-classified based on region primarily speaking romance languages |
| Sub-Europe: EU (n=28) | 0.283 | 0.144 | 0.298 | 0.124 | http://europa.eu |
| World Bank income classifications | | | | | |
| Low (n=32) | 0.219 | 0.228 | 0.196 | 0.282 | http://data.worldbank.org |
| Low middle (n=42) | 0.307 | 0.048 | 0.544 | <0.001 | http://data.worldbank.org |
| Upper middle (n=48) | 0.257 | 0.078 | 0.196 | 0.181 | http://data.worldbank.org |
| High (n=48) | 0.196 | 0.181 | 0.076 | 0.608 | http://data.worldbank.org |
| WHO regions | | | | | |
| AFRO (n=40) | 0.679 | <0.001 | 0.745 | <0.001 | http://www.afro.who.int |
| AMRO (n=35) | 0.518 | 0.001 | 0.604 | <0.001 | http://www.paho.org/hq |
| EMRO (n=15) | 0.285 | 0.252 | *-0.010* | *09.68* | *www.emro.who.int* |
| EURO (n=50) | -0.002 | 0.989 | *0.012* | *0.933* | *www.euro.who.int* |
| SEARO (n=10) | 0.408 | 0.241 | *0.413* | *0.235* | *www.searo.who.int* |
| WPRO (n=17) | 0.455 | 0.067 | *0.613* | *0.009* | *www.wpro.who.int* |
| Various economic and cultural country groupings | | | | | |
| APEC (n=17) | 0.640 | 0.006 | 0.689 | 0.002 | http://www.apec.org |
| Arab World (n=17) | 0.427 | 0.087 | 0.140 | 0.593 | http://data.worldbank.org |
| LAC (n=32) | 0.481 | 0.013 | 0.583 | <0.001 | http://www.unesco.org |
| OECD (n=34) | 0.316 | 0.069 | 0.176 | 0.320 | http://www.oecd.org |

Pearson correlation coefficients and Nonparametric Correlations are reported.

Obesity prevalence is expressed in percentage of defined population with a body mass index (BMI) of 30 kg/m2 or higher. Wheat availability is in g/capita/day.

Abbreviations: LAC, Latin America and the Caribbean, OECD, Organisation for Economic Co-operation and Development; APEC, Asia-Pacific Economic Cooperation; SADC, Southern African Development Community; LA, Latin America; EU, European Union; ACD, Asia Cooperation Dialogue.

SADC, ACD, LA and EU are included as the sub macro UN continents of Africa, Asia, Americas and Europe respectively to further investigate the correlation within the succeeding macro areas. We could not select any small international organization within Oceania due to very limited number of countries for us to access data. In our analysis, we only included those countries for which we could access the data for the specific groupings.

We calculated the standard deviations of wheat availability and obesity prevalence in United Nations macro continents to explore the variation in Pearson coefficients between wheat availability and obesity prevalence due to the different geographic distributions of country groupings.

Pearson's correlation coefficient, non-parametric correlation coefficient (rho) of Speaman were calculated between all selected variables and partial correlation analyses keeping some variables statistically constant were conducted using SPSS v. 22 (SPSS Inc., Chicago Il USA). In this study, significance was kept at the 0.01 level (2-tailed).

## Results

In general, obesity prevalence is in significant positive association with wheat availability ($r=0.500$, $p<0.001$), but inversely with rice availability ($r=-0.405$, $p<0.001$) (Table 2). It is also inversely associated with maize availability ($r=-0.227$, $p=0.004$) and mixed cereals availability ($r=-0.132$, $p=0.087$) (Table 2).

We subsequently performed nonparametric correlations in SPSS (Spearman's "rho") with the same set of data to test whether the Pearson's correlations between obesity prevalence and all variables differ due to potentially abnormally distributed variables (Table 2).

Table 2 Correlation analysis between obesity prevalence and all variables

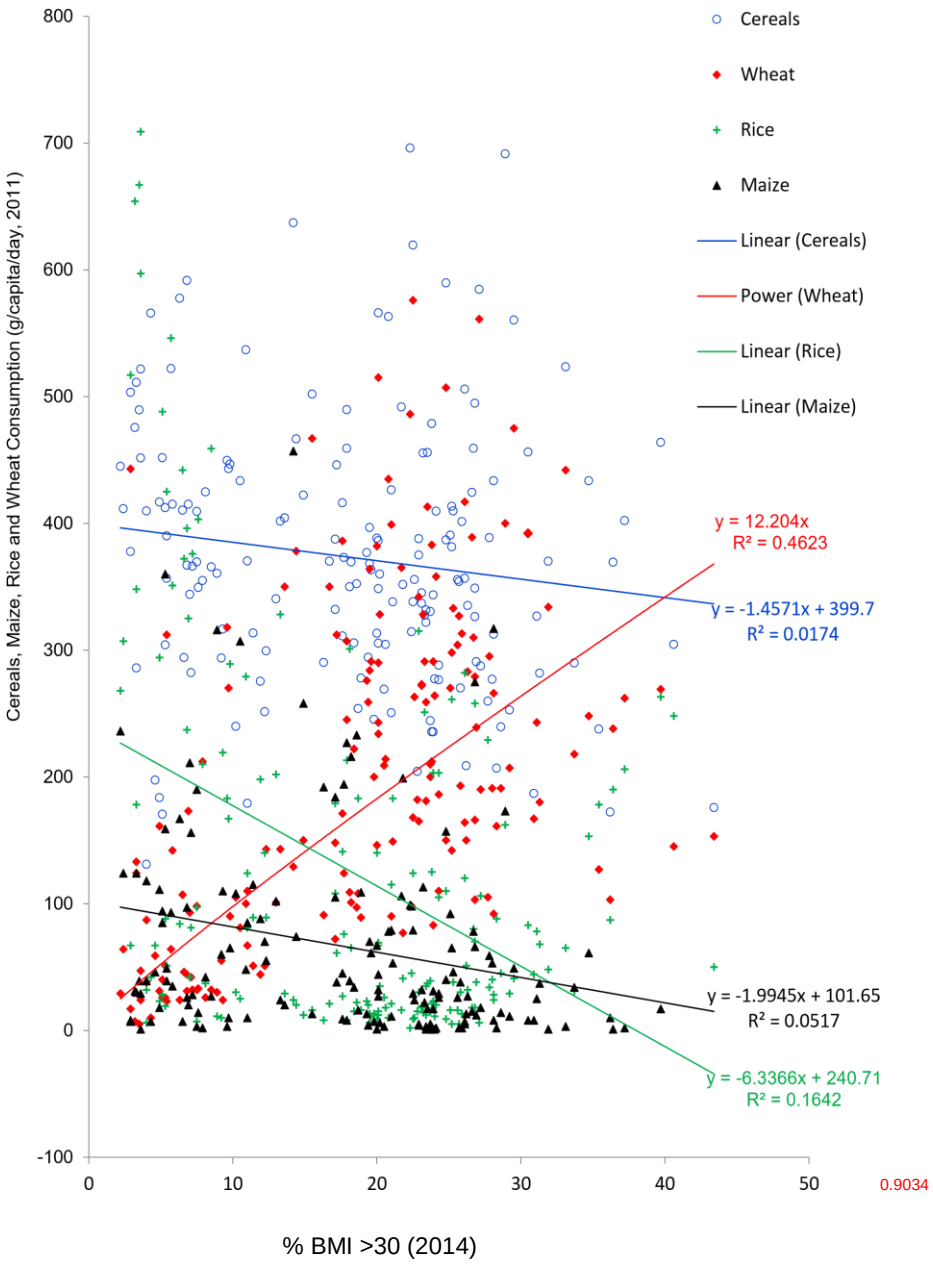
| | BMI≥30 | Total Cereals | Wheat | Rice | Maize | Meat | Total Fat | Sugar | Total Calories | GDP | Soy |
|---|---|---|---|---|---|---|---|---|---|---|---|
| BMI≥30 | 1 | -0.132 | 0.500*** | -0.405*** | -0.227** | 0.624*** | 0.589*** | 0.660*** | 0.593*** | 0.339*** | 0.252*** |
| Cereals | -0.172* | 1 | 0.406*** | 0.164* | 0.216** | -0.336*** | -0.230** | -0.167* | 0.147 | -0.234** | -0.025 |
| Wheat | 0.555*** | 0.288*** | 1 | -0.465*** | -0.347*** | 0.293*** | 0.429*** | 0.380*** | 0.574*** | 0.224** | 0.063 |
| Rice | -0.252*** | 0.052 | -0.473*** | 1 | -0.145 | -0.349*** | -0.392*** | -0.308*** | -0.280*** | -0.246*** | 0.029 |
| Maize | -0.300*** | 0.116 | -0.418*** | -0.021 | 1 | -0.313*** | -0.381*** | -0.260*** | -0.302*** | -0.291*** | -0.038 |
| Meat | 0.637*** | -0.325*** | 0.420*** | -0.321*** | -0.321*** | 1 | 0.811*** | 0.627*** | 0.666*** | 0.648*** | 0.314*** |
| Total Fat | 0.631*** | -0.210** | 0.550*** | -0.363*** | -0.437*** | 0.807*** | 1 | 0.613*** | 0.818*** | 0.731*** | 0.270*** |
| Sugar | 0.664*** | -0.182* | 0.466*** | -0.250*** | -0.337*** | 0.636*** | 0.641*** | 1 | 0.624*** | 0.480*** | 0.343*** |
| Total Calories | 0.603*** | 0.095 | 0.635*** | -0.338*** | -0.314*** | 0.679*** | 0.823*** | 0.627*** | 1 | 0.615*** | 0.313*** |
| GDP | 0.657*** | -0.243** | 0.513*** | -0.287*** | -0.388*** | 0.833*** | 0.819*** | 0.704*** | 0.757*** | 1 | 0.188* |
| Soy | 0.339*** | -0.052 | 0.099 | 0.072 | -0.034 | 0.317*** | 0.317*** | 0.358*** | 0.319*** | 0.428*** | 1 |

Pearson correlation coefficients and non-parametric correlation coefficients (rho) are reported. Number of countries included in the analysis range from 148 to170. *P< 0.05; **P< 0.01; ***P< 0.001

BMI≥30 is for obesity prevalence which is expressed in percentage of defined population with a body mass index (BMI) of 30 kg/$m^2$ or higher. Availabilities of wheat, rice, maize, meat, soy, cereals, sugar and total fat are expressed in g/capita/day. GDP is in per capita USD per year. Total calories is in kcal/capita/day.

Figure 1 presents the relationships between obesity prevalence and each cereal food type. Relationships between obesity prevalence and availabilities of total cereals, rice and maize are linear, and wheat availability shows power relationship with obesity prevalence.

Figure 1 Relationships between obesity prevalence and cereals, maize, rice and wheat availabilities respectively



When the potential confounders which are the availabilities of meat, sugar, fat and soy, the intake of calories and GDP are controlled for partial correlation analysis, obesity prevalence is still in strong positive association with wheat availability ($r=0.368$, $p<0.001$) (Table 3). The association between obesity prevalence and rice availability is relatively strong, but negative ($r=-0.276$, $p=0.001$). No association between cereals availability ($r=0.065$, $p=0.436$) or maize availability ($r=-0.004$, $p=0.963$) is observed (Table 3).

Table 3: Global association between obesity prevalence and total cereals and each cereal food availabilities controlled for several confounders in partial correlation analysis

| Correlation | Cereals | Wheat | Rice | Maize |
|---|---|---|---|---|
| Partial correlation (r) | -0.065 | 0.368 | -0.276 | 0.004 |
| Significance | 0.436 | <0.001 | 0.001 | 0.963 |
| df | 145 | 145 | 145 | 145 |

Partial correlation coefficients are reported. Keeping intake of meats, total fats, sugar, soy, total calories and GDP constant.

Obesity prevalence which is expressed in percentage of defined population with a body mass index (BMI) of 30 $kg/m^2$ or higher. Availabilities of wheat, rice, maize, meat, soy, cereals, sugar and total fat are expressed in g/capita/day. GDP is in per capita USD per year. Total calories is in kcal/capita/day.

The correlation of wheat availability to obesity prevalence in different country groupings is also observed (Table 1). Within the UN macro geographical regions, Africa ($r=0.790$, $p<0.001$), Americas ($r=0.518$, $p=0.001$) and Asia ($r=0.616$, $p<0.001$) have a significant positive association with wheat availability. The association based on Europe region is positive, but not significant. These trends are also observed in Africa sub-grouping (SADC), Americas sub-grouping (LA), Asia sub-grouping (ACD) and Europe sub-grouping (EU) respectively (Table 1).

For the UN macro region of Oceania (Table 1), sample size is small and variation of wheat availability is limited. This renders correlation coefficients uninformative (Table 4).

Table 4 Standard deviations of obesity prevalence and wheat availability in UN macro continents

| Continents | Obesity prevalence | | Wheat availability | |
|---|---|---|---|---|
| | Mean | Std. Deviation | Mean | Std. Deviation |
| Africa (n=47) | 11.75 | 7.36 | 126.13 | 142.21 |
| Americas (n=35) | 24.49 | 5.26 | 154.54 | 65.06 |
| Asia (n=42) | 14.97 | 11.07 | 241.24 | 163.16 |
| Europe (n=39) | 22.14 | 2.92 | 278.92 | 62.19 |
| Oceania (n=7) | 34.47 | 6.19 | 166.57 | 47.15 |

Obesity prevalence is expressed in percentage of defined population with a body mass index (BMI) of 30 kg/m$^2$ or higher. Wheat availability is in g/capita/day.

The association between obesity prevalence and wheat availability exists in all the country groupings categorized by the World Bank based on per capita GDP.

Based on the WHO region classifications, AFRO ($r=0.679$, $p<0.001$) and AMRO ($r=0.518$, $p=0.001$) have the significant correlation between obesity prevalence and wheat availability. However, there is nearly no association in Europe. The similar correlation can be observed in Africa sub-grouping (SADC), Americas sub-grouping (LA) and Europe sub-grouping (EU) respectively. Wheat availability is also positively correlated to obesity prevalence in EMRO and SEARO. Lack of correlation in Europe is a result of small variation in obesity prevalence and wheat availability (Table 4).

The general trend that obesity prevalence is positively associated with wheat availability can be observed in country groupings regardless of cultural backgrounds, economic levels and geographic locations of the clustered countries. The trends are also present in two functional alliances, OECD and APEC although the former comprises developed countries only and the latter is comprised of both developing and developed countries.

## Discussion

The worldwide secular trend of increased obesity prevalence likely has multiple etiologies, which may act through multiple mechanisms. By examining the data collected for 170 countries, we have shown that globally obesity prevalence is significantly associated with wheat availability independent of other food components (total fat, soy products, sugar and meat), total calories and GDP. Although results of ecological analysis must be treated cautiously, our results indicated relationship similar to those

found in three recent empirical surveys in children (31), young women (24) and general adults (23).

Early in history, barley and rye were much more prominent as dietary grains. However, during agriculture modernization and evolution of our culture, wheat has been recognized as the finest grain (40). Wheat has a pleasant flavour, an extensive shelf life and unique properties because of gluten-forming proteins (40). About 95% of the wheat that is grown and consumed globally is bread wheat (*Triticum aestivum*). Bread wheat is a relatively new species, having arisen in southeast Turkey about 9,000 years ago (21, 41). Extensive wheat breeding by modern agricultural techniques, such as seed selection, hybridization and radiation, has aimed to increase crop yield, improve quality, diversify the strains and develop disease and insect resistance and tolerance to abiotic stresses. Modern agriculture techniques have made hereditary factor of wheat changed. These changes of genetic material may have brought new substances in modern wheat in comparison to those from wheat decades ago (25, 30).

Wheats are subjected to many different processes during their preparation for human consumption. Whole wheat products have been generally accepted as food groups that are rich in fiber, micronutrients and minerals. Therefore, it has been argued that whole wheat consumption is beneficial to human health (17, 18). Refined wheat products have been considered as desired food in the past due to its purity, but it contains practically only carbohydrates, which is less beneficial nutritionally (42). Due to their appearance and good taste, food items containing refined wheat may be over consumed. Since we could not obtain separate data on whole wheat and refined wheat consumption, their respective contributions to obesity should be subject of separate study. There are many varieties of wheat gluten proteins which may have structural, metabolic, protective or storage functions (43). Analysis of proteins expressed by a wheat hybrid compared to its two parent strains have demonstrated that 5% of proteins in general (44) and 14% in gluten proteins (45) were present in either parent. From the evolutionary perspective, new types of crops or food components, such as soy (7), when massively introduced into human diet, may be able to change human nutritional environment with the consequence of contributing to obesity prevalence (46).

Gluten proteins are the major storage components in wheat and may account for up to 80% of the total cereal protein (47, 48). Anti-nutrients are natural or synthetic compounds that interfere with nutrient absorption. The gluten complex has been considered as one of the anti-nutrients causing inflammation (25, 30, 32) which has been associated with body weight increase in a number of studies in humans (25, 27, 30) and animals (28). A couple of studies found that gluten consumption was inversely correlated with BMI

increase (27, 49), but the two cohorts in the studies had been under clinical treatments due to celiac disease and Crohn's disease respectively.

In this ecological analysis, wheat contributes to obesity. Wheat has greater energy density than another staple food, rice (23). For instance, eaten in Asia steamed wheat bread doubles the energy from the same amount of steamed rice (50). It has been well established that energy-dense diets increase risk of obesity because they tend to increase total energy intake (51, 52). Since we do not have data on whole wheat and refined wheat availability. It can only be suggested that the observed correlation may be a result of refined grain consumption (53).

European countries are culturally and socio-economically relatively homogenous. In Europe, the correlation coefficient between per capita wheat availability and obesity prevalence does not reach the significant level, though it is still positive in our study. This is most likely due to small variances of wheat availability and obesity prevalence in this region (SD of wheat availability= 62.19, SD of obesity prevalence=2.92, Table 4) that may reduce the co-variance. It may also be that types of wheat products consumed in Europe differ from those consumed in other parts of the world.

Although the association between wheat availability and the obesity prevalence is not significant in the South East Asia (WHO region), it is significant in macro Asia (both UN and WHO definitions) and the Asia sub-ACD. It may be because the obesity prevalence in South East Asia should be assessed with the region specific BMI≥ 28, instead of the universal BMI≥ 30 (54, 55). The universal obesity determining level (BMI≥30) used by the WHO to calculate the obesity prevalence may not be able to determine the actual prevalence of obesity in that area.

In the modern diet, a majority of cereals are refined by the removal of germ, and bran, so that the remaining endosperm is mostly carbohydrate (56). Refined endosperm may be metabolized to satisfy human daily energy requirement earlier than the other two macro-nutrients which are fat and protein (46, 57-59). Globally, wheat, rice and maize supply around 93% of total daily energy from cereals and 50% of all food calories (59, 60). A number of other studies also show that mixed cereals availability is not associated with obesity prevalence. In our study, obesity prevalence has been positively associated with wheat availability, but inversely with both rice and maize availability. These positive and negative associations may have been neutralized which may make mixed cereals appear as the healthy food types for body weight control.

In terms of the association between obesity prevalence and rice availability only, our ecological study findings differ from the results of epidemiological studies in Japanese young women (61), in American Hispanic elders (62) and in Korean adult women (63).

There is a similarity among the three studies that the cohorts were on diet patterns which not only mainly consisted of rice, but also contained significant amount of soy products, such as miso soup, tofu or bean *etc.* Soy products contain anti-nutrients which may alter human metabolism and may contribute to obesity (7). Interestingly, a cohort of Brazilian adults in Sichieri's studies mainly relied on rice and beans being protected against obesity (64). The underlying reason may be that the diet pattern for that particular cohort was that of low fat and low energy, or that too much rice and high fibre from rice and beans overcompensated the soy's effect on causing obesity.

There are a few limitations in this study. Firstly, lack of separate data of whole wheat and refined wheat availability. Secondly, although we attempted to remove confounding effects of variables such as GDP, caloric, fat availability *etc.* by means of partial correlation analysis, some confounding factors may still influence correlation we found. Some residual curvilinearity may remain even though a relationship between all studied variables appeared to be linear as indicated by similarity of values of Pearson and spearman nonparametric correlation coefficients. Secondly, there may be some variables not included in our analysis that influence found correlation. For instance, fruits availability correlated, however, poorly with prevalence of overweight in a similar ecological study (6). Other possible variables to be considered would be vegetables especially starch-rich varieties. Thirdly, we could only use an international food database that tracks the general market availability of different food types, not the actual human consumption. There are no direct measures of actual human consumption that can account for food wastage and provide precise measures of food consumption internationally. As Siervo *et al.* analysed (34), food disappearance data may not reflect precisely food available to household individual for consumption. Disappearance data may overestimate consumption by some ¼ of total amount. Finally, the data analysed are calculated per capita in each country, so we can only demonstrate a relationship between food group availability and obesity at a country level. Within a country, relationships between characteristics of individuals and their diets maybe different due to specific circumstances. —Cohort studies exploring cohort the association between different cereal species consumption and obesity prevalence would be useful.

## Conclusions

Associations between obesity prevalence and wheat availability at country and regional levels suggest that wheat consumption may contribute to obesity. In ecological study, it may not be possible to reach reliable conclusions and therefore, there is a need to explore the association we found by specific studies of the type that can be more conclusive.

### Competing interests

The authors declare no conflict of interest.

### Authors' contributions

WY and MH formulated the hypothesis. WY collected the data and conducted preliminary analyses. MH performed the final statistical analysis and WY constructed tables and figures. WY interpreted the data and drafted the text with contributions from MH. All authors reviewed, edited and approved the final manuscript.

### Acknowledgements

Financial support to MH was provided by the Wood Jones Bequest to the University of Adelaide.

## References

1. Rokholm B, Baker JL, Sørensen TIA. The levelling off of the obesity epidemic since the year 1999 – a review of evidence and perspectives. Oxford, UK2010. p. 835-46.

2. Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developing countries. Nutrition reviews. 2012;70(1):3-21.

3. Stevens G, Singh G, Lu Y, Danaei G, Lin J, Finucane MM, et al. National, regional, and global trends in adult overweight and obesity prevalences. Population Health Metrics. 2012;10.

4. WHO. Obesity: Preventing and Managing the Global Epidemic. Geneva World Health Organization 2000.2004.

5. Nguyen DM, El-Serag H. The Epidemiology of Obesity. Gastroenterol Clin North Am. 2010;39(1):1-7.

6. You W, Henneberg M. Meat consumption providing a surplus energy in modern diet contributes to obesity prevalence: an ecological analysis. BMC Nutrition. 2016;2(1).

7. Roccisano D, Henneberg M. Soy Consumption and Obesity. Food and Nutrition Sciences. 2012;03(02):260-6.

8. Weeratunga P, Jayasinghe S, Perera Y, Jayasena G, Jayasinghe S. Per capita sugar consumption and prevalence of diabetes mellitus – global and regional associations. BMC public health. 2014;14:186-91.

9. Basu S, Yoffe P, Hills N, Lustig RH. The relationship of sugar to population-level diabetes prevalence: an econometric analysis of repeated cross-sectional data. PloS one. 2013;8(2):e57873: 1-8.

10. Slavin JL, Martini MC, Jacobs DR, Marquart L. Plausible mechanisms for the protectiveness of whole grains. The American journal of clinical nutrition. 1999;70(3 Suppl):459S.

11. Cancer Council Australia. Position statement - Fibre, wholegrain cereals and cancer - National Cancer Prevention Policy [Available from: http://wiki.cancer.org.au.

12. De La Hunty A, Ashwell M. Are people who regularly eat breakfast cereals slimmer than those who don't? A systematic review of the evidence. Oxford, UK2007. p. 118-28.

13. de la Hunty A, Gibson S, Ashwell M. Does Regular Breakfast Cereal Consumption Help Children and Adolescents Stay Slimmer? A Systematic Review and Meta- Analysis. Obesity Facts. 2013;6(1):70-85.

14. Bazzano LA, Song Y, Bubes V, Good CK, Manson JE, Liu S. Dietary intake of whole and refined grain breakfast cereals and weight gain in men. Obesity research. 2005;13(11):1952.

15. van de Vijver LPL, van den Bosch LMC, van den Brandt PA, Goldbohm RA. Whole- grain consumption, dietary fibre intake and body mass index in the Netherlands cohort study. European journal of clinical nutrition. 2009;63(1):31.

16. Cho SS, Qi L, Fahey GC, Jr., Klurfeld DM. Consumption of cereal fiber, mixtures of whole grains and bran, and whole grains and risk reduction in type 2 diabetes, obesity, and cardiovascular disease. The American journal of clinical nutrition. 2013;98(2):594-619.

17. Ye EQ, Chacko SA, Chou EL, Kugizaki M, Liu S. Greater whole-grain intake is associated with lower risk of type 2 diabetes, cardiovascular disease, and weight gain.(Nutritional Epidemiology)(Author abstract)(Report). The Journal of Nutrition. 2012;142(7):1304.

18. Slavin J, Jacobs D, Marquart L. Whole- grain consumption and chronic disease: Protective mechanisms. Nutr Cancer1997. p. 14-21.

19. Jacobs D, Marquart L, Slavin J, Kushi LH. Whole- grain intake and cancer: An expanded review and meta-analysis. Nutr Cancer1998. p. 85-96.

20. Aune D, Chan DS, Lau R, Vieira R, Greenwood DC, Kampman E, et al. Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. BMJ. 2011;343:d6617.

21. FAO. FAO plant production and protection series. Rome: Rome : FAO; 2002.

22. Brouns FJPH, van Buul VJ, Shewry PR. Does wheat make us fat and sick? Journal of Cereal Science. 2013;58(2):209-15.

23. Shi Z, Taylor AW, Hu G, Gill T, Wittert GA. Rice intake, weight change and risk of the metabolic syndrome development among Chinese adults: the Jiangsu Nutrition Study (JIN). Asia Pacific journal of clinical nutrition. 2012;21(1):35.

24. Zhang JG, Wang ZH, Wang HJ, Du WW, Su C, Zhang J, et al. Dietary patterns and their associations with general obesity and abdominal obesity among young Chinese women. European journal of clinical nutrition. 2015.

25. Davis WR. Wheat Belly: Lose the Wheat, Lose the Weight, and Find Your Path Back to Health. Rodale Books. 2011.

26. Jönsson T, Olsson S, Ahrén B, Bøg-Hansen T, Dole A, Lindeberg S. Agrarian diet and diseases of affluence–Do evolutionary novel dietary lectins cause leptin resistance? BMC Endocrine Disorders. 2005;5(1):10.

27. Cheng J, Brar PS, Lee AR, Green PHR. Body mass index in celiac disease: beneficial effect of a gluten-free diet. Journal of clinical gastroenterology. 2010;44(4):267-71.

28. Soares FL, de Oliveira Matoso R, Teixeira LG, Menezes Z, Pereira SS, Alves AC, et al. Gluten-free diet reduces adiposity, inflammation and insulin resistance associated with the induction of PPAR-alpha and PPAR-gamma expression. The Journal of nutritional biochemistry. 2013;24(6):1105-11.

29. Kabbani TA, Goldberg A, Kelly CP, Pallav K, Tariq S, Peer A, et al. Body mass index and the risk of obesity in coeliac disease treated with the gluten-free diet. Aliment Pharmacol Ther. 2012;35(6):723-9.

30. Hyman M. The Blood Sugar Solution: The Bestselling Programme for Preventing Diabetes, Losing Weight and Feeling Great: Hodder & Stoughton General Division; 2012.

31. Zhanga J, Wanga H, Wanga Y, Xue H, Wang Z, Dua W, et al. Dietary patterns and their associations with childhood obesity in China. British Journal of Nutrition. 2015:1-7.

32. de Punder K, Pruimboom L. The dietary intake of wheat and other cereal grains and their role in inflammation. Nutrients. 2013;5(3):771-87.

33. FAO. Food Balance Sheets. A Handbook. Rome: Food and Agriculture Organization; 2001.

34. Siervo M, Montagnese C, Mathers JC, Soroka KR, Stephan BCM, Wells JCK. Sugar consumption and global prevalence of obesity and hypertension: an ecological analysis. 2014;17(3):587-96.

35. Basu S, Stuckler D, McKee M, Galea G. Nutritional determinants of worldwide diabetes: an econometric study of food markets and diabetes prevalence in 173 countries. Public health nutrition. 2013;16(1):1-8.

36. Maskova E, Paulickova I, Rysova J, Gabrovska D. Evidence for Wheat, Rye, and Barley Presence in Gluten Free Foods by PCR Method - Comparison with Elisa Method. Czech Journal Of Food Sciences. 2011;29(1):45-50.

37. Davis B, Wansink B. Fifty years of fat: news coverage of trends that predate obesity prevalence. BMC public health. 2015;15:629:1-6.

38. den Engelsen C, Gorter KJ, Salome PL, Rutten GE. Development of metabolic syndrome components in adults with a healthy obese phenotype: a 3-year follow-up. Obesity. 2013;21(5):1025-30.

39. Trøseid M, Seljeflot I, Weiss TW, Klemsdal TO, Hjerkinn EM, Arnesen H. Arterial stiffness is independently associated with interleukin-18 and components of the metabolic syndrome. Atherosclerosis. 2010;209(2):337-9.

40. Nelson JH. Wheat: its processing and utilization. The American journal of clinical nutrition. 1985;41(5 Suppl):1070-76.

41. Feldman N, Norenberg C, Voet H, Manor E, Berner Y, Madar Z. Enrichment of an Israeli ethnic food with fibres and their effects on the glycaemic and insulinaemic responses in subjects with non-insulin-dependent diabetes mellitus. The British journal of nutrition. 1995;74(5):681-8.

42. Latham MC. Human nutrition in the developing world (Food and Nutrition Series No. 29). Rome: Food and Agriculture Organization of the United Nations; 1997.

43. Shewry PR. Wheat. Journal of experimental botany. 2009;60(6):1537.

44. Song X, Ni Z, Yao Y, Zhang Y, Sun Q. Identification of differentially expressed proteins between hybrid and parents in wheat (Triticum aestivum L.) seedling leaves. TAG Theoretical and applied genetics Theoretische und angewandte Genetik. 2009;118(2):213-25.

45. Gao X, Liu SW, Sun Q, Xia GM. High frequency of HMW-GS sequence variation through somatic hybridization between Agropyron elongatum and common wheat. Planta. 2010;231(2):245-50.

46. Henneberg M, Grantham J. Obesity - a natural consequence of human evolution. Anthropological Review. 2014;77(1):1-10.

47. Garg S, Pandey D, Taj G, Goel A, Kumar A. TRIPATH: A Biological Genetic and Genomic Database of Three Economically Important Fungal Pathogen of Wheat - Rust: Smut: Bunt. Bioinformation. 2014;10(7):466-8.

48. Pomeranz Y. Wheat : chemistry and technology. 3rd ed. ed. St. Paul, Minn., USA: St. Paul, Minn., USA: American Association of Cereal Chemists; 1988.

49. Cecinato P, Fuccio L, Sabattini E, Laterza L, Caponi A, Azzaroli F, et al. An unusual cause of weight loss in a young Caucasian man. Common variable immunodeficiency (CVI) associated with diffuse enteral nodular lymphoid hyperplasia (NLH) and CD. Gut. 2014;63(5):856, 9.

50. Yang Y. Chinese Food Composition Table 2004: Beijing Medical University Press; 2005.

51. Prentice AM, Jebb SA. Fast foods, energy density and obesity: a possible mechanistic link. Oxford, UK2003. p. 187-94.

52. Savage JS, Marini M, Birch LL. Dietary energy density predicts women's weight change over 6 y. The American journal of clinical nutrition. 2008;88(3):677.

53. Sun Q, Spiegelman D, van Dam RM, Holmes MD, Malik VS, Willett WC, et al. White rice, brown rice, and risk of type 2 diabetes in US men and women. Archives of Internal Medicine. 2010;170(11):961.

54. James WPT, Chunming C, Inoue S. Appropriate Asian body mass indices? (Editorial). Obesity Reviews 2002;3(3):139-.

55. WHO/IASO/IOTF. The Asia-Pacific perspective: redefining obesity and its treatment: Health Communications Australia Pty: Melbourne.; 2000.

56. Cordain L, Eaton SB, Sebastian A, Mann N, Lindeberg S, Watkins BA, et al. Origins and evolution of the Western diet: health implications for the 21st century. American Journal Of Clinical Nutrition. 2005;81(2):341-54.

57. Grantham JP, Staub K, Rühli FJ, Henneberg M. Modern diet and metabolic variance--a recipe for disaster? Nutrition journal. 2014;13:15:01-10.

58. Speth JD. Early hominid hunting and scavenging: the role of meat as an energy source. Journal of Human Evolution. 1989;18(4):329-43.

59. FAO and WHO. Carbohydrates in human nutrition-report of a joint FAO/WHO expert consultation. Rome: World Health Organization and Food and Agriculture Organization; 1998.

60. FAO. FAOSTAT [Available from: http://faostat3.fao.org.

61. Okubo H, Sasaki S, Murakami K, Kim MK, Takahashi Y, Hosoi Y, et al. Three major dietary patterns are all independently related to the risk of obesity among 3760 Japanese women aged 18–20 years. International journal of obesity. 2007;32(3):541-9.

62. Lin H, Bermudez OI, Tucker KL. Dietary Patterns of Hispanic Elders Are Associated with Acculturation and Obesity. Dietary Patterns of Hispanic Elders Are Associated with Acculturation and Obesity. 2003;133(11):3651-7.

63. Kim J-H, Lee JE, Jung I-K. Dietary Pattern Classifications and the Association with General Obesity and Abdominal Obesity in Korean Women. Academy of Nutrition and Dietetics. 2012;112(10):1550-9.

64. Sichieri R. Dietary Patterns and Their Associations with Obesity in the Brazilian City of Rio de Janeiro. Obesity Research. 2002;10(1):42-8.

# Chapter 3: Birth behaviour & global gynecological cancers

Article 8/10: Greater family size associated with less cancer risk: an ecological analysis of 178 countries

Wenpeng You[1]*, Frank J. Rühli[2], Renata J. Henneberg[1], Maciej Henneberg[1,2]

[1] Adelaide Medical School, The University of Adelaide, Adelaide, Australia

[2] Institute of Evolutionary Medicine, University of Zürich, Zürich, Switzerland

**Under review:** Wenpeng You, Frank J. Rühli, Renata Henneberg, Maciej Henneberg. Greater family size associated with less cancer risk: an ecological analysis of 178 countries. Under review by BMC Cancer after three revisions

✉ **Correspondence:** Wenpeng You wenpeng.you@adelaide.edu.au

## Contexture Statement

A large number of studies concluded that less childbearing may be a major risk of gynaecological cancers. A limited number of studies associated the number of children born into a family (family size) with the risk of non- gynaecological cancers of other family members. Studies also revealed that greater family size may create more positive psychological well-being for family members, which may offer the protection against cancer development.

Ecological study has the advantages that more objective variables, which can even be backdated for predicting prolonged disease presentation, can be included for analysis. With these advantages, we collected the country specific incidence rates of all cancers and individual site cancers and analysed their relationships with family size (indexed by total fertility rate (TFR)). We found that TFR inversely correlated to the incidence rates of all cancers of both sexes, male and female and most of the individual site cancers. This may suggest that greater total fertility rate may protect all the family members from developing cancers.

**Statement of Authorship**

# Statement of Authorship

| | |
|---|---|
| Title of Paper | Greater family size is associated with less cancer risk: an ecological analysis of 178 countries |
| Publication Status | ☐ Published ☐ Accepted for Publication<br>☑ Submitted for Publication ☐ Publication Style |
| Publication Details | Under review by BMC Cancer, Manuscript ID BCAN-D-16-02324R2<br>We have completed two revisions as required by the editorial office. |

## Principal Author

| | | | |
|---|---|---|---|
| Name of Principal Author (Candidate) | Wenpeng You | | |
| Contribution to the Paper | Conceived the hypothesis, collected data, performed analysis on all samples, interpreted data, wrote manuscript and acted as corresponding author. | | |
| Overall percentage (%) | 50 | | |
| Signature | | Date | 22/12/2017 |

## Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

| | | | |
|---|---|---|---|
| Name of Co-Author | Frank J. Rühli | | |
| Contribution to the Paper | Discussed and formalized the hypothesis with all other authors, performed and interpreted the data analysis | | |
| Signature | | Date | 19/01/2018 |

| | | | |
|---|---|---|---|
| Name of Co-Author | Renata Henneberg | | |
| Contribution to the Paper | Discussed and formalized the hypothesis with all other authors, performed and interpreted the data analysis, evaluated and edited manuscript | | |
| Signature | | Date | 19/01/2018 |

| | | | |
|---|---|---|---|
| Name of Co-Author | Maciej Henneberg | | |
| Contribution to the Paper | Supervised development of work, formalized and interpreted data analysis, and evaluated and edited manuscript | | |
| Signature | | Date | 22/12/2017 |

## Abstract

Background: Greater family size measured with total fertility rate (TFR) and with household size, may offer more life satisfaction to the family members. Positive psychological well-being has been postulated to decrease cancer initiation risk. This ecological study aims to examine the worldwide correlation between family size, used as the measure of positive psychological well-being, and total cancer incidence rates.

Methods: Country specific estimates obtained from United Nations agencies on total cancer incidence rates (total, female and male rates in age range 0-49 years and all ages respectively), all ages site cancer incidence (bladder, breast, cervix uteri, colorectum, corpus uteri, lung, ovary and stomach), TFR, household size, life expectancy, urbanization, per capita GDP PPP and self-calculated Biological State Index ($I_{bs}$) were matched for data analysis. Pearson's, non-parametric Spearman's, partial correlations, independent T-test and multivariate regressions were conducted in SPSS.

Results: Worldwide, TFR and household size were significantly and negatively correlated to all the cancer incidence variables. These correlations remained significant in partial correlation analysis when GDP, life expectancy, $I_{bs}$ and urbanization were controlled for. TFR correlated to male cancer incidence rate (all ages) significantly stronger than it did to female cancer incidence rate (all ages) in both Pearson's and partial correlations. Multivariate stepwise regression analysis indicated that TFR and household size were consistently significant predictors of all cancer incidence variables.

Conclusions: Countries with greater family size have lower cancer risk in both females, especially males. It may be worthwhile to include more family life satisfaction as part of strategic plan of cancer prevention.

**Trial registration:** Not applicable

**Keywords:** Total fertility rate, Household size, Psychological well-being, Family life, Cancer initiation

## Introduction

Total fertility rate (TFR) representing the total number of births during a lifetime of a female [1, 2] has been used to measure childbearing and family size [3-5] in a number of studies. The prevalent conclusions were that more childbearing (greater TFR) may protect against female breast cancer [6], corpus uteri cancer [7] and ovarian cancer [8] due to less oestrogen production or less menstrual cycles [9] and more oxytocin secretion [10, 11], but may contribute to cervix uteri cancer because of more exposure to infection risk [12].

The number of children born into a family does not only influence the mother's physiological health of her reproductive system, but also has effects on health, including cancer development, and on her other systems and on all the other family members. For instance, greater family size has been postulated to protect family members from developing colorectal cancer [5], melanoma of skin [5], bladder cancer [5], breast cancer [5] and stomach cancer (in males only) [13]. Relationships between greater family size /household size and lung cancer [5, 14] and stomach cancer (females) [13] were explored, but without much success of seeing a clear trend. Aldrich *et al.* [14] reported that household size was in significant association with a risk of developing lung cancer in African Americans, but not in Latinos.

These controversial and circumstantial correlations between reproductive behaviour and a comprehensive health effect on all family members directed our attention to seeking alternative explanation of the relationship between TFR and risk factors for cancer. Psychological factors have been suggested to be linked with cancer initiation, but the mechanism has been intriguing professionals and laypeople for decades [15, 16]. Although studies on the possible effects of positive and negative psychological factors arising from life events on cancer incidence and prognosis are numerous, the literature remains contradictory as to methods and impacts [14, 17-19]. Extensive studies have suggested that adverse life events and the associated psychological stress may predispose to cancers in various body sites [20-24]. Everson *et al.* [25] reported that more stress may increase the cancer risk. This might be because people tend to recall adverse life events, but easily forget those positive ones, which constantly happen in the daily life. Cancer patients may more easily recall those negative life events which have been considered as cancer risks [26-30]. Only a limited number of studies have addressed the relationship between life satisfaction and cancer risk [14, 31], but the conclusions were controversial.

Research conducted into health effects of positive psychological well-being has concluded that family life satisfaction may stimulate oxytocin production in the human body [32-38], which may have the inhibitory effect on specific cancers [10, 11, 39, 40]. For example, positive psychological wellbeing has been postulated to protect against cancer risk in Israeli women [31], reduce the number of American cancer patients from going into metastasis [41] and help cancer patients with cancer's detection, treatment, and survival [41]. Large families have greater life satisfaction in both Western and Eastern populations [42, 43]. Nan *et al.* [43] have also concluded that the bigger family size is, the higher Subjective Happiness Scale (SHS) result is in the family, regardless of cultural backgrounds.

Therefore, in this study, we assessed, from a global perspective, whether greater family size, measured with TFR [3] and household size may lower cancer risk using empirical, macro-level data obtained from international organizations.

## Materials and Methods

Data Sources

The population specific data were collected for this ecological study.

1. The GLOBOCAN 2012 estimates of incidence rates (age standardised, world) of all cancers excluding non-melanoma skin cancer (C00-97, but C44) in total, and separately for males and females of all ages [44] were used. Crude estimates of incidence rates of all cancers excluding non-melanoma skin cancer (C00-97, but C44) in total, and for males and females in age group 0-49 years were also obtained.

The incidence rates of the individual site-specific cancers (bladder, breast, cervix uteri, colorectum, corpus uteri, lung, melanoma and ovary) were extracted as the dependent variables for data analysis in this study. The results from this study were aligned with the findings of previous studies of the relationships between family/household size and each of these site-specific cancers namely lung cancer [14], bladder cancer, melanoma and colon cancer [5].

GLOBOCAN is a project conducted by the International Agency for Research on Cancer (IARC) of the World Health Organization. This project provides contemporary population level estimates by cancer site and sex using the best available data in each population and uses nine comprehensive methods of estimation [45].

2. The United Nations Statistics Division estimates of the life expectancy [46], the total population in households and the number of households [47].

Life expectancy is the average number of years a person of a given age, residing in a given country is expected to live. We extracted the life expectancy at age of 60 years old ($e_{60}$, 2005-2010) from abridged life tables (1950-2100) [46] published online. Ageing has been a significant risk predictor of cancer. In this study, life expectancy ($e_{60}$) was considered as the indicator of ageing.

As instructed by the United Nations Statistics Division, we created a new variable, household size, through dividing the total population in households [47] by the number of households [47] in each country.

3. The World Bank published data [1] on Gross Domestic Product (GDP), total fertility rate (TFR) and urbanization.

GDP was expressed in per capita purchasing power parity (PPP in current international $) in 2010. The World Bank also clusters countries into 4 classifications in terms of their GDP per capita (High Income, Upper Middle Income, Low Middle Income and Low Income). In this study, we grouped countries with High Income and Upper Middle Income as developed countries, and countries with Low Middle Income and Low Income as developing countries.

Urbanization was expressed with the percentage of total population living in urban areas in 2010.

Total Fertility Rate (TFR) represents total births per woman during her lifetime. It indicates the number of children that would be born to a woman if she were to live to the end of her childbearing years and bear children in accordance with age-specific fertility rates of the specified year.

Total births per woman have been used to indicate the family size in studies at an individual level [3, 4]. Therefore, we used TFR as the measure of family size in this study, and terms “TFR” and “family size” were interchangeably used thereafter. Household size is used as the proxy of family size in this study that has been calculated from data independent from those used for TFR.

4. Biological State Index ($I_{bs}$) was self-calculated [48, 49] with the fertility data of each country published by United Nations in 2008 [50] and the mortality data of life tables (2009) published by World Health Organization (WHO) in 2012 [51].

$I_{bs}$ was included as one of the confounding factors that indicates the level of adaptation of a population [52]. The $I_{bs}$ values range between 0 and 1.0. A greater $I_{bs}$ value of a population means less opportunity for natural selection, and vice versa. The simplest interpretation of $I_{bs}$ is that it indicates a probability with which an average person born into a population is able to pass her/his genes to the next generation. Recent studies have postulated that $I_{bs}$ may indicate the magnitude of deleterious gene/mutation accumulation in a population due to relaxed natural selection [52-55]. The greater $I_{bs}$ value means that a population has accumulated more deleterious gene/mutations of cancer [53], obesity [54] and type 1 diabetes [55], and vice versa [52-55]. Inclusion of $I_{bs}$ as a confounder may remove the influence of cancer gene/mutation accumulation on the correlation between family size and cancer incidence.

Data Selection

We used country specific cancer incidence rates, TFR, GDP, urbanization, household size and life expectancies for all countries where the most updated and recent data were available (N=178). In order to capture as many countries as we could for this study, we

aligned country specific TFR with all cancer incidence rates, and then we matched other country-specific variables with the TFR.

Each country was treated as an individual subject in the analysis. Numbers of countries included in analyses of relationships with other variables may differ somewhat because all information was not uniformly available for all countries. The list of countries included in this study can be found in Supplementary File 1, Table S1.

We singled out the population segment aged 0-49 years because females enter menopause at around 50 years of age and since then they produce less and less female hormones. Numerous studies have associated female oestrogen level with cancer risk [9, 10].

All the aforementioned data were freely available from the websites of the UN agencies. No ethical approval or written informed consent for participation was required.

Data analysis

Scatter plots were produced in Excel (Microsoft® 2016) to explore and visualize the correlations between family size and cancer incidence in total population, males and females respectively. Scatter plots allowed us to assess data quality and distributions of variables. In the supplemental material, family size was replaced with household size for performing the scatter plots (Supplemental File 2, Figure S1).

Prior to correlation/regression analyses all data were log-transformed (ln) in order to reduce non-homoscedascity of their distributions and possible curvilinearity of regressions. To assess the relationships between each cancer incidence rate and family size, the analysis proceeded in four steps.

1. Pearson's and nonparametric correlations (Spearman's rho) were used to evaluate the strength and direction of the associations between family size and all other variables, including independent variables and confounders.

2. Partial correlation of Pearson's moment-product approach was used to assess the relationship between each cancer incidence rate and family size respectively while we controlled for GDP PPP [44, 45], urbanization [56, 57], $I_{bs}$ and life expectancy [58] which have been commonly considered as the contributing factors of cancer.

Fisher's r-to-z transformation was performed to test significance of differences between correlation coefficients.

3. Standard multiple linear regression (stepwise) was performed to identify the most significant predictor(s) of cancer risk. The dependent variables included cancer incidence rate by sex (total, male and female, age group 0-49 and all ages respectively).

The independent variables/predictors entered into analyses were family size, urbanization, GDP, $I_{bs}$ and life expectancy (not for cancer variables for 0-49 years).

4. The independent samples t-test was performed to compare the means of each cancer variable in high and low fertility countries divided at the cut point of TFR=2.36. We used 2.36 as the cut point because it is the world average TFR published by the United Nations for the period of 2010-2015 [59].

Socioeconomic level plays a critical role in family happiness. In parallel to the analyses of the relationship between family size and cancer variables worldwide, the relationships between family size and each cancer variable in developed and developing country groupings were also examined respectively. Descriptive statistics including standard deviations of all variables were calculated for analysing and comparing the covariance (relationship between family size and cancer incidence) in all countries (n=178), in developed world (n=98) and in developing world (n=80).

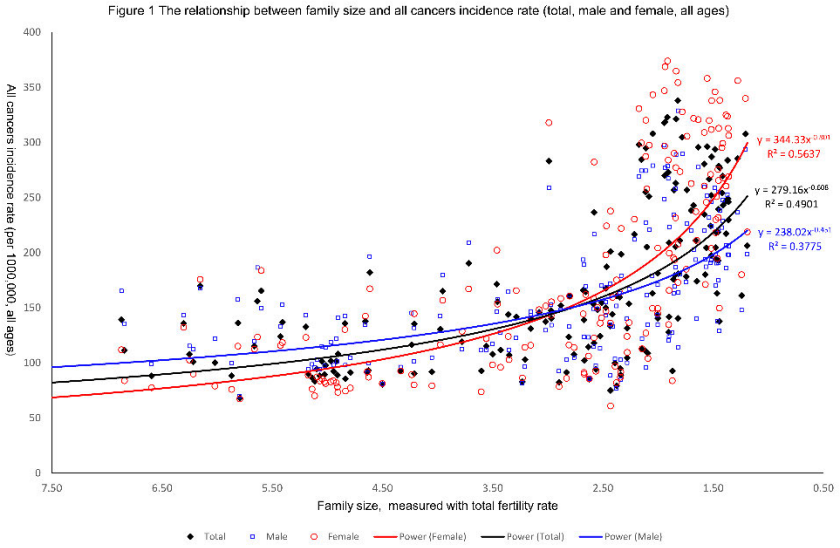
Subsequently, family size was substituted with household size for reanalysing the associations and regressions. The results were reported in Supplementary Files 3 (Table S2) and 4 (Tables S3). There was no stratification of country grouping in the supplemental analyses due to limited sample size of countries for which household size was available (n=58).

Pearson's, non-parametric Spearman's rho correlations, partial correlation, stepwise multiple linear regression, independent samples *t*-test analyses and descriptive statistics were calculated using SPSS v. 22 (SPSS Inc., Chicago Il USA). To increase homoscedasticity of data distributions, log-transformed variables were used for correlation analyses. The significance was reported when P-value was <0.05, but the significance levels of $p < 0.01$ and $p<0.001$ were also indicated in the tables. Regression analysis criteria were set at probability of F to enter ≤ 0.05 and probability of F to remove ≥ 0.10. The raw data were used for scatter plots.

## Results

Figure 1 shows a negative and strong correlation of family size to cancer incidence rates in total population and in males and females separately (all ages). The non-linear relationships between family size and group cancer incidence variables identified in the scatterplots show the strong correlation between family size and cancer incidence rate ($R^2$=0.4901, 0.5637 and 0.3755). Household size as the proxy of family size has shown the similar correlation to all cancers incidence rates (total, female and male) (Supplementary File 2: Figure 1).

Figure 1: The relationship between family size and all cancers incidence rates (total, male and female, all ages)




The subsequent analyses of log-transformed data proved these relationships. Globally (n=178), Spearman's rank correlation showed that family size was in significant negative correlation to all cancers incidence rates (both sexes) in all ages ($r = -0.716$, $p < 0.001$) and in age group 0-49 years ($r = -0.752$, $p < 0.001$), separately in females of all ages ($r = -0.640$, $p < 0.001$) and age group 0-49 years ($r = -0.762$, $p < 0.001$) and in males of all ages ($r = -0.761$, $p < 0.001$) and age group 0-49 years ($r = -0.765$, $p < 0.001$)(Table 1). Pearson's *r* showed quite similar relationship trends between family size and the cancer variables (Table 1).

When family size was replaced with household size for supplemental data analysis, household size also showed significant, negative and strong correlation to each cancer variable (both sexes, female and male in age groups, 0-49 and all ages respectively) (Supplementary File 3: Table S1).

In developed countries grouping (n=98), Spearman's rank correlation showed that family size was in significant negative correlation to all cancers incidence rates (both sexes) in all ages ($r = -0.540$, $p < 0.001$) and age group 0-49 years ($r = -0.705$, $p < 0.001$),

separately in females of all ages ($r$ = -0.477, $p$ < 0.001) and age group 0-49 years ($r$ = -0.581, $p$ < 0.001) (Table 1) and in males of all ages ($r$ =- 0.582, $p$ < 0.001) and age group 0-49 years ($r$ = -0.705, $p$ < 0.001). Pearson's $r$ showed quite similar relationship trends between family size and the cancer variables (Table 1).

In developing countries grouping (n=80), Spearman's rank correlation showed that family size was in significant negative correlation to all cancers incidence rates in all ages ($r$ = -0.334, $p$ < 0.001) and age group 0-49 years ($r$ = -0.498, $p$ < 0.001), separately in females of age group 0-49 years ($r$ = -0.482, $p$ < 0.001) but not at all ages ($r$ = -0.140) and in males of all ages ($r$ =- 0.457, $p$ < 0.001) and age group 0-49 years ($r$ = -0.449, $p$ < 0.001) (Table 1). Pearson's $r$ showed quite similar relationship trends between family size and the cancer variables, except for all cancers incidence rate in females at all ages (r=-0.200, r<0.05) (Table 1).

When, in the partial correlation analyses, we controlled for the major confounders (GDP, urbanization, life expectancy (not for cancer variable in age group 0-49 years) and $I_{bs}$): 1) globally (n=178), family size remained in the significant correlation to all cancer incidence rates (both sexes) in all ages ($r$ = -0.362, $p$ < 0.001) and the age group 0-49 years ($r$ = -0.534, $p$ < 0.001), in females of all ages (r = -0.230p < 0.001) and age group 0-49 years (r = -0.492, p < 0.001) and in males of all ages (r = -0.449, p < 0.001) and age group 0-49 years (r = -0.542, p < 0.001) (Table 1). Family size correlated stronger with male cancers incidence than with female cancers in all ages group (n=178). This difference was shown to be statistically significant by Fisher's r-to-z transformation in both Pearson's (z= 2.43, p=0.015) and partial (z= 2.22, p=0.026) correlations. 2) In developed world (n=98), family size also remained in the significant correlation to all cancers incidence rates (both sexes) in all ages ($r$ = -0.625, $p$ < 0.001) and the age group 0-49 years ($r$ = -0.658, $p$ < 0.001), in males of all ages (r = -0.470, p < 0.001) and age group 0-49 years (r = -0.581, p < 0.001) and in females of all ages (r = -0.362, p < 0.001) and the age group 0-49 years (r = -0.534, p < 0.001) (Table 1). 3) In developing world (n=80), family size remained in the significant correlation to all cancer incidence rates

(both sexes) in the age group 0-49 years ($r = -0.430$, $p < 0.001$) but not at all ages group, in females of age group 0-49 years ($r = -0.384$, $p < 0.001$) but not at all ages and in males of all ages ($r = -0.303$, $p < 0.05$) and the age group 0-49 years ($r = -0.430$, $p < 0.001$) (Table 1).

Table 1 also shows that, globally (n=178) and in developed world (n=98), each of the incidence rates (all ages) of individual site cancers in bladder, breast, colorectum, corpus uteri, lung, skin (melanoma), ovary and stomach was in significant, negative and strong correlation to family size in both Pearson's and partial correlation analyses (Table 1). Globally (n=178), cervix uteri cancer correlated with family size significantly and positively in both Pearson's r and non-parametric correlation, but the correlation was neither strong nor significant in partial correlation (Table 1). In developed world (n=80), cervix uteri cancer did not show correlation (partial) or very weak correlation (Pearson's r) with family size (Table 1) although it statistically significantly correlated with family size ($r=0.223$, $p<0.05$). In developing world (n=80), only correlations between family size and lung cancer and cervix uteri cancer were consistent with those revealed globally and in developed world (Table 1).

Table 1: Pearson's, nonparametric and partial correlations between family size and each cancer variable

| | Total Fertility Rate | Household size | Cancer incidence, 0-49, Total | Cancer incidence, 0-49 years, Male | Cancer incidence, 0-49 years, Female | Cancer incidence, All ages, Total | Cancer incidence, All ages, Male | Cancer incidence, All ages, Female | Life expectancy at 60 years | GDP PPP | Urbanization |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Pearson's & nonparametric correlations** | | | | | | | | | | | |
| Total Fertility Rate | 1 | 0.491*** | -0.746*** | -0.717*** | -0.745*** | -0.695*** | -0.747*** | -0.607*** | -0.704*** | -0.764*** | -0.574*** |
| Household size | 0.437*** | 1 | -0.668*** | -0.646*** | -0.634*** | -0.654*** | -0.615*** | -0.641*** | -0.477*** | -0.427*** | -0.381** |
| Cancer incidence,0-49 years, Total | -0.775*** | -0.742*** | 1 | 0.955*** | 0.978*** | 0.928*** | 0.905*** | 0.900*** | 0.629*** | 0.586*** | 0.397*** |
| Cancer incidence,0-49 years, Male | -0.759*** | -0.698*** | 0.952*** | 1 | 0.878*** | 0.891*** | 0.896*** | 0.827*** | 0.584*** | 0.547*** | 0.354*** |
| Cancer incidence,0-49 years, Female | -0.758*** | -0.731*** | 0.981*** | 0.885*** | 1 | 0.909*** | 0.870*** | 0.908*** | 0.641*** | 0.619*** | 0.425*** |
| Cancer incidence, All ages, Total | -0.713*** | -0.741*** | 0.931*** | 0.896*** | 0.911*** | 1 | 0.980*** | 0.968*** | 0.657*** | 0.605*** | 0.433*** |
| Cancer incidence, All ages, Male | -0.758*** | -0.691*** | 0.908*** | 0.903*** | 0.870*** | 0.978*** | 1 | 0.902*** | 0.658*** | 0.625*** | 0.436*** |
| Cancer incidence, All ages, Female | -0.635*** | -0.739*** | 0.903*** | 0.832*** | 0.909*** | 0.967*** | 0.902*** | 1 | 0.618*** | 0.575*** | 0.419*** |
| Life expectancy at 60 years old | -0.690*** | -0.522*** | 0.626*** | 0.579*** | 0.640*** | 0.653*** | 0.660*** | 0.611*** | 1 | 0.720*** | 0.585*** |
| GDPPPP | -0.729*** | -0.595*** | 0.608*** | 0.571*** | 0.629*** | 0.614*** | 0.635*** | 0.591*** | 0.731*** | 1 | 0.730*** |
| Urbanization | -0.559*** | -0.283* | 0.456*** | 0.398*** | 0.484*** | 0.487*** | 0.479*** | 0.481*** | 0.662*** | 0.793*** | 1 |
| **Partial correlations, keeping DGP, urbanisation and life expectancy constant** | | | | | | | | | | | |
| Total Fertility Rate | 1 | 0.214 | -0.511*** | -0.500*** | -0.473*** | -0.367*** | -0.466*** | -0.222** | - | - | - |
| Household size | | 1 | -0.526*** | -0.514*** | -0.459*** | -0.509*** | -0.448*** | -0.491*** | - | - | - |
| Cancer incidence,0-49, Total | | | 1 | 0.927*** | 0.961*** | 0.871*** | 0.827*** | 0.827*** | - | - | - |
| Cancer incidence,0-49, Male | | | | 1 | 0.795*** | 0.818*** | 0.827*** | 0.715*** | - | - | - |
| Cancer incidence,0-49, Female | | | | | 1 | 0.831*** | 0.753*** | 0.839*** | - | - | - |
| Cancer incidence, All ages, Total | | | | | | 1 | 0.962*** | 0.946*** | - | - | - |
| Cancer incidence, All ages, Male | | | | | | | 1 | 0.825*** | - | - | - |
| Cancer incidence, All ages, Female | | | | | | | | 1 | - | - | - |

Pearson's, nonparametric and partial correlation reported. *** p<0.001, ** p<0.01, * p<0.05; n range: 52 or 58 for the correlation of household size to other variables, n range: 163 or 176 for the correlation of TFR (total fertility rate) to other variables.

All variables were log-transformed. In partial correlation analysis, life expectancy (e60), GDP and Urbanization were controlled for.

Data sources and variable meanings:

The data from the International Agency for Research on Cancer: cancer incidence rate (per 100,000 in 2012) by sex (total, male and female, age group (0-49 and all ages)

the World Bank data: Total Fertility Rate (the mean number of children born to a woman between 2009-2011), GD PPPP (per capita purchasing power parity in current international $ in 2010) and Urbanization (the percentage of total population living urban areas in 2010)

The United Nations data: Household size (the number of persons living in a household in 2010) and life expectancy ($e_{60}$, 2005-2010)

The correlations, especially the partial correlations between family size and cancer variables in all countries (n=178) and developed world (n=98) were stronger and more significant than those in developing world. Variances of cancer incidence variables in developing world (n=80) were smaller than their counterparts in the developed world and all countries grouping (Supplementary File 5).

Table 2 shows that, globally (n=178), the mean incidence rate of each cancer variable in country group (n=95) with TFR ≥ 2.36 was significantly ($p<0.001$) lower than that of country group (n=83) with TFR < 2.36 except cervix uteri cancer. This trend remained in the developed country grouping (n=98) except cervix uteri and stomach cancers, and in developing grouping (n=80) except cancers in breast, cervix uteri, melanoma (skin) and ovary.

Table 2 T-Test to compare the difference between means in two country groups with the cut point of 2.36 (TFR)

| | All countries, n=178 | | | | Developed Countries, n=98 | | | | Developing Countries, n=80 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Sample Size | Mean | Mean Difference | t | Sample Size | Mean | Mean Difference | t | Sample Size | Mean | Mean Difference | t |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) - all ages: both sexes | 95 | 123.81 | -96.58 | -12.50*** | 26 | 141.57 | -88.17 | -6.67*** | 69 | 117.11 | -42.01 | -3.89*** |
| | 83 | 220.39 | | | 72 | | | | 11 | | | |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44)- all ages: female | 95 | 129.73 | -72.58 | -10.67*** | 26 | 143.80 | -66.91 | -5.88*** | 69 | 124.42 | -22.87 | -2.23* |
| | 83 | 202.31 | | | 72 | | | | 11 | | | |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) - all ages: male | 95 | 120.09 | -128.54 | -13.45*** | 26 | 144.41 | -114.97 | -7.00*** | 69 | 110.93 | -67.41 | -5.03*** |
| | 83 | 248.64 | | | 72 | | | | 11 | | | |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) – 0-49: both sexes‡ | 95 | 35.98 | -36.82 | -13.37*** | 26 | 43.04 | -32.18 | -6.49*** | 69 | 33.32 | -23.58 | -6.73*** |
| | 83 | 72.80 | | | 72 | | | | 11 | | | |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) – 0-49: female‡ | 95 | 47.01 | -45.90 | -12.87*** | 26 | 58.15 | -38.23 | -6.06*** | 69 | 42.82 | -27.40 | -6.06*** |
| | 83 | 92.91 | | | 72 | | | | 11 | | | |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) – 0-49: male‡ | 95 | 25.24 | -28.28 | -12.68*** | 26 | 28.57 | -26.50 | -6.52*** | 69 | 23.99 | -19.38 | -6.41*** |
| | 83 | 53.52 | | | 72 | | | | 11 | | | |
| Bladder (C67), all ages | 95 | 2.91 | -4.76 | -9.48*** | 26 | 4.03 | -4.12 | -4.75*** | 69 | 2.49 | -2.03 | -2.70** |
| | 83 | 7.67 | | | 72 | | | | 11 | | | |
| Breast(C50), all ages | 95 | 31.01 | -29.29 | -10.43*** | 26 | 38.65 | -25.47 | -5.46*** | 69 | 28.13 | -7.15 | -1.80 |
| | 83 | 60.30 | | | 72 | | | | 11 | | | |
| Cervix uteri (C53), all ages | 95 | 26.18 | 13.24 | 7.17*** | 26 | 20.85 | 8.29 | 4.02*** | 69 | 28.19 | 12.77 | 2.61* |
| | 83 | 12.94 | | | 72 | | | | 11 | | | |
| Colorectum (C18-21), all ages | 95 | 7.63 | -16.04 | -13.12*** | 26 | 11.34 | -13.91 | -6.68*** | 69 | 6.24 | -7.11 | -4.51*** |
| | 83 | 23.67 | | | 72 | | | | 11 | | | |
| Corpus uteri (C54), all ages | 95 | 5.15 | -6.91 | -8.91*** | 26 | 7.23 | -5.21 | -4.06*** | 69 | 4.37 | -5.24 | -3.76*** |
| | 83 | 12.06 | | | 72 | | | | 11 | | | |
| | 95 | 7.43 | -17.12 | -12.62*** | 26 | 11.42 | -13.76 | -6.16*** | 69 | 5.93 | -14.55 | -5.94*** |

| | All countries, n=178 | | | | Developed Countries, n=98 | | | | Developing Countries, n=80 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Sample Size | Mean | Mean Difference | t | Sample Size | Mean | Mean Difference | t | Sample Size | Mean | Mean Difference | t |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) - all ages: both sexes | 95 | 123.81 | -96.58 | -12.50*** | 26 | 141.57 | -88.17 | -6.67*** | 69 | 117.11 | -42.01 | -3.89*** |
| | 83 | 220.39 | | | 72 | | | | 11 | | | |
| Lung (C33-34), all ages | 83 | 24.55 | | | 72 | | | | 11 | | | |
| Melanoma of skin (C43), all ages | 95 | 1.16 | -4.80 | -6.35*** | 26 | 1.67 | -5.05 | -3.38*** | 69 | 0.97 | -0.07 | -0.25 |
| | 83 | 5.96 | | | 72 | | | | 11 | | | |
| Ovary (C56), all ages | 95 | 4.67 | -3.39 | -9.27*** | 26 | 5.09 | -3.31 | -5.56*** | 69 | 4.51 | -1.28 | -1.95 |
| | 83 | 8.06 | | | 72 | | | | 11 | | | |
| Stomach (C16), all ages | 95 | 6.29 | -3.32 | -3.71*** | 26 | 7.24 | -2.10 | -1.50 | 69 | 5.92 | -5.38 | -2.90** |
| | 83 | 9.61 | | | 72 | | | | 11 | | | |

Note:

Significance level: *** p<0.001, ** p<0.01, * p<0.05

Data sources and variable meanings:

The International Agency for Research published cancer incidence rates (per 100,000 in 2012) of all cancers incidence rate by sex (total, male and female, 0-49 years and all ages respectively); bladder, breast, cervix uteri, colorectum, corpus uteri, ovary and stomach.

The World Bank data: Total Fertility Rate (TFR, the mean number of children born to a woman between 2009-2011). The TFR 2.36 was used as the cut point to stratify the total countries (n=178).

For each cancer variable, the mean of countries with TRR ≥2.36 was reported in the first line, and the mean of countries with TFR < 2.36 was reported in the second line. Sample sizes: n=95

In the standard multiple linear (stepwise) regression analyses, family size was the significant predictor of the total, male and female cancer incidence rates (with exception of all female cancers at all ages) in samples of all ages and in age group 0-49 years respectively when family size, GDP, life expectancy (not for age group 0-49 years) and $I_{bs}$ were entered as the independent variables/predictors (Table 3). Although family size is a significant predictor of the variable of all cancers in females at all ages, the value of its beta coefficient was smaller than for the variable of all cancers in males at all ages. This finding was consistent with those reported in Table 1 that family size was in significantly stronger negative association with all cancers in males at all ages than it was with all cancers in females at all ages in both Pearson's ($z= 2.43$, $p=0.015$) and partial ($z= 2.22$, $p=0.026$) correlation analyses. This means that greater family size may have more protective effects on male cancer risk than on female cancer risk.

Table 3 Stepwise multiple regression results to identify the significant predictors of cancer incidence risk

| Cancer incidence, 0-49 years | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rank | Total | | | | Male | | | | Female | | | |
| | Predictor | Beta | Sig. | Adjusted $R^2$ | Predictor | Beta | Sig. | Adjusted $R^2$ | Predictor | Beta | Sig. | Adjusted $R^2$ |
| 1 | Family size | -0.623 | <0.001 | 0.626 | Family size | -0.619 | <0.001 | 0.608 | Family size | -0.550 | <0.001 | 0.625 |
| 2 | Household Size | -0.352 | <0.001 | 0.716 | Household Size | -0.338 | <0.001 | 0.690 | Household Size | -0.249 | 0.005 | 0.688 |
| 3 | | | | | | | | | Life expectancy | 0.207 | 0.029 | 0.709 |
| **Cancer incidence, All ages** | | | | | | | | | | | | |
| Rank | Total | | | | Male | | | | Female | | | |
| | Predictor | Beta | Sig. | Adjusted $R^2$ | Predictor | Beta | Sig. | Adjusted $R^2$ | Predictor | Beta | Sig. | Adjusted $R^2$ |
| 1 | Family size | -0.370 | <0.001 | 0.487 | Family size | -0.597 | <0.001 | 0.559 | Life expectancy | 0.360 | 0.002 | 0.438 |
| 2 | Household Size | -0.334 | <0.001 | 0.605 | Household Size | -0.319 | <0.001 | 0.631 | Household Size | -0.348 | <0.001 | 0.567 |
| 3 | Life expectancy | 0.289 | 0.006 | 0.652 | | | | | Family size | -0.247 | 0.029 | 0.597 |

Variables (log-transformed) entered for multiple linear regression (stepwise) analysis: Family size (measured with total fertility rate), Household Size, Life expectancy at 60 years($e_{60}$), GD PPPP, Urbanization. Stepwise Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100.

Data sources and variable meanings:

The data from the International Agency for Research on Cancer: cancer incidence rate (per 100,000 in 2012) by sex (total, male and female, age group (0-49 and all ages)

The World Bank data: Total Fertility Rate (the mean number of children born to a woman between 2009-2011), GD PPPP (per capita purchasing power parity in current international $ in 2010) and Urbanization (the percentage of total population living urban areas in 2010)

The United Nations data: Household size (the number of persons living in a household in 2010) and life expectancy (e60, 2005-2010)

Comparing with those correlations in all countries and in developed world, correlations between family size and cancer variables became weak and/or insignificant in the developing country grouping when standard deviation of cancer variable became low (Supplementary File 4: Table S2).

As the proxy of family size, household size was identified as the significant predictor of the total, male and female cancer incidence rates in samples of all ages and in age group 0-49 years respectively when household size, GDP, urbanisation, life expectancy (not for age group 0-49 years) and $I_{bs}$ were entered as the independent variables/predictors in the standard multiple linear (stepwise) regression analysis (Supplementary File 5: Table S3).

## Discussion

Cancer risk has been associated with multiple aetiologies, which may act through various mechanisms. Our results showed that: 1) Worldwide, smaller family size may be an independent determinant of increased cancer risk. 2) increased family size may show more protecting effects on cancer risk in males than females.

It is necessary to note the limitations of our work before analysing the public health implications of this study:

First, we must highlight the ecological fallacy (intrinsic limitation) arising from the ecological study approach which was adopted in this study. The data included in this study were calculated for country/populations as a whole. Thus, values for risk-modifying factors do not always hold true for individuals to predict their cancer risk. However, we would like to note that it is nearly impossible to test the relationships at the individual family level due to rare occurrence rate of cancers, and even rarer in some individual site-specific cancers, such as ovarian cancer.

Second, it is true that family size and family attitudes are influenced by many cultural, religious, economic, and social factors that vary substantially across different countries. However, there are no measures of such differeneces that can be used as confounders in our data analysis.

Third, data compiled and/or collected by the major international agencies (WHO, IARC, the United Nations and the World Bank) might be crude and may contain some random errors arising from methods of reporting incidence of specific diseases, reliability of diagnoses and possible administrative errors.

Finally, the observational data were used in our work, which makes the results subject to inherent bias, i.e. “correlation between two variables does not imply causality”.

Despite these limitations, findings from different data analyses in this study consistently show that country with greater TFR (family size) has lower cancer incidence rate regardless of age range and sex. This relationship trend has been observed in the correlations between family size and not only the individual site cancers, but also all cancers (in males, females and both sexes). The broad correlations between family size and cancers expressed in different individual sites, sexes and groups, may not be simply explained by the female hormonal fluctuation due to pregnancy and breastfeeding.

The relationship between psychological well-being and diseases (body and mind) has been an old issue. In the past, research into well-being has mainly focused on negative attitudes and affects. The majority of the studies documented that negative life events (death, divorce, injury, car crash *etc*.,), stressful life style, depression and/or anxiety, may lead to developing cancers [31, 60]. However, there is a documented bias in the data collected from the individual based surveys. In general, cancer patients tend to report negative events in excess compared to other people with average or positive attitudes [27, 29, 30]. This has been reported or reflected in a number of studies [26-29] regarding the relationship between cancer risk and adverse life events. According to the ancient Chinese medicine textbook, which was compiled 2,200 years ago, it has been believed that people have five internal organs of five gases (five emotions), i.e. happiness, anger, sadness, worry and fear. Among these five gases, only happiness makes the gas smooth [61], which keeps people healthy.

Family has long been cited as a health promoting factor [62, 63], and family size has been associated with life satisfaction [42, 43, 64]. From the perspective of evolution, humans have adapted early to cooperative breeding [65, 66], and then evolved alloparental care [67], and biological foundations of such human love may be heritable generation by generation [68]. Our study has revealed that greater family size, and possibly its associated positive psychological well-being, may play a protective role against cancer initiation. The mechanisms may include following aspects:

***1. Physiological and pathological functions of oxytocin in human health***

Positive psychological well-being may make the functions of neuroendocrine and immune systems more efficient, which may reduce the risk of developing cancer [60, 69-71].

Oxytocin is a peptide hormone and neuropeptide. Its production is associated with good feelings and emotions [72]. Males and females can produce and release similar quantities of oxytocin [73] within the hypothalamo-pituitary magnocellular systems. Researches constantly revealed that family related activities are the major promotors of oxytocin production. A stream of studies in the last decade reported that oxytocin release is not only associated with giving birth [74] and lactation [75], but also with daily interactions between family members, such as

spouses [32-34], mother and children [35], and father and children [36]. Oxytocin may be able to keep family happy and stable as it makes females and males stay monogamous [37, 38] and as it may bring positive psychological well-being to the family members. A self-reinforcing cycle is formed between family members interactions and oxytocin production.

Concurrently with the research into oxytocin production, physiological and pathological functions of oxytocin in humans have been the foci of numerous studies. Oxytocin has been postulated to have a role in inhibiting proliferation of human cancer cells, which may offer protective role in preventing cancer initiation [40]. The inhibitory role of oxytocin has been tested in individual site specific cancers, such as human breast cancer [40, 76] and ovarian cancer (animal model) [11]. A recent study reported that oxytocin, selectively activated by peptidylglycine α -amidating monooxygenase (PAM), may play a role in preventing and controlling a small cell lung cancer [77].

Bai *et al.* [26] reported that women with overall life satisfaction had less chance developing breast cancer. This may partly be true because life satisfaction may promote women to produce more oxytocin to prevent breast cancer cell initiation and proliferation. Another mechanism may be that greater TFR may make women produce less oestrogen and less menstrual cycles [9].

***2. Less cancer genes/mutations accumulated in population with greater TFR/family size***

Natural selection acts on each population [52, 78]. The total opportunity for natural selection in each population has been previously measured with the Biological State Index ($I_{bs}$) [48, 49, 52-55, 79]. An $I_{bs}$ value of one indicates total adaptation of the population to their environment. An $I_{bs}$ value of zero signifies a total lack of adaptation (inability to overcome natural selection pressures that are present), and an impossibility to give life to the next generation [48, 49, 52-55, 79].

Our study indicated that Biological State Index ($I_{bs}$) was in negative, strong and significant correlation to TFR/family size globally, in developed world and developing country groups respectively (Table 1). This means that population with greater TFR/family size is subject to more effective natural selection. As the consequence of less fitness, mortality rate due to various diseases, such as cancers, may increase [48, 49, 52-55, 78, 79]. Thus, cancer genes/mutations would be more often eliminated from a population with greater TFR/family size. Moreover, greater total fertility rates indicate less birth control therefore allowing more biological variation in fertility [80]. A portion of this variation, however small, provides opportunity for natural selection [80].

### *3. Family support and healthy lifestyle*

Family members from the greater family size may interact with each other more often to create life satisfaction [42, 43]. Meanwhile, one family member can remind and/or recommend other members to have necessary medical examination and have a healthy lifestyle [41].

Bai *et al.* [26] reported that people with positive psychological well-being may practice healthy lifestyle, have the knowledge of cancer risks and benefits of regular physical examination. It was reported that such positive psychological well-being may decrease the risk in the development of breast cancer [26, 31, 81].

In this study, we have also observed in Fisher's analysis that family size was in significantly stronger correlation with all cancers incidence in males (all ages) than it was with all cancers incidence in females (all ages). This finding is supported by the studies which found that males psychologically benefited more from having an extended kinship network than females [41, 62, 63]. However, this finding is inconsistent with Feller's finding that reduced life satisfaction was more related to the development of cancer in women than in men [82]. The reason for this inconsistency might be that Feller's data collection was based on the individual survey, which could be easily biased [30].

Family size has been implicated in the aetiologies of several individual site cancers, in previous studies based on the data collected at the individual level. Our findings were in agreement with the conclusions from the previous studies that greater family size was negatively correlated to the risks of developing bladder cancer [5], breast cancer [5, 6], colorectum cancer [5] and melanoma of skin. Although correlation does not necessarily imply causality, it may be suggested that increased family size may protect against the incidence of corpus uteri cancer and ovary cancer, but increase the risk of developing cervical cancer. These findings were in agreement with the prevailing dogma about the relationship between parity and gynecologic cancers, that is that more childbearing (greater TFR) may protect against corpus uteri cancer [7] and ovary cancer [8] due to less oestrogen production (less menstrual cycles) [9] but may contribute to cervix uteri cancer because of more exposure to infection risk [12]. However, our results were not supported by the findings from the study conducted by Hemminki *et al.* [5] that there were no reportable significant correlations between family size and risks of cervix uteri cancer, corpus uteri cancer and ovarian cancer. A number of studies have reported that ageing is one of the major contributors of corpus uteri cancer [83] and ovary cancer [41]. that findings of Hemminki *et al.* [5] were not compatible with our findings may be because only young females (aged mostly 5-43 years, up to 55 years) were included in their studies.

Poor hygiene level related infection with human papillomavirus is associated with cervical cancer initiation [84]. In the developed world, like Sweden, high level of hygiene or sanitation is accessible to almost all the residents. This reduces risk for females to have human papilloma virus infection, which may decrease the cervical cancer risk. This may be the explanation why Hemminki *et al.* [5] did not find the correlation between family size and cervical cancer incidence.

Blaser *et al.* [13] have reported that greater family size increased the risk of developing stomach cancer only for male family members, but not for all family members or female family members [13]. Aldrich *et al.* [14] reported that greater household size correlated with higher risk of lung cancer only in African Americans, but not in Latinos. However, sex specific or ancestry specific site cancer incidence was not included in our study. Thus, we may not be able to align our findings with the conclusions drawn by Blaser *et al.* [13] or Aldrich *et al.* [14].

The correlations, especially the partial correlations between family size and cancer variables in developing world were not as strong or significant as those identified in all countries (n=178) and developed world (n=98). This may be due to small variances (low standard deviations) of cancer incidence variables, which may reduce the covariance (correlation between family size and cancer variable), compared to those in the developed world and all countries grouping.

We must note an important strength of our study. Cancer risk studies based on surveys of individual persons have demonstrated a bias that is, in general, cancer patients tend to exaggerate negative life events in comparison to people with average or positive attitudes [30]. The methods employed in this study may have excluded this major bias because: 1) we used the objective measurement (TFR), instead of individual subjective psychological feeling assuming that TFR may be the family happiness index; 2) ecological study at population/group level, rather than individual based research method was adopted in this study. Ecological studies are based on aggregated quantitative data, not on the interviews with individual patients, so they are often used to determine the presence of effect of cancer risk-modifying factors in advance of, or impossible to identify in other epidemiological or laboratory approaches. Therefore, ecological study may be a better method to conduct the study of cancer incidence and its potential predictors, as cancer is one of the relatively rare diseases.

## Conclusions

In this study of the relationship between the family size and cancer incidence in 178 countries, we have identified that countries with greater family size have lower cancer incidence rates in males and females. This indicates that in terms of cancer prevention it may be worthwhile to consider whether both females, and especially males may benefit from greater family size.

Having more family life satisfaction may be included as a part of strategic plan of cancer prevention.

**List of abbreviations:**

TFR: Total fertility rate

FS: Family size

IARC: The International Agency for Research on Cancer of World Health Organization

GDP: Gross Domestic Product

UN: United Nations

WHO: World Health Organization

**Declarations**

Ethics approval and consent to participate: All the data supporting our findings in this paper were freely downloaded from the UN agencies' websites. No ethical approval or written informed consent for participation was required.

**Consent for publication:** Not applicable.

**Availability of data and materials**

All data for this study are publicly available and are ready for the public to download at no cost from the official websites of the World Bank, IARC and the WHO. Use of these data for this research falls within these UN agency's public permission in their terms and conditions. There is no need to have the formal permission to use the data for this study. The sources and data robustness have been described in the section of "Materials and Methods".

**Competing interest:** All the authors declared that there is no conflict of interest.

**Funding:** The work had no specific funding.

**Authors' contributions:** WPY conceived the idea and all authors consolidated the hypothesis together. WPY drafted the text with contributions from MH, FR and RH. WPY obtained the data, MH, WPY and FR conducted analyses, all authors interpreted the data analysis result and provided suggestions for their interpretations. MH performed the final statistical analysis. All authors reviewed, edited and approved the final manuscript.

**Acknowledgments**

The authors express appreciation to Jacques Ferlay from the International Agency for Research on Cancer of World Health Organization for the assistance in locating and defining the data.

## Reference

1. **World Bank Open Data** [http://data.worldbank.org/]
2. United Nations-Department of Economic and Social Affairs-Population Division: **World Population Prospects: The 2015 Revision, DVD Edition**. In.; 2015.
3. Bevier M, Weires M, Thomsen H, Sundquist J, Hemminki K: **Influence of family size and birth order on risk of cancer: a population-based study**. *BMC cancer* 2011, **11**:163.
4. **Australian Social Trends, 1996**
5. Hemminki K, Mutanen P: **Birth order, family size, and the risk of cancer in young and middle-aged adults**. *British Journal of Cancer* 2001, **84**(11):1466–1471.
6. Collaborative Group on Hormonal Factors in Breast Cancer: **Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50 302 women with breast cancer and 96 973 women without the disease**. *The Lancet* 2002, **360**(9328):187-195.
7. Saso S, Chatterjee J, Georgiou E, Ditri AM, Smith JR, Ghaem-Maghami S: **Endometrial cancer**. *BMJ* 2011, **343**:d3954.
8. **Ovarian cancer risks and causes** [http://www.cancerresearchuk.org]
9. WHO: **Bulletin of the World Health Organization: The breast cancer conundrum**. In: *Bull World Health Organ.* 2013: 626–627.
10. Imanieh MH, Bagheri F, Alizadeh AM, Ashkani-Esfahani S: **Oxytocin has therapeutic effects on cancer, a hypothesis**. *European Journal of Pharmacology* 2014, **741**:112-123.
11. Morita T, Shibata K, Kikkawa F, Kajiyama H, Ino K, Mizutani S: **Oxytocin inhibits the progression of human ovarian carcinoma cells in vitro and in vivo**. *International journal of cancer* 2004, **109**(4):525-532.
12. **Risk factors for cervical cancer** [http://www.cancer.ca]
13. Blaser MJ, Nomura A, Lee J, Stemmerman GN, Perez-Perez GI: **Early-Life Family Structure and Microbially Induced Cancer Risk**. *PLoS medicine* 2007, **4**(1):0053-0058.
14. Aldrich MC, Selvin S, Wrensch MR, Sison JD, Hansen HM, Quesenberry CP, Seldin MF, Barcellos LF, Buffler PA, Wiencke JK: **Socioeconomic status and lung cancer: unraveling the contribution of genetic admixture**. *American journal of public health* 2013, **103**(10):e73.
15. Garssen B: **Psychological factors and cancer development: Evidence after 30 years of research**. *Clinical Psychology Review* 2004, **24**(3):315-338.
16. Levenson JL, Bemis C: **The Role of Psychological Factors in Cancer Onset and Progression**. *Psychosomatics* 1991, **32**(2):124-132.

17. Chen CC, David AS, Nunnerley H, Michell M, Dawson JL, Berry H, Dobbs J, Fahy T: **Adverse life events and breast cancer: case-control study**. *BMJ* 1995, **311**(7019):1527.

18. Lillberg K, Verkasalo P, Kaprio J, Teppo L, Helenius H, Koskenvuo M: **A prospective study of life satisfaction, neuroticism and breast cancer risk (Finland)**. *Cancer Causes Control* 2002, **13**(2):191-198.

19. Eskelinen M, Ollonen P: **Life stress and losses and deficit in adulthood as breast cancer risk factor: a prospective case-control**. *In Vivo* 2010, **Nov-Dec; 24**(6):899-904.

20. Lillberg K: **Stressful Life Events and Risk of Breast Cancer in 10,808 Women: A Cohort Study**. *American journal of epidemiology* 2003, **157**(5):415-423.

21. Chorot P, Sandín B: **Life events and stress reactivity as predictors of cancer, coronary heart disease and anxiety disorders**. *Int J Psychosom **Res*** 1994, **41**:34–40.

22. Ginsberg A, Price S, Ingram D, Nottage E: **Life events and the risk of breast cancer: a case- control study**. *European Journal of Cancer* 1996, **32**(12):2049-2052.

23. Kune S, Kune GA, Watson LF, Rahe RH: **Recent life change and large bowel cancer. Data from the Melbourne Colorectal cancer study**. *Journal of Clinical Epidemiology* 1991, **44**(1):57-68.

24. Levav I, Kohn R, Iscovich J, Abramson JH, Tsai WY, Vigdorovich D: **Cancer incidence and survival Following Bereavement**. *American journal of public health* October 2000, **90**(10):160-1607.

25. Everson SA, Goldberg DE, Kaplan GA, Cohen RD, Pukkala E, Tuomilehto J, Salonen JT: **Hopelessness and risk of mortality and incidence of myocardial infarction and cancer**. *Psychosomatic medicine* 1996, **58**(2):113.

26. Bai A, Li H, Huang Y, Liu X, Gao Y, Wang P, Dai H, Song F, Hao X, Chen K: **A survey of overall life satisfaction and its association with breast diseases in Chinese women**. *Cancer medicine* 2016, **5**(1):111-119.

27. Courtney JG, Longnecker MP, Theorell T, Deverdier MG: **Stressful Life Events and the Risk of Colorectal Cancer: Results from a case-control study**. *Epidemiology* 1993, **138**(8):628-628.

28. Cohen LH, Towbes LC, Flocco R: **Effects of Induced Mood on Self- Reported Life Events and Perceived and Received Social Support**. *Journal of Personality and Social Psychology* 1988, **55**(4):669-674.

29. Brett JF, Brief AP, Burke MJ, George JM, Webster J: **Negative Affectivity and the Reporting of Stressful Life Events**. *Health Psychology* 1990, **9**(1):57-68.

30. Blaney PH: **Affect and Memory: A Review**. *Psychological Bulletin* 1986, **99**(2):229-246.

31. Peled R, Carmil D, Siboni-Samocha O, Shoham-Vardi I: **Breast cancer, psychological distress and life events among young women**. *BMC cancer* 2008, **8**:245.

32. Carmichael MS, Humbert R, Dixen J, Palmisano G, Greenleaf W, Davidson JM: **Plasma oxytocin increases in the human sexual response**. *The Journal of Clinical Endocrinology & Metabolism* 1987, **64**(1):27-31.

33. Carmichael MS, Warburton VL, Dixen J, Davidson JM: **Relationships among cardiovascular, muscular, and oxytocin responses during human sexual activity**. *Archives of sexual behavior* 1994, **23**(1):59-79.

34. Gordon Jr G, Burch RL, Platek SM: **Does semen have antidepressant properties?** *Archives of Sexual Behavior* 2002, **31**(3):289-293.

35. Kendrick KM: **The neurobiology of social bonds**. *Journal of neuroendocrinology* 2004, **16**(12):1007-1008.

36. Weisman O, Zagoory-Sharon O, Feldman R: **Oxytocin administration to parent enhances infant physiological and behavioral readiness for social engagement**. *Biological psychiatry* 2012, **72**(12):982-989.

37. Insel TR, Hulihan TJ: **A gender-specific mechanism for pair bonding: oxytocin and partner preference formation in monogamous voles**. *Behavioral neuroscience* 1995, **109**(4):782.

38. Young LJ, Murphy Young AZ, Hammock EA: **Anatomy and neurochemistry of the pair bond**. *Journal of Comparative Neurology* 2005, **493**(1):51-57.

39. Sapino A, Macri L, Tonda L, Bussolati G: **Oxytocin enhances myoepithelial cell differentiation and proliferation in the mouse mammary gland**. *Endocrinology* 1993, **133**(2):838-842.

40. Cassoni P, Sapino A, Negro F, Bussolati G: **Oxytocin inhibits proliferation of human breast cancer cell lines**. *Virchows Archiv* 1994, **425**(5):467-472.

41. Aizer AA, Chen MH, McCarthy EP, Mendu ML, Koo S, Wilhite TJ, Graham PL, Choueiri TK, Hoffman KE, Martin NE *et al*: **Marital status and survival in patients with cancer**. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2013, **31**(31):3869-3876.

42. Angeles L: **Children and Life Satisfaction**. *Journal of Happiness Studies* 2009, **11**(4):523-538.

43. Nan H, Ni MY, Lee PH, Tam WW, Lam TH, Leung GM, McDowell I: **Psychometric evaluation of the Chinese version of the Subjective Happiness Scale: evidence from the Hong Kong FAMILY Cohort**. *International journal of behavioral medicine* 2014, **21**(4):646-652.

44. **GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]** [http://globocan.iarc.fr]

45. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F: **Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012**. *International journal of cancer* 2015, **136**(5):E359-386.

46. **Abridged life table, for females, by major area, region and country, 1950-2100**

47. **Demographic and Social Statistics** [http://unstats.un.org]

48. Henneberg M: **Reproductive possibilities and estimations of the biological dynamics of earlier human populations**. *Journal of Human Evolution* 1976, **5**:41-48.

49. Henneberg M, Piontek J: **Biological state index of human groups**. *Przeglad Anthropologiczny* 1975, **XLI**:191-201.

50. **World Fertility Data 2008** [http://www.un.org]

51. WHO: **World Health Statistics 2012**. Geneva: World Health Organization; 2012.

52. Stephan CN, Henneberg M: **Medicine may be reducing the human capacity to survive**. *Medical hypotheses* 2001, **57**(5):633-637.

53. You W, M H: **Cancer incidence increasing globally: The role of relaxed natural selection**. *Evol Appl* 2017, **00:1–13**.

54. Budnik A, Henneberg M: **Worldwide Increase of Obesity is Related to the Reduced Opportunity for Natural Selection**. *PloS one* 2017, **12**(1).

55. You W-P, Henneberg M: **Type 1 diabetes prevalence increasing globally and regionally: the role of natural selection and life expectancy at birth**. *BMJ Open Diabetes Research & Care* 2016, **4**(1):e000161.

56. Allender S, Foster C, Hutchinson L, Arambepola C: **Quantification of urbanization in relation to chronic diseases in developing countries: a systematic review**. *Journal of urban health : bulletin of the New York Academy of Medicine* 2008, **85**(6):938-951.

57. Greenberg M: **Urbanization and Cancer: Changing Mortality Patterns?** *International Regional Science Review* 1983, **8**(2):127.

58. Majeed A, Babb P, Jones J, Quinn M: **Trends in prostate cancer incidence, mortality and survival in England and Wales 1971–1998**. *BJU International* 2000, **85**(9):1058-1062.

59. United Nations DoEaSA, Population Division,,: **World Population Prospects: The 2017 Revision, DVD Edition**. In.

60. Williams RB, Schneiderman N: **Resolved: Psychosocial interventions can improve clinical outcomes in organic disease (pro)**. *Psychosomatic Medicine* 2002, **64**:552–557.

61. Huangdi, Qibo, Other ministers: **Huangdi Neijing** Wuhan: Wuhan Publishing House; 2013.

62. Ploubidis GB, Silverwood RJ, DeStavola B, Grundy E: **Life-Course Partnership Status and Biomarkers in Midlife: Evidence From the 1958 British Birth Cohort**. *American journal of public health* 2015, **105**(8):1596-1603.

63. Cable N, Bartley M, Chandola T, Sacker A: **Friends are equally important to men and women, but family matters more for men's well-being**. *Journal of epidemiology and community health* 2013, **67**(2):166-171.

64. Institute of Population and Labor Economy of Chinese Academy of Social Sciences: **Chinese Family Happiness Hot Issues Investigation Report**, 1 edn. Beijing: Xinhua Publishing House; 2015.

65. Clark G, Henneberg M: **The life history of Ardipithecus ramidus: a heterochronic model of sexual and social maturation**. *Anthropological Review* 2015, **78**(2).

66. K H, RR P: **The Evolution of Human Life History**. Santa Fe: School of American Research Press; 2006.

67. Isler K, van Schaik CP: **Allomaternal care, life history and brain size evolution in mammals**. *J Hum Evol* 2012, **63**(1):52-63.

68. Henneberg M, Saniotis A: **The Dynamic Human**. Sharjah, U.A.E.: Bentham Science Publishers Ltd.; 2016.

69. Antonova L, Mueller CR: **Hydrocortisone down-regulates the tumor suppressor gene BRCA1 in mammary cells: a possible molecular link between stress and breast cancer**. *Genes, chromosomes & cancer* 2008, **47**(4):341-352.

70. Cohen S, Rodriguez MS: **Pathways Linking Affective Disturbances and Physical Disorders**. *Health Psychology* 1995, **14**(5):374-380.

71. Diener E, Chan MY: **Happy People Live Longer: Subjective Well-Being Contributes to Health and Longevity**. *Applied Psychology: Health and Well-Being* 2011, **3**(1):1-43.

72. Magon N, Kalra S: **The orgasmic history of oxytocin: Love, lust, and labor**. *Indian journal of endocrinology and metabolism* 2011, **15**(Suppl3):S156.

73. Ivell R, Balvers M, Rust W, Bathgate R, Einspanier A: **Oxytocin and male reproductive function**. *The fate of the male germ cell* 1997.

74. Takayanagi Y, Yoshida M, Bielsky IF, Ross HE, Kawamata M, Onaka T, Yanagisawa T, Kimura T, Matzuk MM, Young LJ: **Pervasive social deficits, but normal parturition, in oxytocin receptor-deficient mice**. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(44):16096-16101.

75. White-Traut R, Watanabe K, Pournajafi-Nazarloo H, Schwertz D, Bell A, Carter CS: **Detection of salivary oxytocin levels in lactating women**. *Developmental psychobiology* 2009, **51**(4):367-373.

76. Murrell TG: **The potential for oxytocin ( OT) to prevent breast cancer: a hypothesis**. *Breast cancer research and treatment* 1995, **35**(2):225.

77. Cao F, Gamble AB, Kim H-k, Onagi H, Gresser MJ, Kerr J, Easton CJ: **Potent and selective inhibitors of human peptidylglycine - amidating monooxygenase**. *Med Chem Commun* 2011, **2**(8):760-763.

78. Darwin C: **The descent of man, and selection in relation to sex**, New edn. London: John Murray; 1901.

79. Rühli F, Henneberg M: **Biological future of humankind – ongoing evolution and the impact of recognition of human biological variation**. In: *On Human Nature Biology, Psychology, Ethics, Politics, and Religion 2016.* edn. Edited by Tibayrenc M, Ayala FJ: Elsevier; 2016: 263-275.

80. Henneberg M: **Natural selection through differential fertility in human populations: Quantitative evaluation of selection intensity**. *Przeglad Antropologiczny* 1980, **46**:21-60.

81. Kim Y, Duhamel KN, Valdimarsdottir HB, Bovbjerg DH: **Psychological distress among healthy women with family histories of breast cancer: effects of recent life events**. *Psycho-oncology* 2005, **14**(7):555-563.

82. Feller S, Teucher B, Kaaks R, Boeing H, Vigl M: **Life satisfaction and risk of chronic diseases in the European prospective investigation into cancer and nutrition (EPIC)-Germany study**. *PloS one* 2013, **8**(8):e73462.

83. Purdie DM, Green AC: **Epidemiology of endometrial cancer**. *Best practice & research clinical obstetrics & gynaecology* 2001, **15**(3):341-354.

84. Snijders PJ, Steenbergen RD, Heideman DA, Meijer CJ: **HPV-mediated cervical carcinogenesis: concepts and clinical implications**. *The Journal of pathology* 2006, **208**(2):152-164.

## Supplementary files

Supplementary File 1:

Table S1: The whole set of data for this study

| Country/Population | Total Fertility Rate Mean of 2009-2011 | Household Si | All cancers excl. non-melanoma skin cancer (C00-97, but C44) - all ages total* | All cancers excl. non-melanoma skin cancer (C00-97, but C44)- all ages female | All cancers excl. non-melanoma skin cancer (C00-97, but C44) - all ages male | All cancers excl. non-melanoma skin cancer (C00-97, but C44) – 0-49 total | All cancers excl. non-melanoma skin cancer (C00-97, but C44) – 0-49 female | All cancers excl. non-melanoma skin cancer (C00-97, but C44) – 0-49 male | Bladder (C67), all ages | Breast(C50), all ages | Cervix uteri (C53), all ages | Colorectum (C18-21), all ages | Corpus uteri (C54), all ages | Lung (C33-34), all ages | Melanoma of skin (C43), all ages | Ovary (C56), all ages | Stomach (C16), all ages | GDP PPP 2010 | Urbanization 2010 | Life Expectancy ($e_{60}$, 2005-2010) | Biological State Index ($I_{bs}$) | Income group, The World Bank |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Afghanistan | 5.661 | | 115.2 | 119.5 | 112.4 | 30.830 | 36.136 | 25.897 | 3.3 | 35.1 | 8.8 | 4.9 | 7.9 | 6.9 | 0.6 | 3.8 | 12.7 | 1604.1915 | 24.689 | 16 | 0.719929791 | Low income |
| Albania | 1.744 | | 178.3 | 173.2 | 185 | 66.513 | 86.508 | 46.831 | 10 | 53.9 | 5 | 8.4 | 11.1 | 26.2 | 0.9 | 3.2 | 20.1 | 9480.8862 | 52.163 | 20 | 0.975085402 | Upper middle income |
| Algeria | 2.809 | 5.54229439 | 123.5 | 132.7 | 116.2 | 49.966 | 65.988 | 34.498 | 5.9 | 48.5 | 8.5 | 11.5 | 1.5 | 9.9 | 0.5 | 5 | 6 | 12241.441 | 67.526 | 18 | 0.954124186 | Upper middle income |
| Angola | 6.216 | | 100.8 | 112.2 | 89.9 | 22.851 | 31.967 | 13.676 | 2.3 | 23.5 | 35.5 | 5 | 3.2 | 2 | 1.7 | 3.5 | 3.8 | 6904.5626 | 40.097 | 15 | 0.76887096 | Lower middle income |
| Argentina | 2.215 | | 216.7 | 211.8 | 230.4 | 63.639 | 82.286 | 45.312 | 6 | 71.2 | 20.8 | 23.8 | 7.8 | 20.9 | 2.9 | 8.7 | 6.7 | | 90.966 | 21 | 0.976303124 | Upper middle income |
| Armenia | 1.737 | 4.12331532 | 257 | 226.4 | 305.6 | 84.665 | 107.687 | 60.532 | 12.3 | 74.1 | 13.8 | 19.3 | 26.7 | 35.9 | 1.8 | 8.5 | 15.1 | 6376.3262 | 63.58 | 20 | 0.976385611 | Lower middle income |
| Australia | 1.914 | 2.61385088 | 323 | 278.6 | 373.9 | 99.536 | 121.391 | 78.321 | 7.6 | 86 | 5.5 | 38.4 | 12.1 | 27 | 34.9 | 7.6 | 4.8 | 39048.153 | 88.733 | 25 | 0.989885167 | High income |
| Austria | 1.420 | 2.4055178 | 254.1 | 222.7 | 295.2 | 92.790 | 111.646 | 74.461 | 11.4 | 68 | 5.8 | 26 | 10.4 | 27.5 | 9.9 | 7.3 | 6.8 | 41677.506 | 65.852 | 23 | 0.990942956 | High income |
| Azerbaijan | 1.900 | 4.89121243 | 141.9 | 124 | 165.8 | 55.212 | 66.053 | 44.674 | 2.7 | 25.4 | 9.8 | 6.7 | 2.8 | 11.5 | 0.8 | 2.5 | 13 | 15627.748 | 53.401 | 18 | 0.959516066 | Upper middle income |
| Bahamas | 1.899 | | 208.9 | 223.4 | 199.5 | 74.060 | 118.840 | 29.428 | 2 | 98.9 | 20.6 | 20.3 | 15.5 | 9.7 | 0.9 | 8.9 | 7 | 22468.542 | 82.549 | 22 | 0.973918012 | High income |
| Bahrain | 2.145 | | 112.4 | 121.9 | 112.8 | 26.302 | 39.310 | 18.488 | 6.3 | 42.5 | 5.9 | 11.3 | 4.7 | 15.5 | 0.3 | 4.4 | 3.9 | 39732.914 | 88.535 | 19 | 0.98111564 | High income |
| Bangladesh | 2.280 | | 104.4 | 100 | 109.4 | 34.224 | 45.799 | 22.973 | 1.6 | 21.7 | 19.2 | 3.6 | 1.5 | 10 | 0.1 | 4.4 | 5.7 | 2409.5485 | 30.462 | 18 | 0.920698299 | Lower middle income |
| Barbados | 1.839 | | 263.1 | 258.1 | 277.2 | 90.265 | 118.901 | 63.281 | 4.4 | 94.7 | 25.4 | 28.4 | 34.1 | 4.9 | 0 | 7 | 6.7 | 15025.027 | 32.06 | 19 | 0.981286541 | High income |
| Belarus | 1.457 | 2.99571216 | 218.7 | 190.6 | 275.5 | 82.496 | 102.905 | 62.079 | 6.5 | 45.9 | 13.2 | 24.4 | 17.1 | 26.2 | 4.1 | 10.9 | 18.8 | 15385.599 | 74.615 | 17 | 0.982938077 | Upper middle income |
| Belgium | 1.837 | 2.43570188 | 321.1 | 288.9 | 364.8 | 111.733 | 145.022 | 79.294 | 17.5 | 111.9 | 8.6 | 36.7 | 13.2 | 36.8 | 12.1 | 8.1 | 5.8 | 39215.713 | 97.641 | 23 | 0.990156593 | High income |
| Belize | 2.803 | | 160.7 | 161.2 | 160.6 | 32.694 | 42.400 | 22.761 | 3.2 | 39.6 | 32.7 | 9 | 15.9 | 11.8 | 0.7 | 3.2 | 5.7 | 8042.9802 | 44.963 | 21 | 0.969397579 | Upper middle income |
| Benin | 5.096 | | 94.3 | 102.7 | 87.2 | 26.129 | 32.181 | 20.044 | 1.7 | 30.2 | 27.6 | 4.4 | 3.4 | 1.5 | 0.8 | 3.2 | 3.7 | 1603.4629 | 41.854 | 15 | 0.834449996 | Low income |
| Bhutan | 2.378 | | 79.2 | 77.1 | 82 | 22.868 | 26.828 | 19.361 | 1 | 4.6 | 12.8 | 3.5 | 0 | 6.9 | 0.9 | 5.2 | 17.2 | 6383.7627 | 34.793 | 19 | 0.894327782 | Lower middle income |
| Bolivia | 3.359 | 2.9777716 | 143.9 | 164.3 | 123.9 | 40.765 | 58.469 | 23.384 | 2.4 | 19.2 | 47.7 | 9.1 | 4 | 5.1 | 2.5 | 6.9 | 7.8 | 5172.148 | 66.426 | 18 | 0.930648766 | Lower middle income |
| Bosnia and Herzeg | 1.242 | | 161.1 | 147.8 | 180 | 61.941 | 78.340 | 45.209 | 7 | 37.4 | 13.7 | 16.6 | 9.6 | 26.5 | 1.9 | 8.1 | 8 | 8746.0806 | 39.226 | 20 | 0.930648766 | Upper middle income |
| Botswana | 2.761 | | 107.6 | 104.7 | 113.9 | 36.991 | 43.901 | 30.442 | 0.9 | 19.9 | 30.3 | 3.5 | 4 | 4.8 | 1.7 | 3.1 | 0.9 | 13079.289 | 56.235 | 16 | 0.883605129 | Upper middle income |
| Brazil | 1.841 | 3.7937745 | 205.5 | 186.8 | 231.6 | 58.890 | 76.672 | 41.123 | 4.5 | 59.5 | 16.3 | 15.8 | 5.6 | 16.3 | 2.8 | 5.2 | 9.2 | 13759.42 | 84.335 | 21 | 0.974586305 | Upper middle income |
| Brunei Darussalam | 2.051 | | 163.2 | 179 | 149.5 | 46.048 | 56.384 | 35.704 | 1.6 | 48.6 | 16.9 | 25 | 12.6 | 22.7 | 2.5 | 8.8 | 7.4 | 69208.221 | 75.51 | 21 | 0.984968274 | High income |
| Bulgaria | 1.580 | 2.80556541 | 234.8 | 220.1 | 260.5 | 93.372 | 119.497 | 68.249 | 11 | 58.5 | 24.5 | 31.5 | 17.8 | 28.1 | 3.4 | 14 | 10.3 | 14690.492 | 72.302 | 18 | 0.983594781 | Upper middle income |
| Burkina Faso | 5.868 | 5.9392007 | 88.2 | 99.8 | 75.9 | 23.147 | 30.954 | 15.445 | 2.9 | 22.7 | 23.3 | 2.5 | 2.3 | 2.4 | 0.7 | 3.1 | 3 | 1438.3186 | 25.665 | 15 | 0.634764538 | Low income |
| Burundi | 6.304 | | 135.8 | 143 | 132.2 | 35.729 | 44.700 | 26.673 | 2.1 | 23.5 | 49.3 | 6 | 2.7 | 1.7 | 1.7 | 4.5 | 4 | 710.62884 | 10.641 | 16 | 0.735917366 | Low income |
| Cambodia | 2.967 | | 140.4 | 134.1 | 155.3 | 41.912 | 50.133 | 33.721 | 2.2 | 19.3 | 23.8 | 8.2 | 2.5 | 12.4 | 0.7 | 4.5 | 5 | 2462.2446 | 19.81 | 23 | 0.887938675 | Lower middle income |
| Cameroon | 5.017 | | 97.6 | 114.1 | 81.2 | 43.681 | 56.881 | 30.634 | 0.8 | 35.2 | 30 | 3.3 | 2.8 | 1.5 | 1.1 | 4.9 | 2.4 | 2518.7852 | 51.516 | 16 | 0.760388823 | Lower middle income |
| Canada | 1.635 | | 295.7 | 277.4 | 320.8 | 92.742 | 121.021 | 65.372 | 11.5 | 79.8 | 6.3 | 35.2 | 16.3 | 37.9 | 9.6 | 8.6 | 4.9 | 39972.336 | 80.937 | 24 | 0.989113532 | High income |
| Cape Verde | 2.431 | | 74.9 | 88.4 | 60.9 | 22.180 | 32.974 | 11.790 | 1.2 | 25.1 | 29 | 3.5 | 3.7 | 0.8 | 0.4 | 4 | 4 | 5883.5288 | 61.833 | 19 | 0.963496649 | Lower middle income |
| Central African Rep | 4.627 | | 92.9 | 99.7 | 86.9 | 27.875 | 34.555 | 21.138 | 1.3 | 31.4 | 21 | 4.5 | 2.5 | 1.5 | 1.2 | 4.9 | 2.3 | 882.90446 | 38.828 | 15 | 0.713390767 | Low income |
| Chad | 6.595 | | 88.1 | 99.2 | 77.4 | 28.071 | 36.501 | 19.666 | 1.4 | 34.1 | 18.8 | 4.2 | 3 | 1.2 | 0.8 | 5.4 | 2 | 1913.051 | 21.983 | 15 | 0.705486484 | Low income |
| Chile | 1.858 | 3.87655766 | 175.7 | 163.3 | 195.3 | 48.107 | 59.627 | 36.798 | 3.9 | 34.8 | 12.8 | 15 | 5.4 | 13.3 | 1.5 | 6.6 | 15.6 | 18234.984 | 88.586 | 23 | 0.985685173 | High income |
| China | 1.650 | | 174 | 139.9 | 211.2 | 56.991 | 63.380 | 51.223 | 3 | 22.1 | 7.5 | 14.2 | 8.6 | 36.1 | 0.6 | 4.1 | 22.7 | 9043.7984 | 49.226 | 20 | 0.96889093 | Upper middle income |
| Colombia | 2.376 | 4.25625479 | 160.6 | 151.5 | 175.2 | 41.793 | 54.139 | 29.419 | 2.9 | 35.7 | 18.7 | 12.9 | 3.6 | 11 | 3.3 | 5.9 | 13.4 | 10558.984 | 75.036 | 21 | 0.973771988 | Upper middle income |
| Comoros | 4.918 | | 101.5 | 121.8 | 81.9 | 30.214 | 44.098 | 16.682 | 1.7 | 17.4 | 61.3 | 2.7 | 1.4 | 1.2 | 0.9 | 0.8 | 1.1 | 1349.0557 | 27.918 | 16 | 0.850014346 | Low income |
| Congo Democratic | 6.250 | | 107.8 | 115.2 | 102.5 | 20.825 | 27.029 | 14.628 | 2 | 23.5 | 33.1 | 5.3 | 3.8 | 1 | 2.2 | 3.7 | 5.4 | 671.00952 | 39.937 | 15.01 | 0.729168867 | Low income |
| Congo Republic of | 5.069 | | 88.2 | 94.1 | 83.7 | 16.653 | 21.016 | 12.354 | 0.5 | 31.7 | 25.2 | 5.6 | 2.4 | 1.1 | 2 | 3.3 | 2.7 | 5421.3133 | 63.228 | 17 | 0.810288325 | Lower middle income |
| Costa Rica | 1.849 | | 179.3 | 169.2 | 193.5 | 51.208 | 67.778 | 35.401 | 3.2 | 45.4 | 11.4 | 16.4 | 7.3 | 7.3 | 2.3 | 5.4 | 17.3 | 12073.493 | 71.734 | 23 | 0.983623262 | Upper middle income |
| Cote d'Ivoire | 4.909 | | 89 | 101 | 78.2 | 24.944 | 32.160 | 17.877 | 1.2 | 33.7 | 21.7 | 4.8 | 3.9 | 1.7 | 0.5 | 3.9 | 4.4 | 2833.7598 | 50.557 | 14 | 0.783166113 | Lower middle income |
| Croatia | 1.537 | 3.11524218 | 266.9 | 231.6 | 319.9 | 90.729 | 104.680 | 77.170 | 10.9 | 60.9 | 10 | 32.9 | 12.5 | 34.3 | 8.9 | 10.3 | 9.8 | 18968.886 | 57.537 | 20 | 0.988890534 | Upper middle income |
| Cuba | 1.469 | 3.23198513 | 218 | 190.3 | 250.8 | 63.545 | 84.844 | 43.188 | 7.4 | 50.4 | 17.1 | 19.7 | 14 | 32.9 | 0.8 | 6.9 | 5.9 | 17921.087 | 76.597 | 22 | 0.988866186 | Upper middle income |
| Cyprus | 1.478 | 3.0813039 | 204.7 | 198.2 | 218.2 | 64.199 | 91.459 | 39.332 | 11.6 | 78.4 | 4.1 | 24.5 | 10.7 | 16.2 | 3.2 | 7 | 5.4 | 31089.861 | 67.551 | 21 | 0.994060627 | High income |
| Czech Republic | 1.483 | 2.67265428 | 293.8 | 258.9 | 345.9 | 82.564 | 101.532 | 64.630 | 11.5 | 70.3 | 14.1 | 38.9 | 18 | 32.5 | 12.6 | 11.1 | 7.4 | 27053.983 | 73.255 | 21 | 0.991737284 | High income |
| Denmark | 1.820 | | 338.1 | 328.8 | 354.3 | 102.631 | 131.343 | 74.805 | 14.4 | 105 | 10.6 | 40.5 | 13.5 | 39.2 | 19.2 | 10.3 | 5.6 | 41811.606 | 86.795 | 22 | 0.991323324 | High income |
| Djibouti | 3.604 | | 92.7 | 111.3 | 73.7 | 30.065 | 42.310 | 18.016 | 2.2 | 35.9 | 17.3 | 6.1 | 2.5 | 2.7 | 0 | 7.4 | 2.7 | 2611.4591 | 76.996 | 17 | 0.845853923 | Lower middle income |
| Dominican Republi | 2.584 | 2.76402074 | 153.4 | 149.1 | 158.5 | 44.818 | 62.222 | 27.630 | 1.1 | 38.1 | 30.7 | 10.2 | 4 | 12 | 0.3 | 1.6 | 7.3 | 10862.084 | 73.752 | 21 | 0.956152384 | Upper middle income |
| Ecuador | 2.656 | | 164.5 | 169.2 | 162 | 48.158 | 65.180 | 31.468 | 2.2 | 32.7 | 29 | 10.7 | 3.8 | 7.2 | 1.7 | 5.2 | 16.9 | 9122.6938 | 62.69 | 23 | 0.962401931 | Upper middle income |
| Egypt | 2.882 | | 152 | 147.8 | 158.4 | 42.865 | 50.584 | 35.363 | 13.1 | 49.5 | 2.3 | 5.6 | 3.8 | 7.2 | 0.2 | 6.4 | 2.5 | 10405.85 | 43.019 | 17 | 0.970774956 | Lower middle income |
| El Salvador | 2.263 | | 153.4 | 167.2 | 136.6 | 47.819 | 66.709 | 27.604 | 1.5 | 23.7 | 24.8 | 8.5 | 16.3 | 5.9 | 0.2 | 4 | 16.4 | 7090.6338 | 64.286 | 22 | 0.969379018 | Lower middle income |
| Equatorial Guinea | 5.138 | | 86.4 | 98.5 | 76.1 | 38.608 | 50.651 | 26.997 | 2.6 | 25.2 | 25.1 | 4.3 | 5.4 | 3.9 | 0.2 | 4 | 2.3 | 33765.48 | 39.223 | 15 | 0.766811221 | Upper middle income |
| Eritrea | 4.969 | | 101.7 | 118.6 | 82.8 | 30.871 | 41.143 | 20.623 | 2.2 | 35.9 | 17.4 | 6 | 2.2 | 2.5 | 0.5 | 7.6 | 2.4 | 1061.3342 | 20.572 | 15 | 0.914840478 | Low income |
| Estonia | 1.677 | 8.42041474 | 242.8 | 202.7 | 321.9 | 72.368 | 93.414 | 51.564 | 7 | 51.6 | 19.9 | 27.2 | 14.6 | 24.4 | 7.4 | 11.8 | 13.8 | 21085.2 | 68.094 | 20 | 0.98914898 | High income |
| Ethiopia | 4.905 | 4.88315386 | 108 | 140.9 | 73.2 | 38.247 | 52.386 | 24.145 | 2.5 | 41.8 | 26.4 | 7.3 | 2 | 3.2 | 0.1 | 8.6 | 3 | 1059.4511 | 17.319 | 17 | 0.815116195 | Low income |
| Fiji | 2.671 | | 139.1 | 189.3 | 91.3 | 70.771 | 112.069 | 31.759 | 3.1 | 65 | 37.8 | 7 | 11.3 | 6 | 0.8 | 14.9 | 2.4 | 6954.0226 | 51.828 | 17 | 0.967330789 | Upper middle income |
| Finland | 1.853 | 2.01672397 | 256.8 | 234.2 | 290.1 | 75.169 | 95.695 | 55.500 | 8.3 | 89.4 | 4.3 | 23.5 | 13.9 | 20.1 | 12.6 | 8.4 | 5.2 | 38300.788 | 83.558 | 23 | 0.990882332 | High income |
| French Polynesia | 2.111 | | 255 | 227.3 | 287.4 | 69.026 | 104.407 | 35.419 | 6 | 92.2 | 8.2 | 13.3 | 15.2 | 37.1 | 5.9 | 9 | 7.4 | | 56.479 | 19 | | High income |
| FYR Macedonia | 1.452 | | 239.3 | 220.8 | 265.5 | 79.421 | 97.722 | 62.162 | 11.5 | 76.2 | 12.4 | 24.3 | 29 | 40.8 | 4.8 | 11.3 | 16.5 | 11366.125 | 56.992 | 19 | 0.986258559 | Upper middle income |
| Gabon | 4.213 | | 90.2 | 101.5 | 79.9 | 31.457 | 41.726 | 21.346 | 3.8 | 16.1 | 19.9 | 5.2 | 7.5 | 6.2 | 1.2 | 4.8 | 2.4 | 16387.647 | 85.697 | 18 | 0.889527438 | Upper middle income |
| Georgia | 1.817 | 3.51647578 | 181 | 163.7 | 207.8 | 72.862 | 95.825 | 49.485 | 4.1 | 44 | 14.2 | 8.5 | 14.2 | 15.9 | 1.7 | 3.7 | 9.6 | 5818.2471 | 52.869 | 19 | 0.962737814 | Lower middle income |
| Germany | 1.370 | 2.22772901 | 283.8 | 252.5 | 323.7 | 106.685 | 128.829 | 85.307 | 13.4 | 91.6 | 8.2 | 30.9 | 11.9 | 27.5 | 11.4 | 7.4 | 7.8 | 39557.603 | 74.291 | 23 | 0.992090967 | High income |
| Ghana | 4.052 | 3.45922966 | 91.7 | 104.8 | 79.2 | 27.423 | 37.305 | 17.966 | 2.1 | 25.6 | 35.4 | 4.1 | 3 | 2.4 | 0.8 | 4.1 | 2.9 | 3002.9708 | 50.713 | 15 | 0.884556472 | Lower middle income |
| Greece | 1.463 | 3.08614485 | 163 | 138 | 194.8 | 53.175 | 65.879 | 41.181 | 9.1 | 43.9 | 5.2 | 13.5 | 7 | 28.5 | 2.4 | 8.4 | 5.3 | 28904.089 | 76.292 | 23 | 0.991007185 | High income |

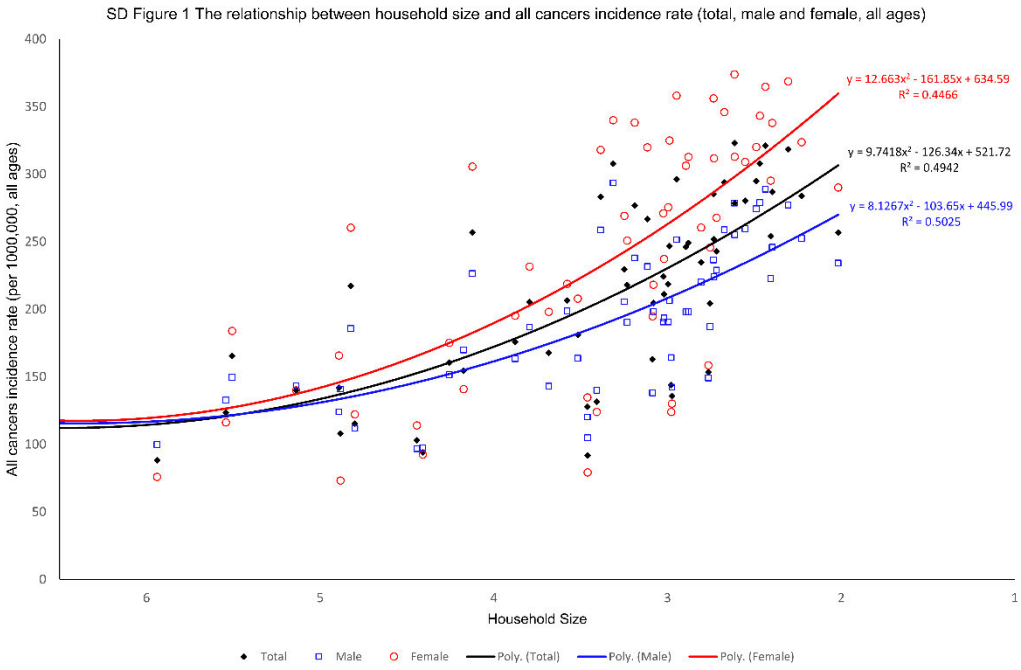
| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Guam | 2.473 | 3.68355304 | 167.7 | 143 | 198 | 48.963 | 61.739 | 36.826 | 2.9 | 49.4 | 9 | 20.5 | 12.2 | 34.7 | 2.9 | 2.8 | 1.6 | | 94.099 | 21 | | High income |
| Guatemala | 3.975 | | 130.4 | 142.7 | 116.4 | 32.471 | 42.725 | 21.818 | 0.6 | 11.9 | 22.3 | 4.3 | 17.4 | 6.5 | 1.1 | 2.3 | 23.7 | 6710.795 | 49.323 | 21 | 0.938546115 | Lower middle income |
| Guinea | 5.174 | | 90 | 94 | 88.9 | 15.556 | 22.503 | 8.842 | 0.9 | 14.5 | 38.4 | 1.8 | 2.6 | 1.4 | 1.3 | 3 | 3.2 | 1161.4787 | 34.856 | 15 | 0.79172525 | Low income |
| Guinea-Bissau | 5.116 | | 83.1 | 96 | 70 | 22.395 | 30.461 | 14.295 | 2.2 | 26 | 29.8 | 3.5 | 2.4 | 1.6 | 0.6 | 3.8 | 3.3 | 1340.2172 | 45.221 | 15 | 0.733617368 | Low income |
| Guyana | 2.679 | | 165.9 | 193.5 | 144.4 | 46.466 | 72.969 | 20.851 | 0.5 | 50.4 | 46.9 | 9.3 | 22.6 | 4.1 | 0 | 7.9 | 3.9 | 5430.0032 | 28.239 | 16 | 0.949433557 | Upper middle income |
| Haiti | 3.351 | | 106.9 | 111.5 | 102.9 | 29.720 | 42.055 | 17.373 | 1.2 | 22 | 24.9 | 6.8 | 3.2 | 7.1 | 0.1 | 3 | 7.7 | 1487.0111 | 52.016 | 17 | 0.870884327 | Low income |
| Honduras | 3.156 | | 131.2 | 146.7 | 116 | 36.924 | 51.660 | 22.305 | 1 | 19.9 | 29.4 | 6.9 | 14.8 | 6.4 | 0.7 | 2.5 | 17 | 4183.7238 | 51.696 | 21 | 0.954928722 | Lower middle income |
| Hungary | 1.277 | 2.73506387 | 285.4 | 236.5 | 356.1 | 100.964 | 106.925 | 95.192 | 14.1 | 54.5 | 18 | 42.3 | 7.5 | 51.6 | 7.1 | 10.6 | 9.5 | 21480.001 | 68.859 | 19 | 0.989011398 | High income |
| Iceland | 2.150 | | 284.3 | 274.2 | 299.5 | 71.976 | 102.447 | 42.739 | 11.5 | 96.3 | 7.9 | 28.4 | 11.8 | 29.8 | 12.1 | 6 | 5 | 38663.421 | 93.624 | 24 | 0.994339886 | High income |
| India | 2.565 | 4.40924943 | 94 | 97.4 | 92.4 | 30.835 | 41.162 | 21.307 | 1.6 | 25.8 | 22 | 6.1 | 2.3 | 6.9 | 0.2 | 4.9 | 6.1 | 4544.2929 | 30.93 | 17 | 0.898263498 | Lower middle income |
| Indonesia | 2.432 | | 133.5 | 134.4 | 136.2 | 44.899 | 58.558 | 31.407 | 3.2 | 40.3 | 17.3 | 12.8 | 5.6 | 16.3 | 0.5 | 8.4 | 2.8 | 7864.4155 | 49.924 | 17 | 0.945065241 | Lower middle income |
| Iran Islamic Repub | 1.903 | 3.46087562 | 127.7 | 120.1 | 134.7 | 39.578 | 47.429 | 31.956 | 8.3 | 28.1 | 2.8 | 11.1 | 2.5 | 7.7 | 0.8 | 4.8 | 15.2 | 15387.319 | 70.626 | 19 | 0.963201795 | Upper middle income |
| Iraq | 4.211 | | 135.3 | 131.7 | 144.6 | 30.885 | 38.674 | 23.392 | 11.4 | 42.6 | 2.8 | 7.1 | 1.4 | 14 | 0.3 | 4.3 | 5.3 | 12080.445 | 69.034 | 17 | 0.935835896 | Upper middle income |
| Ireland | 2.047 | 2.4675299 | 307.9 | 278.9 | 343.3 | 85.865 | 111.643 | 60.941 | 8.9 | 92.3 | 13.6 | 34.9 | 11.1 | 31.3 | 13.7 | 11.2 | 6.5 | 42904.849 | 61.84 | 23 | 0.98988536 | High income |
| Israel | 2.990 | 3.38411446 | 283.2 | 258.7 | 318 | 66.907 | 84.028 | 50.199 | 12.6 | 80.5 | 4.6 | 35.9 | 15.4 | 21.2 | 11.4 | 7.3 | 7.1 | 28588.83 | 91.824 | 24 | 0.991156199 | High income |
| Italy | 1.450 | 2.61320392 | 278.6 | 255.2 | 312.9 | 111.089 | 141.483 | 82.253 | 11.8 | 91.3 | 6.7 | 33.9 | 14 | 24.5 | 11.4 | 10.2 | 8.2 | 34719.963 | 68.327 | 24 | 0.992006133 | High income |
| Jamaica | 2.334 | | 198.5 | 179.2 | 222 | 52.991 | 73.362 | 32.341 | 3.3 | 55.8 | 26.3 | 14.4 | 12 | 18.2 | 0.9 | 6 | 9.1 | 8201.1989 | 53.743 | 21 | 0.956146048 | Upper middle income |
| Japan | 1.383 | 4.82303535 | 217.1 | 185.7 | 260.4 | 68.254 | 97.217 | 40.359 | 5.6 | 51.5 | 10.9 | 32.2 | 10.6 | 24.6 | 0.6 | 8.4 | 29.9 | 33916.472 | 90.522 | 26 | 0.993244417 | High income |
| Jordan | 3.457 | | 155.4 | 157.8 | 153.3 | 38.825 | 49.437 | 28.846 | 7.1 | 61 | 2.4 | 25.6 | 5.2 | 15.7 | 0.5 | 5.4 | 5.9 | 11028.5 | 82.473 | 19 | 0.961171052 | Lower middle income |
| Kazakhstan | 2.580 | | 236.5 | 216.7 | 282.2 | 75.212 | 99.458 | 51.066 | 6.7 | 63 | 29.4 | 22.8 | 12.9 | 27.9 | 3 | 9.7 | 21.6 | 19204.759 | 53.732 | 16 | 0.954558068 | Upper middle income |
| Kenya | 4.615 | | 181.8 | 196.6 | 167.2 | 41.046 | 48.641 | 33.546 | 1.7 | 38.3 | 40.1 | 8.6 | 6.6 | 2.6 | 1.2 | 6.4 | 9.5 | 2451.7771 | 23.571 | 17 | 0.871652279 | Lower middle income |
| Korea Democratic | 2.002 | | 181.2 | 170.8 | 204.2 | 64.089 | 75.337 | 53.232 | 4.6 | 36.8 | 12.4 | 21.8 | 5 | 44.2 | 0.2 | 6.8 | 14.3 | | 60.21 | 16 | 0.952413886 | Low income |
| Korea Republic of | 1.206 | 3.31220901 | 307.8 | 293.6 | 340 | 154.930 | 220.523 | 93.005 | 5.2 | 52.1 | 9.5 | 45 | 5.8 | 28.7 | 0.9 | 6.8 | 41.8 | 30422.952 | 81.936 | 23 | 0.988347461 | High income |
| Kuwait | 2.669 | | 102.1 | 123.3 | 89.8 | 23.350 | 32.780 | 16.836 | 5.5 | 46.7 | 4 | 12.8 | 7.5 | 8 | 0 | 4.7 | 2.6 | 76319.392 | 98.263 | 17 | 0.980383745 | High income |
| Kyrgyzstan | 3.020 | | 137.6 | 129.4 | 151.6 | 39.059 | 52.804 | 25.401 | 2.8 | 27.3 | 23.7 | 8.2 | 8.4 | 15.6 | 1.1 | 6.3 | 21.4 | 2733.7542 | 35.303 | 17 | 0.951036801 | Lower middle income |
| Lao PDR | 3.286 | | 141.8 | 122.4 | 165.5 | 34.924 | 39.893 | 30.027 | 2.1 | 19 | 12.5 | 8.8 | 3.4 | 13.2 | 0.5 | 5.2 | 2.3 | 3821.876 | 33.123 | 17 | 0.903514396 | Lower middle income |
| Latvia | 1.383 | 2.98778811 | 246.8 | 206.5 | 325 | 77.276 | 95.052 | 59.678 | 9.2 | 52.1 | 17.3 | 23.7 | 16.7 | 27.8 | 5.6 | 14.2 | 14.3 | 17592.15 | 67.692 | 19 | 0.985139438 | High income |
| Lebanon | 1.517 | | 197.4 | 192.8 | 203.9 | 65.304 | 83.585 | 46.772 | 16.6 | 78.7 | 4.6 | 16.1 | 7.7 | 19.8 | 1.1 | 7.5 | 5.5 | 15934.14 | 87.183 | 22 | 0.979697241 | Upper middle income |
| Lesotho | 3.208 | 4.4430915 | 103 | 96.7 | 114 | 36.142 | 40.299 | 32.088 | 1.2 | 9 | 38.4 | 2 | 2.5 | 2.8 | 0.5 | 1.7 | 1.1 | 2183.4685 | 24.753 | 15 | 0.78463986 | Lower middle income |
| Liberia | 5.025 | | 89.2 | 97 | 82.9 | 19.909 | 27.414 | 12.612 | 1.1 | 24.1 | 30.1 | 3.2 | 3.2 | 1.5 | 1 | 3.4 | 3.7 | 674.6061 | 47.801 | 15 | 0.818162224 | Low income |
| Libya | 2.525 | | 124.1 | 113.1 | 135.9 | 33.461 | 44.257 | 22.566 | 8.6 | 24.1 | 9.7 | 14.5 | 3.8 | 15.6 | 0.4 | 5 | 3.6 | 29649.271 | 77.642 | 19 | 0.966322628 | Upper middle income |
| Lithuania | 1.517 | 2.7312225 | 251.9 | 224 | 311.8 | 85.800 | 115.614 | 56.225 | 8.7 | 48.7 | 26.1 | 23.4 | 17.7 | 26.2 | 5.2 | 12.2 | 13.8 | 19843.44 | 66.757 | 19 | 0.987817471 | High income |
| Luxembourg | 1.580 | 2.55256541 | 280.3 | 259.6 | 309.1 | 88.710 | 112.102 | 66.106 | 8.7 | 89.1 | 4.9 | 31.5 | 24.2 | 28.4 | 11.4 | 7.3 | 7.6 | 84210.015 | 88.547 | 23 | 0.992357704 | High income |
| Madagascar | 4.655 | | 137.5 | 134.3 | 142.4 | 37.494 | 47.056 | 27.942 | 3.2 | 26.6 | 44.6 | 8 | 2.9 | 7.2 | 1.3 | 2.2 | 4.7 | 1362.3112 | 31.929 | 17 | 0.911507927 | Low income |
| Malawi | 5.634 | | 156 | 186.4 | 123.5 | 61.693 | 75.804 | 47.896 | 7 | 16.8 | 75.9 | 3.4 | 2.2 | 0.9 | 1.5 | 2.5 | 2.7 | 722.38784 | 15.54 | 17 | 0.777079695 | Low income |
| Malaysia | 2.004 | | 143.6 | 143.4 | 144.9 | 45.486 | 57.071 | 34.291 | 3.7 | 38.7 | 15.6 | 18.3 | 5.3 | 17.9 | 0.5 | 7.8 | 7.8 | 19985.916 | 70.912 | 19 | 0.984797747 | Upper middle income |
| Maldives | 2.338 | | 88.9 | 84.8 | 91.6 | 19.789 | 28.138 | 11.484 | 2 | 31.6 | 11 | 5.8 | 3.1 | 7.7 | 0 | 7.1 | 3.7 | 10465.204 | 39.984 | 20 | 0.979947754 | Upper middle income |
| Mali | 6.840 | | 111.4 | 135.6 | 83.8 | 32.862 | 44.087 | 21.896 | 6.7 | 29.8 | 44.2 | 6 | 3.5 | 2.7 | 0.9 | 5.2 | 9.2 | 1630.0012 | 35.996 | 15 | 0.774009131 | Low income |
| Malta | 1.410 | 2.71689491 | 242.9 | 228.9 | 267.7 | 73.077 | 95.833 | 51.343 | 15.8 | 85.9 | 3.8 | 31.9 | 15.3 | 20.4 | 6.2 | 11.8 | 8 | 26671.467 | 94.665 | 22 | 0.990072799 | High income |
| Mauritania | 4.836 | | 85.7 | 97.7 | 74.4 | 25.517 | 34.527 | 16.767 | 2.2 | 25.8 | 29.4 | 3.6 | 2.7 | 1.4 | 0.8 | 3.9 | 3.5 | 2619.8656 | 56.682 | 16 | 0.828872462 | Lower middle income |
| Mauritius | 1.570 | | 180.2 | 193.9 | 171.1 | 63.403 | 82.849 | 44.164 | 4.3 | 64.2 | 15 | 18.6 | 11.5 | 9.9 | 0.3 | 8.3 | 8 | 14917.42 | 40.579 | 19 | 0.977322443 | Upper middle income |
| Mexico | 2.281 | 3.4076092 | 131.5 | 139.9 | 123.9 | 48.260 | 63.430 | 32.948 | 2.9 | 35.4 | 23.3 | 7.8 | 4.9 | 7.5 | 1.8 | 5.6 | 6.9 | 14726.446 | 77.825 | 22 | 0.976184901 | Upper middle income |
| Micronesia | 3.461 | | 171.4 | 146.3 | 202.1 | 46.506 | 59.655 | 33.631 | 2.9 | 48.8 | 8.7 | 20.4 | 12.3 | 34.8 | 3.1 | 3.1 | 1.7 | 3269.6711 | 22.298 | 17 | | Lower middle income |
| Mongolia | 2.431 | | 200.9 | 171.9 | 237.7 | 40.423 | 41.639 | 39.209 | 1.1 | 9.4 | 24.3 | 6 | 1.9 | 15.6 | 0.1 | 3.7 | 32.5 | 6344.5119 | 67.567 | 16 | 0.964403327 | Lower middle income |
| Montenegro | 1.698 | | 238.3 | 219.7 | 262.7 | 112.392 | 131.442 | 93.668 | 10.1 | 59.7 | 20.2 | 28.2 | 15.3 | 39.6 | 4.7 | 12 | 9.5 | 13325.031 | 63.096 | 19 | 0.987371013 | Upper middle income |
| Morocco | 2.581 | | 117.8 | 114.4 | 122.7 | 40.690 | 54.102 | 26.776 | 5.8 | 40.8 | 14.3 | 8.5 | 3 | 13.6 | 0.4 | 4.7 | 4 | 6334.9581 | 57.684 | 18 | 0.954920903 | Lower middle income |
| Mozambique | 5.408 | | 136.8 | 153 | 118.3 | 55.340 | 68.501 | 41.798 | 1.5 | 14.5 | 65 | 1.2 | 1.9 | 2.8 | 1.3 | 0.9 | 0.9 | 913.89597 | 30.955 | 16 | 0.752170541 | Low income |
| Myanmar | 2.003 | | 140.5 | 134.6 | 149.4 | 48.005 | 58.192 | 37.732 | 2.1 | 22.1 | 20.6 | 8.7 | 2.4 | 20.2 | 0.3 | 5.5 | 11.2 | | 31.405 | 16 | 0.901564516 | Lower middle income |
| Namibia | 3.230 | | 82.7 | 81.5 | 86.3 | 25.466 | 31.340 | 19.704 | 2.1 | 24.4 | 14.7 | 4.8 | 2.6 | 3.3 | 2.1 | 2.8 | 1.6 | 8267.3391 | 41.616 | 17 | 0.881203108 | Upper middle income |
| Nepal | 2.623 | | 85.2 | 85.6 | 85.6 | 24.848 | 31.708 | 18.047 | 2 | 13.7 | 19 | 3.2 | 0.9 | 12.3 | 0.2 | 5.8 | 5.3 | 1958.6623 | 16.822 | 17 | 0.931676131 | Low income |
| New Caledonia | 2.173 | | 297.9 | 269.3 | 330.7 | 78.178 | 112.809 | 43.885 | 4.5 | 87.6 | 15.3 | 24.1 | 22.1 | 40.1 | 7.2 | 7.6 | 7.8 | | 67.273 | 19.700 | | High income |
| New Zealand | 2.110 | 2.48849734 | 295 | 274.3 | 320.1 | 90.008 | 115.955 | 63.939 | 2.9 | 85 | 5.3 | 37.3 | 13.9 | 25.9 | 35.8 | 8 | 5.2 | 30336.916 | 86.165 | 24 | 0.987538394 | High income |
| Nicaragua | 2.630 | | 114.4 | 123.1 | 106.1 | 33.898 | 47.608 | 20.087 | 0.9 | 23.9 | 36.2 | 7.9 | 2.8 | 7 | 0.5 | 2.1 | 11.1 | 3962.7333 | 57.255 | 21 | 0.964828261 | Lower middle income |
| Niger | 7.583 | | 63.4 | 71 | 56.7 | 20.738 | 26.540 | 14.966 | 1.9 | 23.8 | 8.6 | 4.8 | 3.1 | 0.2 | 0.6 | 7.3 | 1.9 | 824.20535 | 17.559 | 15 | 0.796596007 | Low income |
| Nigeria | 6.020 | | 100.1 | 121.7 | 79 | 29.975 | 42.991 | 17.465 | 1 | 50.4 | 29 | 4.2 | 3.4 | 1.1 | 0.5 | 3.1 | 2 | 5010.3365 | 43.48 | 13 | 0.782234559 | Lower middle income |
| Norway | 1.937 | 2.3047853 | 318.3 | 277.1 | 368.7 | 88.025 | 105.949 | 70.980 | 13.5 | 73.1 | 9.8 | 38.9 | 16.9 | 30 | 18.8 | 9.5 | 4.6 | 57739.041 | 79.102 | 23 | 0.991312752 | High income |
| Oman | 2.898 | | 82.1 | 92.4 | 78.6 | 22.046 | 27.644 | 18.078 | 4.8 | 26 | 5.3 | 7.4 | 3.8 | 5.1 | 0.4 | 3.3 | 5.3 | 48497.048 | 75.161 | 20 | 0.991312752 | High income |
| Pakistan | 3.429 | | 111.8 | 127.7 | 96 | 40.326 | 51.609 | 29.394 | 3.4 | 50.3 | 7.9 | 4 | 3.6 | 5.8 | 0.2 | 5.6 | 3 | 4134.2353 | 36.598 | 17 | 0.876554278 | Lower middle income |
| Panama | 2.549 | | 148.4 | 148.8 | 150.1 | 47.011 | 66.396 | 28.207 | 2.3 | 43 | 18.7 | 12.5 | 7.7 | 9 | 1.4 | 6.4 | 10.6 | 14619.586 | 65.115 | 23 | 0.967054161 | Upper middle income |
| Papua New Guinea | 3.954 | | 165.2 | 179.8 | 156.7 | 48.203 | 71.502 | 26.241 | 1.5 | 33.7 | 34.5 | 8.1 | 9.3 | 8 | 4.2 | 7 | 6.9 | 2071.4873 | 13.019 | 15 | 0.892398564 | Lower middle income |
| Paraguay | 2.970 | | 147.5 | 153 | 143.2 | 43.623 | 64.310 | 23.402 | 2.2 | 43.8 | 34.2 | 12.1 | 5.8 | 14 | 1.5 | 5.1 | 6.3 | 6865.8846 | 58.487 | 21 | 0.966617037 | Upper middle income |
| Peru | 2.512 | 4.17362513 | 154.5 | 169.8 | 140.9 | 46.689 | 65.242 | 28.564 | 3.1 | 28 | 32.7 | 11.1 | 3.7 | 10.1 | 1.6 | 4.9 | 15.8 | 9714.6375 | 76.915 | 21 | 0.967407676 | Upper middle income |
| Philippines | 3.154 | 5.13756125 | 140 | 143.4 | 139.9 | 42.919 | 58.209 | 28.070 | 1.5 | 47 | 16 | 13.1 | 5.6 | 19.3 | 0.4 | 5.9 | 3.8 | 5500.2372 | 45.255 | 17 | 0.952944127 | Lower middle income |

| | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Poland | 1.360 | 3.24831429 | 229.6 | 205.6 | 269.2 | 62.543 | 74.630 | 50.808 | 11 | 51.9 | 12.2 | 27 | 16.9 | 38 | 4.1 | 13.6 | 8.4 | 20683.139 | 60.892 | 20 | 0.988809463 | High income |
| Portugal | 1.360 | 2.89386299 | 246.2 | 198.1 | 306.3 | 101.877 | 110.554 | 93.380 | 12 | 67.6 | 9 | 31.7 | 12.6 | 20.2 | 6.7 | 6.2 | 13.1 | 26927.431 | 60.567 | 22 | 0.990743134 | High income |
| Puerto Rico | 1.657 | 3.01953105 | 211.1 | 193.5 | 237.2 | 72.707 | 98.906 | 46.307 | 5.5 | 57.5 | 11.4 | 24.6 | 16 | 9.2 | 2.1 | 4.6 | 4.1 | 33759.999 | 93.825 | 23 | | High income |
| Qatar | 2.091 | 9.18350419 | 108.8 | 134.5 | 104 | 28.719 | 50.248 | 21.838 | 5.3 | 46.1 | 5.1 | 12.6 | 5.7 | 10.7 | 0.5 | 4.6 | 5.8 | 126613.79 | 98.655 | 21 | 0.985182321 | High income |
| Republic of Moldov | 1.476 | | 194.1 | 170.2 | 230 | 64.194 | 76.157 | 52.032 | 6.3 | 38.7 | 19.6 | 28.3 | 12.8 | 23.5 | 2.3 | 7.5 | 11.5 | 3831.8627 | 44.886 | 16 | 0.979509629 | Lower middle income |
| Romania | 1.523 | 3.02381665 | 224.2 | 190.6 | 271 | 82.738 | 95.967 | 69.951 | 9.7 | 50 | 28.6 | 26.4 | 8.5 | 32.6 | 3.5 | 10.3 | 10.4 | 16252.231 | 53.829 | 19 | 0.98265702 | Upper middle income |
| Russian Federation | 1.563 | 2.75399249 | 204.3 | 187.1 | 245.8 | 69.573 | 90.248 | 48.848 | 5.7 | 45.6 | 15.3 | 24.5 | 16.1 | 24 | 4.1 | 11.3 | 16 | 20541.334 | 73.687 | 17 | 0.976110152 | Upper middle income |
| Rwanda | 4.841 | 2.97399284 | 135.8 | 142.3 | 130.2 | 32.732 | 41.506 | 23.737 | 3.5 | 15.9 | 41.8 | 5.1 | 5.3 | 1.2 | 2.4 | 4.2 | 8.2 | 1236.476 | 23.952 | 17 | 0.839898749 | Low income |
| Samoa | 4.335 | | 92.7 | 96.1 | 92.5 | 29.915 | 45.359 | 15.824 | 1.3 | 23.2 | 17.1 | 6.5 | 7.4 | 4.7 | 0.9 | 3.9 | 9.7 | 5307.7036 | 20.078 | 18 | 0.964750824 | Upper middle income |
| Saudi Arabia | 2.830 | | 91.1 | 102.8 | 85.9 | 28.025 | 38.452 | 19.434 | 3.6 | 29.5 | 2.7 | 11.6 | 5.8 | 5.1 | 0.3 | 3.4 | 3.1 | 45247.385 | 82.084 | 18 | 0.965797508 | High income |
| Senegal | 5.049 | | 101.2 | 115 | 85.5 | 25.216 | 35.331 | 15.068 | 3.9 | 22.4 | 41.4 | 3.9 | 3 | 2.1 | 1.1 | 4.2 | 6.2 | 2137.5803 | 42.23 | 16 | 0.870409233 | Low income |
| Serbia | 1.413 | | 269.7 | 247.6 | 299.2 | 99.341 | 118.559 | 80.972 | 10.6 | 69 | 23.8 | 32.6 | 17.9 | 45.6 | 7.1 | 12.8 | 8.6 | 11805.284 | 55.208 | 18 | 0.988676774 | Upper middle income |
| Sierra Leone | 4.945 | | 92.3 | 97.7 | 83.8 | 21.286 | 29.100 | 13.165 | 1 | 24.3 | 30.2 | 3.4 | 3.2 | 1.5 | 0.9 | 3.3 | 3.8 | 1319.2568 | 38.241 | 12 | 0.727886294 | Low income |
| Singapore | 1.190 | 3.57709406 | 206.4 | 198.7 | 218.8 | 74.461 | 102.602 | 47.684 | 4.3 | 65.7 | 8.1 | 33.7 | 13.9 | 24.9 | 0.5 | 9.9 | 8.2 | 70364.208 | 100 | 24 | 0.993723047 | High income |
| Slovakia | 1.440 | 3.1886377 | 276.9 | 238 | 338.2 | 83.123 | 93.386 | 73.235 | 10.1 | 57.5 | 16.1 | 42.7 | 19 | 28.3 | 9.9 | 11.6 | 9.6 | 24434.94 | 54.685 | 19 | 0.98890098 | High income |
| Slovenia | 1.553 | 2.9476531 | 296.3 | 251.5 | 358.2 | 105.499 | 125.201 | 86.730 | 10.7 | 66.5 | 10.5 | 37 | 15 | 33.9 | 16.2 | 10.4 | 10.4 | 27566.802 | 50.04 | 22 | 0.98958832 | High income |
| Solomon Islands | 4.235 | | 116.3 | 145.2 | 89.3 | 44.369 | 66.601 | 23.875 | 1.5 | 47.6 | 28.5 | 6.9 | 10.2 | 7.6 | 0.6 | 10.1 | 2 | 1750.3006 | 20.048 | 17 | 0.946529387 | Lower middle income |
| Somalia | 6.868 | | 139.1 | 165.2 | 111.9 | 36.113 | 48.428 | 23.736 | 2.1 | 40.6 | 33.4 | 8 | 4.3 | 2.9 | 0.7 | 7.5 | 6.3 | | 37.259 | 16 | 0.726995383 | Low income |
| South African Repu | 2.468 | | 187.1 | 168.9 | 224.3 | 51.815 | 68.617 | 35.533 | 4.4 | 41.5 | 31.7 | 11.9 | 6.9 | 18.5 | 4.5 | 5.5 | 5.1 | 11415.27 | 62.218 | 15 | 0.81134756 | Upper middle income |
| South Sudan | 5.194 | | 132.7 | 143 | 123.1 | 35.665 | 43.847 | 27.578 | 1.9 | 31.8 | 30.4 | 6.6 | 3.8 | 2.3 | 1.2 | 6.4 | 5 | 3760.456 | 17.855 | 16 | | Low income |
| Spain | 1.363 | 2.87917702 | 249 | 198.2 | 312.8 | 88.128 | 109.409 | 67.965 | 13.9 | 67.3 | 7.8 | 33.1 | 11.6 | 30.3 | 6.9 | 7.7 | 7.8 | 32354.127 | 78.442 | 24 | 0.991649103 | High income |
| Sri Lanka | 2.342 | | 94.8 | 102.7 | 86.9 | 32.586 | 41.798 | 23.410 | 1.8 | 30.9 | 13.1 | 3.7 | 1.5 | 6.2 | 0.1 | 5.8 | 5.8 | 7418.9043 | 18.321 | 19 | 0.968771239 | Lower middle income |
| Sudan | 4.638 | | 91.1 | 91 | 92 | 24.468 | 29.736 | 19.286 | 2.3 | 27.8 | 7.9 | 4.6 | 2.4 | 2 | 0.7 | 6.4 | 1.8 | 3259.4073 | 33.08 | 17 | 0.820980152 | Lower middle income |
| Suriname | 2.346 | | 159.6 | 162.7 | 163.8 | 53.954 | 77.399 | 31.369 | 3.1 | 41.4 | 38 | 16.7 | 4 | 12.8 | 0.5 | 8.6 | 5 | 14034.923 | 66.344 | 18 | 0.956288068 | Upper middle income |
| Swaziland | 3.558 | 4.79959 | 115.3 | 111.9 | 122.1 | 39.740 | 47.239 | 32.238 | 1.9 | 10.5 | 53.1 | 2.3 | 5.3 | 3 | 0.6 | 2.5 | 2.1 | 6379.9884 | 21.492 | 16 | 0.80146329 | Lower middle income |
| Sweden | 1.940 | | 270 | 248.7 | 296.8 | 79.343 | 105.049 | 54.799 | 10.3 | 80.4 | 7.4 | 29.2 | 13.5 | 19.1 | 18 | 7.5 | 3.7 | 41731.84 | 85.056 | 24 | 0.992422252 | High income |
| Switzerland | 1.513 | 2.39680956 | 287 | 245.9 | 337.9 | 96.142 | 117.225 | 75.586 | 12.2 | 83.1 | 3.6 | 29.4 | 12.6 | 27.3 | 20.3 | 7.9 | 4.2 | 51327.255 | 73.663 | 25 | 0.992895335 | High income |
| Syrian Arab Republ | 3.080 | | 145.9 | 145.2 | 148.3 | 37.806 | 48.734 | 27.229 | 9.6 | 52.5 | 2.6 | 16.2 | 3.3 | 15.1 | 0.4 | 4.8 | 5.6 | | 55.677 | 20 | 0.975967947 | Lower middle income |
| Tajikistan | 3.777 | | 119.1 | 112.3 | 128.7 | 35.870 | 41.201 | 30.451 | 3 | 20.4 | 9.9 | 5.5 | 12.2 | 7.7 | 1 | 2 | 21.7 | 2067.6194 | 26.516 | 18 | 0.937648873 | Lower middle income |
| Tanzania | 5.426 | | 123.7 | 132.7 | 115.8 | 33.776 | 39.486 | 28.148 | 3.2 | 19.4 | 54 | 4.8 | 1.5 | 0.7 | 1.6 | 2 | 3.1 | 1513.96 | 28.114 | 17 | 0.829695397 | Low income |
| Thailand | 1.443 | | 137.5 | 128.8 | 149.6 | 57.275 | 70.517 | 44.083 | 2.7 | 29.3 | 17.8 | 12.4 | 3.9 | 20.9 | 0.4 | 5.9 | 3.1 | 12562.433 | 44.08 | 21 | 0.970930809 | Upper middle income |
| The Gambia | 5.794 | | 68.2 | 69.6 | 67.3 | 15.846 | 14.789 | 16.936 | 0.8 | 9.8 | 26.1 | 1.3 | 1.1 | 2.4 | 0.9 | 1.2 | 1.7 | 1633.3732 | 56.297 | 15 | 0.848684102 | Low income |
| The Netherlands | 1.780 | | 304.8 | 289.6 | 327.8 | 103.206 | 136.038 | 71.460 | 8.2 | 99 | 6.8 | 40.2 | 12.4 | 37.2 | 19.4 | 6.8 | 5.6 | 44747.976 | 87.061 | 23 | 0.991178479 | High income |
| Timor-Leste | 5.600 | 5.50669848 | 165.4 | 149.6 | 183.9 | 27.088 | 32.684 | 21.764 | 3.5 | 32.6 | 13.3 | 13.4 | 9.3 | 31 | 1.7 | 5.1 | 2.3 | 1741.1962 | 29.507 | 16 | 0.924547722 | Lower middle income |
| Togo | 4.791 | | 91.1 | 104.8 | 77.2 | 28.816 | 38.094 | 19.476 | 1.8 | 27.2 | 21.5 | 3.8 | 4.1 | 1.4 | 1 | 5 | 5.8 | 1220.822 | 37.533 | 14 | 0.853879923 | Low income |
| Trinidad and Tobag | 1.801 | | 210.9 | 180.3 | 273.5 | 60.151 | 82.396 | 37.776 | 3.3 | 56.9 | 24.5 | 23.5 | 14.6 | 12.2 | 1.3 | 10.6 | 4.4 | 28728.157 | 9.092 | 18 | 0.956069163 | High income |
| Tunisia | 2.110 | | 110.6 | 95.7 | 127 | 37.733 | 46.114 | 29.442 | 8.3 | 31.8 | 4.8 | 10.9 | 3.3 | 16 | 0.5 | 4.2 | 4.2 | 10198.464 | 65.934 | 19 | 0.972061666 | Lower middle income |
| Turkey | 2.101 | | 205.1 | 161.6 | 257.8 | 65.428 | 75.226 | 55.813 | 15.2 | 39.1 | 4.3 | 16.6 | 10.1 | 34.7 | 2.1 | 6.3 | 14.2 | 16195.185 | 70.715 | 20 | 0.974579606 | Upper middle income |
| Turkmenistan | 2.413 | | 144 | 132.8 | 159.4 | 52.557 | 59.904 | 45.214 | 3.5 | 26.8 | 13.1 | 9 | 6.1 | 12.7 | 1.2 | 2.6 | 18.2 | 9828.8076 | 48.402 | 17 | 0.936137764 | Upper middle income |
| Uganda | 6.154 | | 169.7 | 167.4 | 175.7 | 40.141 | 44.766 | 35.572 | 0.9 | 27.5 | 44.4 | 7.1 | 4 | 2.7 | 1.8 | 6.9 | 5.1 | 1267.8395 | 14.492 | 17 | 0.817073102 | Low income |
| Ukraine | 1.455 | | 192.9 | 174.7 | 231.9 | 76.978 | 98.174 | 55.728 | 6 | 41.3 | 16.6 | 23.4 | 16.6 | 22.2 | 4 | 10.7 | 14.3 | 7697.9839 | 68.686 | 17 | 0.977451168 | Lower middle income |
| United Arab Emirat | 1.871 | | 92.5 | 127.1 | 83.8 | 19.951 | 38.228 | 11.605 | 4.3 | 39.2 | 9.5 | 8.5 | 6 | 9.4 | 0.2 | 6.4 | 4.8 | 55764.873 | 84.055 | 19 | 0.986916907 | High income |
| United Kingdom | 1.907 | | 272.9 | 267.3 | 284 | 85.956 | 113.088 | 59.511 | 5.8 | 95 | 7.1 | 30.2 | 13.9 | 30 | 14.6 | 11.7 | 4.7 | 35924.014 | 81.302 | 23 | 0.989882467 | High income |
| United States of Ar | 1.943 | | 318 | 297.4 | 347 | 101.350 | 126.577 | 77.022 | 11.6 | 92.9 | 6.6 | 25 | 19.5 | 38.4 | 14.3 | 8 | 3.9 | 48377.394 | 80.772 | 23 | 0.985353656 | High income |
| Uruguay | 2.080 | | 251 | 220.9 | 297.5 | 72.147 | 91.718 | 52.822 | 8.5 | 69.8 | 18.9 | 29.5 | 9.5 | 27.4 | 4.1 | 8.4 | 10 | 16160.788 | 94.414 | 21 | 0.979863168 | High income |
| Uzbekistan | 2.390 | | 99.7 | 103.5 | 96.9 | 37.112 | 47.766 | 26.596 | 2.3 | 27.1 | 13.5 | 5.3 | 5.8 | 8 | 0.8 | 2.1 | 12.5 | 4100.5201 | 36.191 | 18 | 0.9542107 | Lower middle income |
| Vanuatu | 3.501 | | 107.8 | 117 | 98.2 | 39.690 | 53.397 | 26.525 | 1.4 | 31.8 | 19.2 | 6.5 | 10.2 | 9.6 | 1.1 | 4.2 | 3.3 | 2888.8251 | 24.589 | 17 | 0.962863745 | Lower middle income |
| Venezuela | 2.472 | | 150 | 155 | 146.9 | 51.465 | 72.566 | 30.855 | 3.7 | 41.2 | 32.8 | 10.7 | 5.4 | 16 | 1.1 | 5.1 | 9.6 | 16202.142 | 88.769 | 21 | 0.973793981 | Upper middle income |
| Vietnam | 1.819 | | 140.4 | 114.2 | 172.9 | 62.740 | 60.579 | 64.867 | 1.1 | 23 | 10.6 | 10.1 | 5.4 | 25.2 | 0.2 | 2.6 | 16.3 | 4395.5501 | 30.392 | 22 | 0.965838559 | Lower middle income |
| Yemen, Rep. | 4.501 | | 80.4 | 80.7 | 81.2 | 24.521 | 28.304 | 20.838 | 1.8 | 27.4 | 3.1 | 4.5 | 0.1 | 3.8 | 0.3 | 3.8 | 4 | 4442.5431 | 31.732 | 16 | 0.902273096 | Lower middle income |
| Zambia | 5.812 | | 136.2 | 157.8 | 115.1 | 41.861 | 51.846 | 32.093 | 2.8 | 22.4 | 58 | 4.8 | 3.3 | 1.8 | 0.9 | 4 | 4.4 | 3381.0597 | 38.725 | 16 | 0.761352332 | Lower middle income |
| Zimbabwe | 3.719 | | 190.3 | 209.1 | 167 | 44.841 | 48.064 | 41.644 | 2.9 | 28.5 | 56.4 | 8.8 | 9.1 | 4.9 | 1.6 | 6.6 | 8 | 1454.2285 | 33.196 | 18 | 0.8087404 | Low income |

* Age-standardised rates per 100,000. This rate is same for all the other cancer rates.

Supplementary File 2:

Figure S1. The relationship between household size and all cancers incidence rates (total, male and female, all ages)



Supplementary File 3:

Table S2: Pearson, Nonparametric and partial correlation between household size and each cancer variable and confounder

| | All countries, n=58 | | | | | |
|---|---|---|---|---|---|---|
| | Pearson | | Nonparametric | | Partial | |
| | r | n | rho | n | r | Df |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) - all ages: total | -0.635*** | 58 | -0.720*** | 58 | -0.492*** | 50 |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44)- all ages: female | -0.624*** | 58 | -0.720*** | 58 | -0.482*** | 50 |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) - all ages: male | -0.593*** | 58 | -0.671*** | 58 | -0.429* | 50 |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) – 0-49: total‡ | -0.629*** | 58 | -0.711*** | 58 | -0.493*** | 50 |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) – 0-49: female‡ | -0.586*** | 58 | -0.700*** | 58 | -0.422** | 50 |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) – 0-49: male‡ | -0.617*** | 58 | -0.673*** | 58 | -0.486*** | 50 |
| Bladder (C67), all ages | -0.482*** | 58 | -0.593*** | 58 | -0.303* | 50 |
| Breast(C50), all ages | -0.489*** | 58 | -0.672*** | 58 | -0.305* | 50 |
| Cervix uteri (C53), all ages | 0.260* | 58 | 0.331* | 58 | 0.010 | 50 |
| Colorectum (C18-21), all ages | -0.508*** | 58 | -0.583*** | 58 | -0.278* | 50 |
| Corpus uteri (C54), all ages | -0.493*** | 58 | -0.499*** | 58 | -0.324* | 50 |
| Lung (C33-34), all ages | -0.380** | 58 | -0.425*** | 58 | -0.059 | 50 |
| Melanoma of skin (C43), all ages | -0.646*** | 58 | -0.778*** | 58 | -0.564*** | 50 |
| Ovary (C56), all ages | -0.374** | 58 | -0.456*** | 58 | -0.239 | 50 |
| Stomach (C16), all ages | -0.104 | 58 | 0.002 | 58 | 0.133 | 50 |
| GDP PPP 2010 | -0.463*** | 58 | -0.505*** | 58 | ^ | ^ |
| Urbanization 2010 | -0.388** | 57 | -0.578*** | 57 | ^ | ^ |
| Life expectancy ($e_{60}$, 2005-2010) | -0.346** | 58 | -0.269* | 58 | ^ | ^ |
| Biological State Index ($I_{bs}$) | -0.400** | 56 | -0.581*** | 56 | ^ | ^ |

Note:

Pearson, Nonparametric and partial correlation reported. Significance level: *** p<0.001, ** p<0.01, * p<0.05

^ Partial correlations were calculated when GDP, Urbanization, Life expectancy (e60) and Biological State Index ($I_{bs}$) were kept statistically constant.

‡ Life expectancy ($e_{50}$) was not controlled for as it is not relevant in population segment aged 0-49 years old.

Data sources and variable meanings:

The International Agency for Research published cancer incidence rates (per 100,000 in 2012) of all cancers incidence rate by sex (total, male and female, 0-49 years and all ages respectively); bladder, breast, cervix uteri, colorectum, corpus uteri, ovary and stomach.

The World Bank data: GDP PPP (per capita purchasing power parity in current international $ in 2010) and Urbanization (the percentage of total population living urban areas in 2010)

The United Nations data: Life expectancy ($e_{60}$, 2005-2010), the total population in households and the number of households for calculating household size. Household size is expressed as total number of persons in a household.

United Nations published (2008) country specific fertility data and WHO published (2012) life table were used for calculating the Biological State Index ($I_{bs}$).

All variables were log-transformed for analysis in SPSS.

Supplementary File 4:

Table S3: Stepwise multiple linear regression to identify the significant predictors of cancer incidence risk

| | All countries, n=58 | | | |
|---|---|---|---|---|
| | Rank | Predictor | Beta | Adjusted $R^2$ |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) - all ages: both sexes | 1 | Life Expectancy | 0.302** | 0.471 |
| | 2 | Household Size | -0.362*** | 0.591 |
| | 3 | Biological State Index ($I_{bs}$) | 0.331** | 0.645 |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44)- all ages: female | 1 | Life Expectancy | 0.351** | 0.479 |
| | 2 | Household Size | -0.353*** | 0.589 |
| | 3 | Biological State Index ($I_{bs}$) | 0.273* | 0.624 |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) - all ages: male | 1 | Biological State Index ($I_{bs}$) | 0.317* | 0.470 |
| | 2 | Household Size | -0.343*** | 0.583 |
| | 3 | GDP PPP | 0.319* | 0.622 |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) – 0-49: both sexes‡ | 1 | GDP PPP | 0.512*** | 0.450 |
| | 2 | Household Size | -0.431*** | 0.604 |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) – 0-49: female‡ | 1 | GDP PPP | 0.576*** | 0.503 |
| | 2 | Household Size | -0.363*** | 0.601 |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) - all ages: male‡ | 1 | GDP PPP | 0.504*** | 0.435 |
| | 2 | Household Size | -0.423*** | 0.582 |

Note:

Significance level: *** p<0.001, ** p<0.01, * p<0.05

Variables (log-transformed) entered for multiple linear regression (stepwise) analysis: Household Size, Life Expectancy ($e_{60}$), GDP PPP, Urbanization and Biological State Index ($I_{bs}$).

‡ Life expectancy ($e_{50}$) was not included as it is not relevant in population segment aged 0-49 years.

Data sources and variable meanings:

The International Agency for Research published cancer incidence rates (per 100,000 in 2012) of all cancers incidence rate by sex (total, male and female, 0-49 years and all ages respectively); bladder, breast, cervix uteri, colorectum, corpus uteri, ovary and stomach.

The World Bank data: Total Fertility Rate (the mean number of children born to a woman between 2009-2011), GDP PPP (per capita purchasing power parity in current international $ in 2010) and Urbanization (the percentage of total population living urban areas in 2010)

The United Nations data: Life expectancy ($e_{60}$, 2005-2010), the total population in households and the number of households for calculating household size. Household size is expressed as total number of persons in a household.

United Nations published (2008) country specific fertility data and WHO published (2012) life table were used for calculating the Biological State Index ($I_{bs}$).

Supplementary File 5:

Table S4 Data descriptive and summary

**Descriptive Statistics**

| | All countries | | | | | | | Developed Countries | | | | | | | Developing Countries | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | N | Minimum | Maximum | Mean | Std. | Skewness | Std. | N | Minimum | Maximum | Mean | Std. | Skewness | Std. | N | Minimum | Maximum | Mean | Std. | Skewness | Std. | Data |
| | Statistic | Statistic | Statistic | Statistic | Deviation | Statistic | Error | Statistic | Statistic | Statistic | Statist c | Deviation | Statistic | Error | Statistic | Statistic | Statistic | Statistic | Deviation | Statistic | Error | Source |
| Total Fertility Rate (Mean 2009-2011) | 178 | 1.19 | 7.58 | 2.97 | 1.52 | 0.97 | 0.18 | 98 | 1.19 | 5.14 | 2.05 | 0.70 | 1.83 | 0.24 | 80 | 1.45 | 7.58 | 4.09 | 1.49 | 0.12 | 0.27 | The World Bank |
| Household Size | 58 | 2.02 | 9.18 | 3.58 | 1.35 | 2.25 | 0.31 | 46 | 2.02 | 9.18 | 3.37 | 1.38 | 2.94 | 0.35 | 12 | 2.97 | 5.94 | 4.35 | 0.97 | -0.01 | 0.64 | The United Nations |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) - all ages: total | 178 | 63.40 | 338.10 | 168.84 | 70.47 | 0.61 | 0.18 | 98 | 82.10 | 338.10 | 206.35 | 69.52 | -0.06 | 0.24 | 80 | 63.40 | 257.00 | 122.89 | 36.14 | 1.00 | 0.27 | The IARC of WHO |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44)- all ages: female | 178 | 69.60 | 328.80 | 163.57 | 57.93 | 0.63 | 0.18 | 98 | 81.50 | 328.80 | 192.96 | 57.70 | 0.08 | 0.24 | 80 | 69.60 | 226.40 | 127.57 | 32.36 | 0.64 | 0.27 | The IARC of WHO |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) - all ages: male | 178 | 56.70 | 373.90 | 180.03 | 90.31 | 0.57 | 0.18 | 98 | 76.10 | 373.90 | 228.88 | 87.79 | -0.15 | 0.24 | 80 | 56.70 | 305.60 | 120.20 | 47.18 | 1.40 | 0.27 | The IARC of WHO |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) – 0-49: total‡ | 178 | 15.56 | 154.93 | 53.15 | 25.95 | 0.84 | 0.18 | 98 | 19.79 | 154.93 | 66.69 | 25.86 | 0.32 | 0.24 | 80 | 15.56 | 84.66 | 36.56 | 13.49 | 1.30 | 0.27 | The IARC of WHO |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) – 0-49: female‡ | 178 | 14.79 | 220.52 | 68.42 | 32.97 | 0.95 | 0.18 | 98 | 27.64 | 220.52 | 86.24 | 32.26 | 0.58 | 0.24 | 80 | 14.79 | 107.69 | 46.58 | 16.78 | 1.29 | 0.27 | The IARC of WHO |
| All cancers excl. non-melanoma skin cancer (C00-97, but C44) – 0-49: male‡ | 178 | 8.84 | 95.19 | 38.43 | 20.47 | 0.91 | 0.18 | 98 | 11.48 | 95.19 | 48.04 | 21.23 | 0.39 | 0.24 | 80 | 8.84 | 64.87 | 26.65 | 11.43 | 1.32 | 0.27 | The IARC of WHO |
| Bladder (C67), all ages | 178 | 0.50 | 17.50 | 5.13 | 4.09 | 1.01 | 0.18 | 98 | 0.50 | 17.50 | 7.06 | 4.18 | 0.47 | 0.24 | 80 | 0.50 | 13.10 | 2.77 | 2.41 | 2.47 | 0.27 | The IARC of WHO |
| Breast(C50), all ages | 178 | 4.60 | 111.90 | 44.67 | 23.71 | 0.79 | 0.18 | 98 | 16.10 | 111.90 | 57.37 | 23.21 | 0.34 | 0.24 | 80 | 4.60 | 74.10 | 29.11 | 12.39 | 0.85 | 0.27 | The IARC of WHO |
| Cervix uteri (C53), all ages | 178 | 2.30 | 75.90 | 20.01 | 13.94 | 1.19 | 0.18 | 98 | 2.70 | 46.90 | 14.76 | 9.69 | 0.98 | 0.24 | 80 | 2.30 | 75.90 | 26.44 | 15.61 | 0.84 | 0.27 | The IARC of WHO |
| Colorectum (C18-21), all ages | 178 | 1.20 | 45.00 | 15.11 | 11.41 | 0.81 | 0.18 | 98 | 3.50 | 45.00 | 21.56 | 10.97 | 0.21 | 0.24 | 80 | 1.20 | 28.30 | 7.22 | 5.42 | 2.15 | 0.27 | The IARC of WHO |
| Corpus uteri (C54), all ages | 178 | 0.00 | 34.10 | 8.37 | 6.20 | 1.09 | 0.18 | 98 | 1.40 | 34.10 | 11.06 | 6.04 | 0.83 | 0.24 | 80 | 0.00 | 26.70 | 5.09 | 4.64 | 2.24 | 0.27 | The IARC of WHO |
| Lung (C33-34), all ages | 178 | 0.20 | 51.60 | 15.41 | 12.43 | 0.65 | 0.18 | 98 | 3.30 | 51.60 | 21.52 | 11.47 | 0.20 | 0.24 | 80 | 0.20 | 44.20 | 7.93 | 9.03 | 1.96 | 0.27 | The IARC of WHO |
| Melanoma of skin (C43), all ages | 178 | 0.00 | 35.80 | 3.40 | 5.57 | 3.14 | 0.18 | 98 | 0.00 | 35.80 | 5.38 | 6.87 | 2.21 | 0.24 | 80 | 0.00 | 4.20 | 0.98 | 0.81 | 1.81 | 0.27 | The IARC of WHO |
| Ovary (C56), all ages | 178 | 0.80 | 14.90 | 6.25 | 2.96 | 0.61 | 0.18 | 98 | 1.60 | 14.90 | 7.53 | 2.98 | 0.30 | 0.24 | 80 | 0.80 | 10.70 | 4.69 | 2.05 | 0.55 | 0.27 | The IARC of WHO |
| Stomach (C16), all ages | 178 | 0.90 | 41.80 | 7.83 | 6.17 | 2.08 | 0.18 | 98 | 0.90 | 41.80 | 8.79 | 6.18 | 2.34 | 0.24 | 80 | 0.90 | 32.50 | 6.67 | 5.98 | 1.96 | 0.27 | The IARC of WHO |
| GDP PPP 2010 | 178 | 12.35 | 25.51 | 18.90 | 2.87 | 0.23 | 0.18 | 98 | 15.21 | 25.51 | 20.53 | 2.41 | -0.17 | 0.24 | 80 | 12.35 | 23.43 | 16.90 | 2.01 | 0.90 | 0.27 | The World Bank |
| Urbanization 2010 | 170 | 671.01 | 126613.79 | 16232.57 | 18632.70 | 2.32 | 0.19 | 94 | 5307.70 | 126613.79 | 26553.24 | 19622.99 | 2.13 | 0.25 | 76 | 671.01 | 11028.50 | 3467.53 | 2492.11 | 1.11 | 0.28 | The World Bank |
| Life Expectancy ($e_{60}$, 2005-2010) | 178 | 9.09 | 100.00 | 55.88 | 23.28 | -0.06 | 0.18 | 98 | 9.09 | 100.00 | 69.69 | 18.39 | -0.71 | 0.24 | 80 | 10.64 | 82.47 | 38.95 | 16.51 | 0.46 | 0.27 | The United Nations |
| Biological State Index ($I_{bs}$) | 172 | 0.63 | 0.99 | 0.92 | 0.08 | -1.32 | 0.19 | 94 | 0.77 | 0.99 | 0.97 | 0.03 | -3.79 | 0.25 | 78 | 0.63 | 0.98 | 0.87 | 0.09 | -0.46 | 0.27 | Self calculated |
| Valid N (listwise) | 56 | | | | | | | 44 | | | | | | | 12 | | | | | | | |

## 3rd response to reviewers' comments

BCAN-D-16-02324R2

Greater family size is associated with less cancer risk: an ecological analysis of 178 countries Wenpeng You; Frank Rühli; Renata Henneberg; Maciej Henneberg BMC Cancer

**Technical Comments:**

1.Please include e-mail addresses for all authors on the title page.

Authors: Now all the authors' email addresses are included in the title page.

2. Please remove the point-by-point response to the reviewers from the additional files.

Authors: Will not upload the response as the additional file this time.

BMC Cancer operates a policy of open peer review, which means that you will be able to see the names of the reviewers who provided the reports via the online peer review system. We encourage you to also view the reports there, via the action links on the left-hand side of the page, to see the names of the reviewers.

Authors: Thanks for your reminding.

**Reviewer reports:**

Hauke Thomsen (Reviewer 1): Dear Ladies and Gentlemen, although the comments of my last review have been considered in the latest revision of the manuscript, I still a few comments that need to be taken care of:

1. in the third section of the discussion you state: "Our findings were supported by the hypotheses that ..." This is the complete wrong way: Hypotheses should be supported by findings !

Authors: Now this sentence has been revised as

Our findings were in agreement with the conclusions from previous studies that greater family size protected family members from developing bladder cancer------

2. in the next paragraph you state that your study revealed that family size protects against corpus uteri cancer. However, in the beginning of the discussion you state: "correlation between two variables does not imply causality." Therefore, I would like to ask you to change the statements such as "family size protects ..." or "family size increase risk ...." in the third section of the discussion to something like "family size is negatively/positively correlated with cancer risk. One might speculate about the reason for this correlation, but there is clearly no evidence for any protection. By the way: in one of our recent studies (Thomsen et al., European Journal of Human Genetics, Sept. 2014) we have included the number of children as a covariate in our estimation of heritability for Hodgkin lymphoma, because it had a significant effect.

Authors: Thanks. The original descriptions have been amended and the suggestion has been included.

3. The point raised in #2 should also be changed in the conclusion. This holds especially for the last sentence. Considering your statement: what is your recommendation to reduce cancer risk? For males: just having more children? I am not sure whether this should be a "strategic plan".

Authors: The conclusion has been revised. The two points concerning the reviewer have been incorporated in the updated conclusion.

In this study of the relationship between the family size and cancer incidence in 178 countries, we have identified that countries with greater family size have lower cancer incidence rates in males and females. This indicates that it may be worthwhile to consider in terms of cancer prevention whether both females, and especially males, may benefit from greater family size. Having more family life satisfaction may be included as a part of strategic plan of cancer prevention.

Accordingly, the section of Conclusion in the Abstract has been revised as well.

Article 9/10: Decreasing Birth rate determining worldwide incidence and regional variation of female breast cancer (Published at Advances in Breast Cancer Research, 2018)

Wenpeng You[1], Ian Symonds[1], Frank J. Rühli[2], Maciej Henneberg[1,2]

[1.] Adelaide Medical School, the University of Adelaide, Adelaide, Australia

[2.] Institute of Evolutionary Medicine, University of Zürich, Zürich, Switzerland

**Published:** You, W.P., Symonds, I., Rühli, F.J. and Henneberg, M. (2018). Decreasing Birth Rate Determining Worldwide Incidence and Regional Variation of Female Breast Cancer. Advances in Breast Cancer Research, 2018 , 7, 1-14. https://doi.org/10.4236/abcr.2018.71001

✉ **Correspondence:** Wenpeng You, wenpeng.you@adelaide.edu.au

## Contextual Statement

Breast cancer has been associated with decreasing birth rate, urbanization, overweight, ageing and GDP. However, these studies did not identify which risk factor is significant contributor to breast cancer. Additionally, it is confusing that WHO and its cancer research agent constantly stated that breast cancer incidence is greater in the developed world.

This study examined and identified that low birth rate is the significant determinant of breast cancer. We also found that decreasing birth rate, instead of GDP, may determine the significant variations of breast cancer incidence within the WHO regions.

## Statement of Authorship

# Statement of Authorship

| | |
|---|---|
| Title of Paper | Birth rate determining worldwide incidence and regional variation of female breast cancer |
| Publication Status | ☑ Published ☐ Accepted for Publication<br>☐ Submitted for Publication ☐ Publication Style |
| Publication Details | Citation: You, W.P., Symonds, I., Rühli, F.J. and Henneberg, M. (2018) Decreasing Birth Rate Determining Worldwide Incidence and Regional Variation of Female Breast Cancer. Advances in Breast Cancer Research, 7, 1-14. https://doi.org/10.4236/abcr.2018.71001 |

### Principal Author

| | | | |
|---|---|---|---|
| Name of Principal Author (Candidate) | Wenpeng You | | |
| Contribution to the Paper | Conceived the hypothesis, collected data, performed analysis on all samples, interpreted data, wrote manuscript and acted as corresponding author. | | |
| Overall percentage (%) | 50 | | |
| Signature | | Date | 22/12/2017 |

### Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

| | | | |
|---|---|---|---|
| Name of Co-Author | Ian Symonds | | |
| Contribution to the Paper | Discussed and consolidated the hypothesis with other authors, performed and interpreted the data analysis, helped to evaluate and edit the manuscript | | |
| Signature | | Date | 22nd Jan 2018 |

| | | | |
|---|---|---|---|
| Name of Co-Author | Frank J. Rühli | | |
| Contribution to the Paper | Conceived and discussed the hypothesis with other authors, helped to design the data analysis, interpreted the data analysis results, and edited and approved the manuscript | | |
| Signature | | Date | 19/01/2018 |

| | | | |
|---|---|---|---|
| Name of Co-Author | Maciej Henneberg | | |
| Contribution to the Paper | Supervised development of work, formalized and interpreted data analysis, and evaluated and edited manuscript | | |
| Signature | | Date | 22/12/2017 |

Please cut and paste additional co-author panels here as required.

## Abstract

**Purpose**: Urbanization, obesity and ageing and their associated with lifestyle changes (Westernized diet patterns, pollution, physical inactivity) have been proposed as the major contributing factors for the global rise in breast cancer (BCa) and have been the variables used to predict the future breast cancer rate. At the same time, socio-economic level, instead of birth rate, has been proposed for explanation of dramatic regional variations of breast cancer incidence. We sought to determine which factor plays the determining role in predicting worldwide breast cancer incidence rates and regional variations.

**Methods**: Bivariate correlation was conducted to examine the relationships between country-specific estimates of birth rate, BCa incidence, urbanization, overweight, ageing and GDP. Partial correlation was performed to identify the correlation between BCa incidence with each independent variable while we controlled the other four variables. Multiple linear regression was used to identify the most significant predictors of BCa incidence. Post hoc Scheff and independent T-Test analysis were performed to compare mean differences in BCa incidence rates and residuals of BCa standardised on birth rate in the WHO regions, and UN developed and developing regions respectively.

**Results**: Worldwide, BCa incidence rate tends to increase while birth rate decreases and urbanization, overweight, ageing and GDP increase. However, birth rate was the only variable that had a significant correlation with BCa incidence when controlled for the other four variables. Birth rate was the only significant predictor of BCa incidence in regression analysis. Multiple mean differences of BCa incidence between regions were significant, but all disappeared when the contributing effect of birth rate on BCa incidence rate was removed.

**Conclusions**: Birth rate plays a determining role in worldwide BCa incidence rate and regional variations. Current BCa projection methods may estimate future rates of BCa poorly if they fail to incorporate the impact of birth rate.

**Keywords:** Regional variations, Hormones, Breast cancer, Birth rate, Mean difference comparison

## Introduction

The global incidence rate of female breast cancer (BCa) has been on the rise since the 1970s even in the countries in Asia and Africa that had previously reported low rates. BCa is the most common invasive cancer in women, accounting for over 25% of all

cancer cases [1] and affecting about one in eight women during their lives. The WHO has concluded that life expectancy, urbanization and western lifestyles [2] are the major risk factors for BCa.

BCa is a disease with genetic background, but genetics may only explain 5–10% of all cases [3]. Most BCa cases occur due to the mutations caused by the interaction between an environmental factor and a genetically susceptible host [3].

Ageing, which may influence carcinogenesis, has been regarded as a prime contributing factor to BCa [4]. Tobacco smoking has been long postulated as one of the environmental factors to cause BCa [5]. For instance, the risk of BCa may be increased from 35% to 50% in female smokers. Anti-smoking campaigns have reduced the rate of smoking in women in the developed world [6], but the BCa incidence rate remains much greater [7] than in the developing world, and the incidence rate in the developed world continues to rise [8].

Several other alternative hypotheses about the relationships between BCa and contributing environmental factors have been explored in the past decades. Decreasing physical activity has been associated with the increase of BCa risk, although the mechanism of effect of exercise is not fully established. Supplemental to this conclusion, the Lancet Physical Activity Series Working Group [9] reported that BCa risk may be reduced when females become physically active [9]. High-fat [10] and/or high-alcohol [11] diet patterns have been related to BCa risk.

The incidence of BCa varies greatly around the world. The WHO has associated regional variations with country groups due to their different socio-economic levels [12]. Genetic differences between ethnic groups have also been implicated in the genesis of regional variations. Perhaps relaxed natural selection is involved through the accumulation of BCa genes or mutations due to modern medicine advancement, which allows early onset BCa patient to survive, but makes BCa genes inheritable to their next generation [13].

Female reproductive behaviour was initially postulated to be associated with BCa risk since it was greater among nulliparous Catholic nuns 300 years ago [14]. Specifically, the postulation that female childbearing reduces BCa risk was advanced in the 1920s and confirmed in 1970’s [15]. The underlying mechanism for this relationship is that pregnancy breaks menstrual cycles, which reduces breast exposure to estrogen. Studies have identified that estrogen may cause DNA damage and thus initiation of BCa [16]. Recent studies have shown that estrogen receptor (ER) positive BCa may make up approximately 70% of all BCa [17] and that child-bearing may decrease the risk of developing BCa by up to 50% [18].

The incidence of BCa varies greatly around the world. Genetic differences between ethnic groups have been implicated in the genesis of regional variations. Perhaps relaxed natural selection is involved through the accumulation of BCa genes or mutations due to modern medicine advancement, which allows early onset BCa patient to survive, but makes BCa genes inheritable to their next generation [13, 19].

Professionals and laypeople are still intrigued with the mechanisms about how physical activities, diet patterns, genetic background and reproduction behaviour contribute to BCa from the perspective of physiology. However, a number of publications have reported that females with higher socioeconomic levels may be subject to higher risk of BCa [20, 21]. Furthermore, as the directing and coordinating authority for health within the United Nations system, the WHO and its cancer research agent, the International Agency for Research on Cancer (IARC) also support the theory that female's socioeconomic level is associated with BCa risk [12, 22]. Therefore, females in the developed world and those at high socioeconomic levels in developing countries may have wondered what is wrong to be at higher socioeconomic level?

The present study starts with measures of proximal causes of BCa, analysing how BCa incidence rate relates to birth rate, socio-economic factors, urbanization, overweight and ageing. It then assesses which underlying factors, from socio-economic factors, urbanization, overweight to ageing to birth rate, account for significant proximal risks and overall BCa incidence. Finally, it shows that birth rate plays the determining role in contributing to regional variations of BCa incidence rate.

## Materials and Methods

### Data Sources

The country specific variables were collected for this ecological study.

#### The GLOBOCAN 2012 estimates of incidence rate of female BCa (C50) [7]

BCa incidence rate indicates the number per 100,000 females who were diagnosed with BCa in 2012. The rate was age-standardized using the World standard population to increase the comparability.

GLOBOCAN is a project conducted by the International Agency for Research on Cancer (IARC) of the WHO. This project provides contemporary population level estimates by cancer site and sex using the best available data in each population and nine comprehensive methods of estimation [23].

#### The World Bank published data on birth rate, GDP and urbanization

Birth rate indicates the number of live births per 1,000 population occurring at midyear during the year 1992.

Socio-economic levels measure with GDP have been related to BCa incidence rate [7]. GDP is used as the index of socio-economic level and it is expressed in per capita purchasing power parity (PPP in current international $) in 2010.

Urbanization is expressed with the percentage of total population living in urban areas in 2010. Urbanization, representing a major demographic shift, entails lifestyle changes, including diet with more energy dense components, such as high fat and high alcohol in daily diet, and less physical exercise [24]. Therefore, urbanization has been postulated as a major BCa predictor [24].

**The United Nations Statistics Division estimates of the life expectancy**

Life expectancy, indexed as ageing in this study, has been considered as an attributable factor to BCa [25]. Women age 50+ enter menopause, which is leads to fall in estrogen levels. Therefore, life expectancy ($e_{50}$, 2005-2010) was extracted from abridged life tables (1950-2100) published online by the United Nations.

The WHO Global Health Observatory (GHO) has published data on the estimated prevalence rate of women who are overweight. The overweight prevalence is expressed as the percentage of the population (2010) aged 18+ with a BMI ≥ 25 kg/m$^2$. Being overweight also has been postulated as a risk factor of BCa.

**Data Selection**

We used country specific BCa incidence rate, birth rate, GDP (index of socio-economic level), urbanisation, overweight prevalence (Western lifestyle) and life expectancy (ageing) for all countries where data were available. We matched BCa incidence rates and birth rate by country and we obtained a set of data consisting of 179 countries.

Each country was treated as an individual in the analysis. the numbers of countries included in the analysis of relationships with other variables may have differed somewhat because all information was not uniformly available for all countries.

**Data analysis**

Various statistical analysis methods were applied in this study to explore the correlation between birth rate and BCa incidence rate.

**Data robustness and variable distributions check**

Scatter plots were used to explore and visualize the correlations between birth rate and BCa incidence rate. The strength and form of the relationship between BCa and birth rate were analysed using actual values of the two variables. For other analyses, variable values were logarithmically transformed to bring their distributions closer to normality.

To examine the correlation between birth rate and BCa incidence, the underlying contributing factors of BCa risk and the determining role of birth rate in regional variation, the analysis proceeded in four steps:

Pearson's r and nonparametric correlations were used to evaluate the strength and direction of the correlation between all the variables.

The independent relationships between BCa and each of the five independent variables are explored with partial correlation of Pearson's moment-product approach while we controlled for the other four variables. This allows the identification of the strongest correlation and its independency.

Standard multiple linear regression (enter) was performed to describe the relationships between the outcome variable and the explanatory variables. In order to highlight that birth rate is the major population-level contributor to BCa incidence, standard multiple linear regression (enter) was also conducted to calculate the correlation between BCa incidence and the risk factors when birth rate is included and excluded respectively.

The equation of the best fitting trendline (polynomial) displayed in the scatter plots analysis of relationship between birth rate and BCa incidence was used to calculate and remove the contributing effect of birth rate on BCa incidence rate, which allowed the creation of a new dependent variable, "Residual of BCa standardised on birth rate". Means of the BCa incidence rate and the "Residuals of BCa standardised on birth rate" of all the countries were calculated for mean difference comparison. Countries were categorized as per the UN common practice of defining more developed and developing countries and WHO regions for investigating the regional variations based on mean difference.

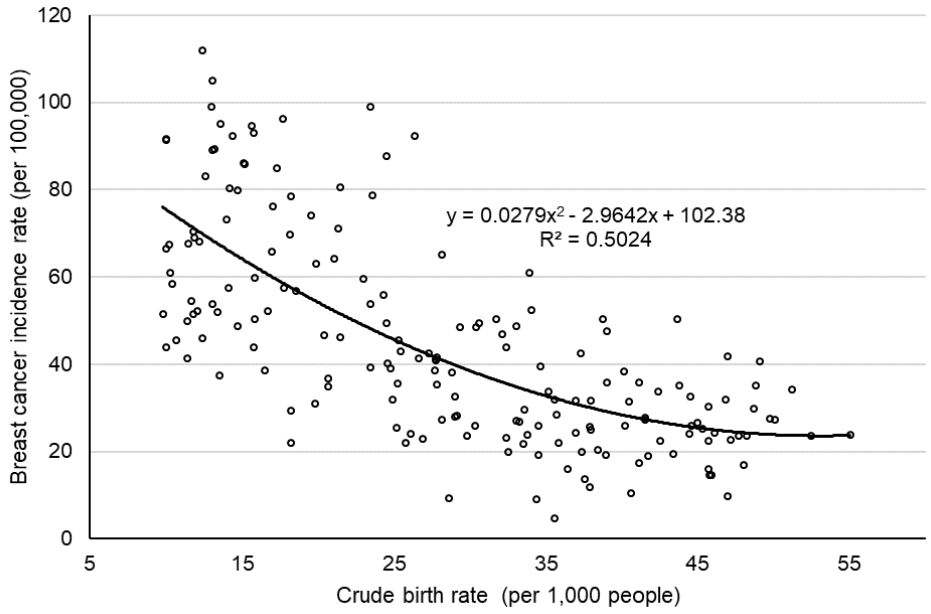
Independent Samples T-test was conducted to compare the means of the two BCa incidence variables of the pairs of UN country groupings. Post hoc Scheffe (Oneway ANOVA) testing was performed to compare difference of multiple means between six WHO regions.

Scatter plots and calculation of means were performed in Excel® (Microsoft 2016). Pearson and partial correlations, multiple linear regression analysis, Independent Samples T-test and Post hoc Scheffe for mean comparison were conducted using SPSS v. 22 (SPSS Inc., Chicago Il USA). The original data was used for scatter plots and mean calculation of BCa incidence rate and "Residual of BCa standardised on birth rate". To increase homoscedasticity of data distributions log transformed variables were used for correlation analyses. The significance was kept at the 0.05 level, but 0.01 and 0.001 levels are also reported. Standard multiple linear regression analysis criteria were set at probability of F to enter ≤ 0.05 and probability of F to remove ≥ 0.10.

## Results

Figure 1 shows that the relationship between the birth rate and BCa incidence rate is polynomial with a strong negative correlation ($R^2$=0.5024).

Figure 1. The relationship between birth rate and breast cancer incidence rate




The non-linear relationship between birth rate and BCa incidence variables identified in the scatterplots shows the strong correlation between birth rate and BCa incidence. This relationship was confirmed by the subsequent analyses of log-transformed data and in nonparametric analysis.

Worldwide, birth rate was significantly correlated to BCa incidence (r=-0.680 and rho= -0.723, p<0.001 respectively in Pearson and non-parametric analyses) (Table 1).

Table 1 showed that not only birth rate, but also GDP, urbanization, overweight prevalence and ageing correlate significantly to BCa incidence rates in both Pearson and non-parametric analyses.

There was a strong and highly significant correlation between GDP and birth rate (r=-0.760 and rho= -0.797, p<0.001 respectively in Pearson and non-parametric analysis).

Table 1: Pearson r (above the diagonal) and nonparametric (below the diagonal) correlation between all variables

| | Birth rate | Breast Cancer | GDP | Urbanization | Overweight | Ageing |
|---|---|---|---|---|---|---|
| Birth rate | 1 | -0.680*** | -0.760*** | -0.557*** | -0.397*** | -0.753*** |
| Breast Cancer | -0.723*** | 1 | 0.639*** | 0.474*** | 0.394*** | 0.611*** |
| GDP | -0.797*** | 0.694*** | 1 | 0.702*** | 0.581*** | 0.766*** |
| Urbanization | -0.619*** | 0.551*** | -0.764*** | 1 | 0.482*** | 0.618*** |
| Overweight | -0.401*** | 0.385*** | 0.519*** | 0.475*** | 1 | 0.414*** |
| Ageing | -0.784*** | 0.649*** | 0.775*** | 0.676*** | -0.392*** | 1 |

The table describes the bivariate correlation between all the variables. *** p<0.001; Country number: 171-179.

Breast cancer incidence rate is from the International Agency for Research on Cancer. Birth rate, GDP and urbanization are from the World Bank. Ageing expressed as life expectancy ($e_{50}$) is from the United Nations. Overweight prevalence is the World Health Organization.

The relationship between dependent variable (BCa) and each independent variable (birth rate, GDP, urbanization, overweight and ageing) was examined by controlling for the other four variables in a partial correlation analysis. Birth rate was the only the independent variable to have a strong and significant correlation (r= -0.330, p< 0.001) with BCa independent of the other four variables (Table 2). None of the other four variables (GDP, urbanization, overweight and ageing) showed a correlation with BCa incidence independent of the other four variables despite the fact that each of them (GDP, urbanization, overweight and ageing) had a strong significant correlation to BCa incidence in Pearson r and non-parametric correlation analysis. This suggests that birth rate is the independent determinant of the secondary association between BCa incidence and environmental factors.

Table 2 Comparison of partial correlation coefficients between breast cancer incidence and each variable when the other four variables are kept constant

| | Birth rate | | | GDP | | | Urbanization | | | Overweight | | | Ageing | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variables | r | p | df | r | p | df | r | p | df | r | p | df | r | p | df |
| Birth rate | -0.330 | <0.001 | 162 | - | - | - | - | - | - | - | - | - | - | - | - |
| GDP | - | - | - | 0.129 | 0.099 | 162 | - | - | - | - | - | - | - | - | - |
| Urbanization | - | - | - | - | - | - | 0.001 | 0.994 | 162 | - | - | - | - | - | - |
| Overweight | - | - | - | - | - | - | - | - | - | 0.077 | 0.327 | 162 | - | - | - |
| Ageing | - | - | - | - | - | - | - | - | - | - | - | - | 0.103 | 0.187 | 162 |

The table describes the partial correlation between breast cancer incidence between each variable while the other four variables are controlled for. - Controlled variable

Breast cancer incidence rate is from the International Agency for Research on Cancer. Birth rate, GDP and urbanization are from the World Bank. Ageing expressed as life expectancy ($e_{50}$) is from the United Nations. Overweight prevalence is the World Health Organization.

Standard multiple linear regression (enter) analysis was applied to further predict BCa incidence when birth rate, GDP, urbanization, overweight and ageing were used as the independent variables.

When birth rate is excluded as one of the independent variables, GDP (β=0.401, p<0.001) and ageing (β=0.300, p<0.001) are the two significant predictors of BCa incidence. However, when birth rate was included as an independent variable, the correlations between BCa incidence and both GDP and ageing become very weak and no longer reach statistical significance (Table 3). This supports our previous suggestion that birth rate is the principal and independent determinant of BCa incidence in partial correlation analysis.

Table 3 Independent predictors of breast cancer incidence rate based on multiple linear regression modelling

| Variable | β | Std. Error | Sig. | β | Std. Error | Sig. |
|---|---|---|---|---|---|---|
| Birth rate | - | - | - | -0.460 | 0.106 | <0.001 |
| GDP | 0.401 | 0.050 | <0.001 | 0.193 | 0.051 | 0.083 |
| Urbanization | -0.015 | 0.093 | 0.856 | -0.008 | 0.086 | 0.916 |
| Overweight | 0.038 | 0.111 | 0.605 | 0.048 | 0.104 | 0.480 |
| Ageing | 0.300 | 0.369 | <0.001 | 0.102 | 0.380 | 0.285 |

The table describes the multiple linear regression analysis results including and excluding birth rate as a predictor of breast cancer. *df* = 167; - excluded variable

Breast cancer incidence rate is from the International Agency for Research on Cancer. Birth rate, GDP and urbanization are from the World Bank. Ageing expressed as life expectancy ($e_{50}$) is from the United Nations. Overweight prevalence is the World Health Organization.

Table 4 shows that the mean BCa incidence rate was lowest in South-East Asia (26.31) and highest in Europe (63.60). The means of BCa in the other four regions are Africa (26.99), Eastern Mediterranean (40.77), Western Pacific (43.03) and Americas (46.98).

A post hoc Scheffe analysis conducted on the multiple mean comparisons revealed that there were a number of significant mean differences in BCa incidence rates between different WHO regions (Table 4). Mean of BCa incidence in Africa was significantly lower than that in Americas, Europe and Western Pacific. Mean of BCa incidence in the Americas was significantly lower than that in Europe and Western Pacific. Mean of BCa incidence in Eastern Mediterranean was significantly lower than that in Europe. The mean BCa incidence in South-Eastern Asia was significantly lower than that in Americas, Europe and Western Pacific. Whilst the mean BCa incidence in Western Pacific was significantly lower than that in Europe.

A subsequent ANOVA with post hoc Scheffe procedure performed on the means of "Residual of BCa standardised on birth rate" in different WHO regions showed no significant differences among and between regions (Table 4).

Interestingly, mean BCa incidence in the developed regions was significantly greater than that in the developing regions (mean difference=9.75, $p<0.001$). However, the difference between the means of the "Residual of BCa standardised on birth rate" in the developed region and developing region is weak and does not reach statistical significance (Table 4).

The results from post hoc Scheffe tests conducted on mean comparison between the WHO regions suggest that regional variations of BCa incidence may only reach statistically significant levels if the contribution of their respective birth rates is included. In other words, except for birth rate, the contribution of the other BCa predicting factors to BCa incidence may not be sufficient for the difference in mean rates to reach significance. This result is supported by the findings identified in our previous partial correlation (Table 2) and multiple linear regression (Table 3) that birth rate is the critical risk factor of BCa.

Table 4 Mean difference between WHO regions, and between UN developed and developing regions

| | | BCa incidence rate | | Residual of BCa standardised on birth rate |
|---|---|---|---|---|
| WHO regions (One-way ANOVA, Post hoc Scheffe) | | | | |
| I | J | Mean difference (I-J) | J | Mean difference (I-J) |
| AF, n=45, BCa incidence mean= 26.99 | AM | -20.33*** | AM | -2.93 |
| | EM | -13.78 | EM | -4.78 |
| | EU | -36.62*** | EU | -0.75 |
| | SEA | 0.68 | SEA | 13.40 |
| | WP | -20.50*** | WP | -4.25 |
| AM, n=31, BCa incidence mean= 46.98 | AF | 20.33*** | AF | 2.93 |
| | EM | 6.55 | EM | -1.85 |
| | EU | -16.28*** | EU | 2.18 |
| | SEA | 21.01* | SEA | 16.33 |
| | WP | -0.17 | WP | -1.32 |
| EM, n=21, BCa incidence mean= 40.77 | AF | 3.78 | AF | 4.78 |
| | AM | -6.55 | AM | 1.85 |
| | EU | -22.83*** | EU | 4.03 |
| | SEA | 14.46 | SEA | 18.19 |
| | WP | -6.72 | WP | 0.53 |
| EU, n=50, BCa incidence mean=63.60 | AF | 36.62*** | AF | 0.75 |
| | AM | 16.28*** | AM | -2.18 |
| | EM | 22.83*** | EM | -4.03 |
| | SEA | 37.29*** | SEA | 14.15 |
| | WP | 16.11* | WP | -3.50 |
| SEA, n=11, BCa incidence mean= 26.31 | AF | -0.68 | AF | -13.40 |
| | AM | -21.01* | AM | -16.33 |
| | EM | -14.46 | EM | -18.19 |
| | EU | -37.29*** | EU | -14.15 |
| | WP | -21.18* | WP | -17.66 |
| WP, n= 21, BCa incidence mean=43.03 | AF | 20.50*** | AF | 4.25 |
| | AM | 0.17 | AM | 1.32 |
| | EM | 6.72 | EM | -0.53 |
| | EU | -16.11* | EU | 3.50 |
| | SEA | 21.18* | SEA | 17.66 |
| United Nations region classifications based on common practice (Independent T-test, two tailed) | | | | |
| | | t | | t |
| Developed, n=44 | Developing | 9.75*** | Developing | 0.56 |

Mean difference comparison results are reported. * p<0.05, ** p<0.01, *** p<0.001

Breast cancer incidence rate is from the International Agency for Research on Cancer. Birth rate, GDP and urbanization are from the World Bank. Ageing expressed as life expectancy (e50) is from the United Nations. Overweight prevalence is the World Health Organization.

Abbreviations: AF, Africa; AM, Americas; EM, Eastern Mediterranean; EU, Europe; SEA, South-East Asia; WP, Western Pacific

## Discussion

The worldwide trend of increased BCa incidence may have multiple aetiologies, which may act through multiple mechanisms. Our ecological analysis suggests that birth rate may be a determining factor of BCa incidence at the population level. This study also reveals that the effect of birth rate on BCa incidence is independent of the effects of socio-economic factors, urbanization, overweight and ageing.

The results of this study show that, a country with greater birth rate may have lower BCa incidence. This supports the observation from previous studies that higher parity is associated with a decreased risk of BCa based on observational approaches. This study used the ecological approach, which has an advantage over the observational studies in terms of obtaining more variables [26] for data analysis. For instance, we were able to use 5 variables, which allowed us to control for four variables, including the socio-economic factor (GDP), which has been used by the WHO to interpret the regional variations of BCa incidence [23].

The prevalent interpretation that greater birth rate protects against female BCa [26] is that the interruption in the normal menstrual cycle during pregnancy and subsequent breast feeding is associated with an interruption in the normal cyclical production of oestrogen [27], but an increase in oxytocin [28]. The public have been extensively educated for decades that oestrogen contributes to BCa as it fuels the growth of most breast cancer tumours [18].

Oxytocin, produced during pregnancy, delivery and breastfeeding, may have a role in the control of mammary cell growth [29] and inhibiting proliferation of human BCa cells, which may offer BCa prevention and treatment [30]. These findings have driven a hypothesis that oxytocin may have therapeutic effects on cancer [28]. Similarly, Misra *et al.* (2012) reported that females with greater parity may reduce their long-term BCa risk because of multiple hormones released during pregnancy that generate genetic changes in the mammary glands which decrease BCa risk in mature breast cells [31].

The WHO and its agent the IARC have endorsed the paradigm that BCa incidence is lower in less-developed countries but greater in the more-developed countries and this has been widely cited in a large body of literature to describe regional variations of BCa incidence [1]. This may lead to the impression that that GDP is the main risk factor of BCa. However, this paradigm is not supported by the results of the three statistical analyses in this study. Firstly, birth rate, other than GDP, is the only predicting factor which is correlated to BCa incidence independent of all the other four confounders in partial correlation analysis. Secondly, in this study, once the effects of birth rate are considered in multiple linear correlation analysis, the correlation between BCa incidence

and GDP and other variables disappears. Finally, it is the contribution of birth rate instead of GDP that accounts for the statistically significant regional variations.

In this study, birth rate was the principal determining predictor of BCa incidence and it may explain the correlation between GDP and BCa incidence. GDP shows a significant and strong correlation to birth rate in both Pearson r and non-parametric correlation analysis. This relationship is consistent with the theory of the demographic transition which proposed that a country or region may transition from high birth to lower birth rate when it is transforming to an industrialized economic system [32].

There are several caveats, including the one conceptualized as the ecological fallacy [33], to this study.

Firstly, each country is considered as a subject in this study. The country-specific data included in this study were aggregated, different from data collected from individual patients. Therefore, values for risk-modifying factors may not hold true for individuals to develop BCa.

Secondly, data aggregated and/or collected by the UN and its agencies (WHO, IARC and the World Bank) may include some random errors arising from methods of reporting incidence of BCa, reliability of diagnoses and possible administrative errors. For instance, data quality of the BCa incidence depends upon the quality and on the amount of the information available for each country. In general, data from developing countries are less complete than those from developed countries.

Finally, there are around 20 sub-types of BCa, such as ductal carcinomas and lobular carcinomas. This study only focuses on the hormone receptor-positive BCa. Recently, scientists at Boston University found that high parity was associated with an increased estrogen and progesterone receptor negative (ER-/PR-) BCa [34]. This suggests that high parity has dual effect on BCa, which our data analysis may not be able to explain.

**Authors' contributions**

WY conceived the idea, and IS, FR and MH consolidated the hypothesis. WY and MH conducted data analysis. All authors interpreted the data analysis result and provided suggestions for the manuscript writing. All authors reviewed, edited and approved the final manuscript.

**Conflict of interest statement**

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could

have influenced its outcome. We confirm that the manuscript has been read and approved by all named authors.

**Funding sources**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Ethical Approval**

All the aforementioned data were freely available from the official websites of the UN agencies. No ethical approval or written informed consent for participation was required.

## References

1. Stewart, B.W., *World Cancer Report 2014*. 2014, Lyon: Lyon, FRA: International Agency for Research on Cancer.
2. Chia, K.S., et al., *Profound changes in breast cancer incidence may reflect changes into a Westernized lifestyle: A comparative population-based study in Singapore and Sweden.* International Journal of Cancer, 2005. **113**(2): p. 302-306.
3. Gage, M., D. Wattendorf, and L. Henry, *Translational advances regarding hereditary breast cancer syndromes.* Journal of surgical oncology, 2012. **105**(5): p. 444-451.
4. McPherson, K., C. Steel, and J. Dixon, *Breast cancer—epidemiology, risk factors, and genetics.* Bmj, 2000. **321**(7261): p. 624-628.
5. Johnson, K.C., et al., *Active smoking and secondhand smoke increase breast cancer risk: the report of the Canadian Expert Panel on Tobacco Smoke and Breast Cancer Risk (2009).* Tobacco control, 2010: p. tc. 2010.035931.
6. Centers for Disease Control and Prevention of USA. *Cigarette Smoking Among Adults and Trends in Smoking Cessation---United States, 2008*. last reviewed: 11/12/2009; Available from: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5844a2.htm.
7. Ferlay, J., et al. *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]*. 2013 28.05.2016]; Available from: http://globocan.iarc.fr.
8. Sloan, F.A. and H. Gelband, *Committee on Cancer Control in Low and Middle Income Countries Board on Global Health*. The cancer burden in low-and middle-income countries and how it is measured. 2007, Washington, D.C: The National Academies Press.
9. Lee, I.-M., et al., *Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy.* The lancet, 2012. **380**(9838): p. 219-229.
10. Gl, B., *Dietary fat reduction and breast cancer outcome: results from the Women's Intervention Nutrition Study (WINS).* The American journal of clinical nutrition., 2007. **86**(3): p. 878S.
11. Sieri, S., et al., *Dietary fat intake and development of specific breast cancer subtypes.* Journal of the National Cancer Institute, 2014: p. dju068.

12. WHO. *Cancer mortality and morbidity*. WHO 2015 2015-03-27 07:42:24 2 November 2016]; Available from: http://www.who.int/gho/ncd/mortality_morbidity/cancer_text/en/.

13. You, W.-P. and M. Henneberg, *Type 1 diabetes prevalence increasing globally and regionally: the role of natural selection and life expectancy at birth.* BMJ Open Diabetes Research & Care, 2016. **4**(1): p. e000161.

14. Britt, K. and R. Short, *The plight of nuns: hazards of nulliparity.* The Lancet, 2012. **379**(9834): p. 2322-2323.

15. MacMahon, B., et al., *Age at first birth and breast cancer risk.* Bulletin of the World Health Organization, 1970. **43**(2): p. 209-221.

16. Miller, K., *Estrogen and DNA damage: the silent source of breast cancer?* Journal of the National Cancer Institute, 2003. **95**(2): p. 100-102.

17. American Cancer Society. *Targeted therapy for breast cancer*. 2016 22.10.2016]; Available from: http://www.cancer.org.

18. Dall, G., G. Risbridger, and K. Britt, *Mammary stem cells and parity-induced breast cancer protection- new insights.* J Steroid Biochem Mol Biol, 2016.

19. You, W. and H. M, *Cancer incidence increasing globally: The role of relaxed natural selection.* Evol Appl., 2017. **00:1–13**.

20. Faggiano, F., et al., *Socioeconomic differences in cancer incidence and mortality.* IARC Scientific Publications, 1997(138): p. 65-176.

21. van Loon, A.J.M., et al., *Differences in cancer incidence and mortality among socio-economic groups.* Scandinavian journal of social medicine, 1995. **23**(2): p. 110-120.

22. WHO. *Breast cancer: prevention and control*. WHO 2016 2016-01-21 11:17:19 2 November 2016]; Available from: http://www.who.int/cancer/detection/breastcancer/en/.

23. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* Int J Cancer, 2015. **136**(5): p. E359-86.

24. Allender, S., et al., *Quantification of urbanization in relation to chronic diseases in developing countries: a systematic review.* J Urban Health, 2008. **85**(6): p. 938-951.

25. Majeed, A., et al., *Trends in prostate cancer incidence, mortality and survival in England and Wales 1971–1998.* BJU International, 2000. **85**(9): p. 1058-1062.

26. Collaborative Group on Hormonal Factors in Breast Cancer, *Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50 302 women with breast cancer and 96 973 women without the disease.* The Lancet, 2002. **360**(9328): p. 187-195.

27. Ma, H., et al., *Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies.* Breast Cancer Research, 2006. **8**(4): p. 1.

28. Imanieh, M.H., et al., *Oxytocin has therapeutic effects on cancer, a hypothesis.* European Journal of Pharmacology, 2014. **741**: p. 112-123.

29. Sapino, A., et al., *Oxytocin enhances myoepithelial cell differentiation and proliferation in the mouse mammary gland.* Endocrinology, 1993. **133**(2): p. 838-842.

30. Cassoni, P., et al., *Oxytocin inhibits proliferation of human breast cancer cell lines.* Virchows Archiv, 1994. **425**(5): p. 467-472.

31. Misra, Y., et al., *Mammary gland morphological and gene expression changes underlying pregnancy protection of breast cancer tumorigenesis.* Physiological genomics, 2012. **44**(1): p. 76-88.

32. Thompson, W., *Population.* American Journal of Sociology, 1929. **34**(4): p. 959.

33. Winzar, H., *The ecological fallacy: How to spot one and tips on how to use one to your advantage.* Australasian Marketing Journal (AMJ), 2015. **23**(1): p. 86-92.

34. Palmer, J.R., et al., *Dual effect of parity on breast cancer risk in African-American women.* Journal of the National Cancer Institute, 2003. **95**(6): p. 478-483.

Article 10/10: Low fertility may be a significant determinant of ovarian cancer worldwide

Wenpeng You[1], Ian Symonds[1], Maciej Henneberg[1,2]

[1.] Adelaide Medical School, the University of Adelaide, Adelaide, Australia

[2.] Institute of Evolutionary Medicine, University of Zürich, Zürich, Switzerland

**Under review:** Wenpeng You, Ian Symonds, Maciej Henneberg. Low parity may be a significant determinant of ovarian cancer worldwide. Under review by BMC Journal of Ovarian Research.

✉ **Correspondence:** Wenpeng You wenpeng.you@adelaide.edu.au

## Contexture Statement

Ovarian cancer (OC56) has multiple aetiologies which may act through different risk factors, such as ageing, GDP, obesity, low parity, reduced natural selection and urbanization.

This study compared the contributing effects of ageing, GDP, obesity, low parity, reduced natural selection and urbanization on OC56, and we identified that low birth rate may be a significant determinant of OC56.

More oestrogen production due to less childbearing has been linked to high risk of OC56. This may be a suggestion to health research authority for prioritizing the studies on how to reduce ovulation rates for those females who do not want to have children or do not want to have more children.

## Statement of Authorship

# Statement of Authorship

| | |
|---|---|
| Title of Paper | Low parity may be a significant determinant of ovarian cancer worldwide |
| Publication Status | ☐ Published ☐ Accepted for Publication<br>☑ Submitted for Publication ☐ Publication Style |
| Publication Details | Under review by BMC Journal of Ovarian Research. Manuscript ID: JOVR-D-18-00051 |

### Principal Author

| | | | |
|---|---|---|---|
| Name of Principal Author (Candidate) | Wenpeng You | | |
| Contribution to the Paper | Conceived the hypothesis, collected data, performed analysis on all samples, interpreted data, wrote manuscript and acted as corresponding author. | | |
| Overall percentage (%) | 60 | | |
| Signature | | Date | 26/02/2018 |

### Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

| | | | |
|---|---|---|---|
| Name of Co-Author | Ian Symonds | | |
| Contribution to the Paper | Discussed and consolidated the hypothesis with other authors, performed and interpreted the data analysis, evaluate and edited the manuscript | | |
| Signature | | Date | 6/3/2018 |

| | | | |
|---|---|---|---|
| Name of Co-Author | Maciej Henneberg | | |
| Contribution to the Paper | Supervised development of work, formalized and interpreted data analysis, and evaluated and edited manuscript. | | |
| Signature | | Date | 6/3/2018 |

Please cut and paste additional co-author panels here as required.

## Abstract

**Background**: Ageing, GDP, obesity, fertility, reduced natural selection (measured by Ibs) and urbanization have been postulated as the risk factors of ovarian cancer (OC56). We sought to identify which factor plays the most significant role in predicting OC56 incidence rate worldwide.

**Methods**: Bivariate correlation was performed to assess the correlations between country-specific estimates of ageing (measured by life expectancy), GDP PPP, obesity prevalence, fertility (indexed by birth rate), $I_{bs}$ and urbanization. Partial correlation was used to compare variables and identify that fertility was the only variable strongly correlated to OC56 independent of the other five variables. Fisher A-to-Z was used to compare the correlation coefficients. Multiple linear regression (Enter and Stepwise) was conducted to identify significant determinants of OC56 incidence. Post hoc Bonferroni analysis was performed to compare mean differences between the means of OC56 incidence rate and residuals of OC56 standardised on fertility and GDP respectively between the six WHO regions.

**Results**: Bivariate analyses revealed that OC56 was significantly and strongly correlated to ageing, GDP, obesity, fertility, $I_{bs}$ and urbanization. However, partial correlation analysis only identified that fertility and ageing were the two variables that had significant and strong correlation to OC56 incidence when the other five variables are kept statistically constant, but Fisher A-to-z revealed that fertility correlated to OC56 significantly stronger than ageing. Both Enter and Stepwise regression analyses indicated that fertility was the only significant variable predicting OC56 risk. Post hoc Bonferroni analysis showed that, between the six WHO regions, multiple mean differences of OC56 incidence were significant, but all disappeared when the contributing effect of fertility on OC56 incidence rate was removed.

**Conclusions**: Low fertility may be the significant determinant of OC56 incidence increase worldwide. The health research authorities need to prioritize the studies into reducing the number of ovulation cycles for protecting those females who choose to be nulliparous or not to have more children from developing OC56.

**Keywords:** Ovarian cancer, Fertility, Oxytocin, Significant predictor, Psychological well-being

## Introduction

Ovarian Cancer (OC56) [1] has been a leading cause of cancer incidence and mortality globally. It ranks among the top ten diagnosed cancers and top five deadliest cancers in most countries [2, 3]. In 2015, OC56 was present in 1.2 million women and resulted in

161,100 deaths worldwide [4]. In the 21st century, a woman's overall lifetime risk of developing OC is around 1.6% [5,2, 6], and her chance of dying of the disease is 1 in 100 [2, 6].

Although OC56 has been known to scientists for over 150 years [7], the etiology of this lethal disease is not well understood. Traditionally, the majority of researches into OC56 has focused directly on the carcinogenic factors, such as talc, pesticides, red meat and alcohol in diet, smoking, and herbicides. However, to date, none of these factors has been consistently considered as the real risk factor of OC56 [8] due to circumstantial study designs and controversial conclusions.

In the past decades, alternative hypotheses have also been explored. Various studies postulated that, overall, obese women (those with a body mass index of at least 30) may have a higher risk of developing OC56 because their rising levels of estrogen circulation [9-11]. Increased age has been considered as a risk factor for OC56 because more mutations in cells can accumulate and eventually cause OC56 [9]. Urbanization may have improved public hygiene, sanitation and access to health care for females [12], but it has been associated with public health issues, including OC56 [13] due to the changes in occupational, dietary and exercise patterns in females [12]. Dietary factors, such as alcohol consumption [6] and low level of Vitamin D [14], and lifestyles, such as physical activities [6] and smoking [15] are also associated with OC56, but so far the available results are not conclusive [6].

Recently, the researchers from the University of Adelaide have conducted a number of studies of the role of the relaxed natural selection (measured by Biological State Index, $I_{bs}$) in accumulating the deleterious genes/mutations of non-communicable diseases, including cancers [16], Type 1 diabetes [17] and obesity [18]. Conclusions from these studies indicate that reduced natural selection may be an important contributor of increasing OC56 incidence globally [16].

An in-depth internet and literature search was conducted for associations between fertility and female behaviours. It has turned out that the OC56 risk increases in women who have ovulated more over their lifetime due to infertility and less fertility as they may produce more oestrogen which increases the OC56 risk [6, 19-21], but less oxytocin, which has been associated with less OC56 risk [22, 23]. Therefore, studies have shown that fertility may outscore the importance of other reproduction related factors [24-26].

To the best of our knowledge, despite that fertility is a well-established risk factor of OC56, no research has compared the contributing effects of fertility on CO56 with other OC56 risk factors, such as ageing, $I_{bs}$ (index of magnitude of OC56 gene accumulation in

human population), obesity and socioeconomic factors (GDP and urbanization) respectively.

Globally, OC56 incidence presents significant variations in different geographic regions [2, 3, 27-29]. This phenomenon has also been observed in different populations [6, 13] within the same countries [30, 31]. A number of publications suggest that the disparity between regions and populations has been associated with socioeconomic level. Some of these studies were published in the reputable journals by the International Agency for Research on Cancer (IARC), the specialized cancer agency of World Health Organization (WHO). Therefore, the empirical findings in those publications may have easily intrigued the professionals and laypeople. Females in the developed regions/nations may wonder why their wealth makes them exposed to high risk of developing OC56?

In this study, we drew on the empirical and macro-level data to test the hypotheses that fertility (measured by birth rate) is the significant determinant of OC56, and that it is fertility, instead of GDP, that is most important factor in shaping the regional variation of OC56 incidence rate.

## Materials and Methods

The country specific data published by the agencies of the United Nations were collected for this study.

1. The GLOBOCAN 2012 estimates of incidence rate of female OC56 [28]

GLOBOCAN provides contemporary population level estimates by cancer site and sex [2]. This project is conducted by the WHO cancer research agency, the International Agency for Research on Cancer (IARC).

OC56 incidence rate is expressed as the number per 100,000 females who were diagnosed with OC56 in 2012. The age-standardized OC56 incidence rate was selected in the interest of the data comparability between countries.

2. The World Bank published data [32] on birth rate, per capita GDP PPP and urbanization

Crude birth rate indicates the number of live births occurring during the year, per 1,000 population estimated at midyear. Crude birth rate (CBR) was used to index the fertility in this study, and it is backdated 20 years (1992) to reflect long exposure with delayed presentation of OC56.

Socio-economic level has been associated with OC56 risk [33] [2, 28, 34]. We chose per capita GDP purchasing power fertility (GDP PPP in 2012 international $) because it takes into account the relative cost of local goods, services and inflation rates of the country.

Urbanization has been postulated as a major OC56 predictor [35, 36] because it represents the major demographic shift entailing lifestyle changes [12, 37, 38]. Urbanization is expressed with the country-specific percentage of total population living in urban areas in 2012.

3. The United Nations Statistics Division estimates of the life expectancy [39]

The country-specific life expectancy, which indexes the ageing, has been well established as the attributable factor to OC56 [40] [41]. Therefore, we selected life expectancy ($e_{65}$, 2005-2010) [39] to index the ageing process at population level.

4. The magnitude of OC56 gene accumulation in a population indexed with the Biological State Index ($I_{bs}$)

Country specific $I_{bs}$ was downloaded from the previous publication [18]. It has been postulated that reduced natural selection (measured by $I_{bs}$) may have accumulated the deleterious genes of non-communicable diseases such as cancers [16], Type 1 diabetes [17] and obesity [18] at population level.

5. The WHO Global Health Observatory (GHO) data on obesity prevalence

Obese females may pose more risk to OC56 than those who are not obese [42]. The country-specific percentage of the females aged 18+ with a BMI ≥ 30 kg/m$^2$ in 2010 was extracted from the GHO data repository [43].

**Data Selection**

We collected country specific OC56 incidence rates, ageing, fertility, GDP, $I_{bs}$, obesity and urbanization for all countries where data were available. We extracted OC56 incidence rates for 182 countries and then the other variables were matched individually with OC56.

Each country is treated as an individual study subject in the data analysis. Not all the countries (subjects) have all the information for all the variables. The numbers of countries(subjects) included for analysing the correlations to other variables may differ as such.

The relevant United Nations agencies offer free online access to data required for the analyses in this study. There are no individual patients involved in the study. Therefore, there is no need to obtain the ethical approval or consent during our entire study process.

**Data multicollinearity check**

In order to avoid the inter-correlation between predictor variables, the multicollinearity statistics were calculated to test the correlations among the variables. Each variable was alternated as the dependent variable, and all the others were considered as the predictor variables in our analysis with the regression model. It turned out that collinearities between variables are insignificant since the tolerance of less than 0.20 and a VIF of more than 5 indicates a multicollinearity problem [44]. Details see the Additional File, AF 1 Collinearity among the variables.

**Data analysis**

To assess the population level determinants of OC56, the analysis proceeded in five steps.

1. Scatter plots were produced with the original data in Microsoft Excel® to explore and visualize the strength, shape and direction of correlations of OC56 to fertility and GDP respectively.

2. Bivariate (Pearson's r and nonparametric) correlations were performed to evaluate the direction and strength of the correlations between all the variables of all the subjects.

3. Partial correlation of Pearson's moment-product approach was performed to identify the strongest correlation and its independency. We alternated each of the six variables (ageing, fertility, GDP, $I_{bs}$, obesity and urbanization) as the independent predictor when all the five variables were included as the potential confounding factors.

Fisher's r-to-z transformation was conducted to assess the significance level of difference between correlation coefficients.

4. Standard multiple linear regression (Enter) was performed to describe the correlations between the dependent variable (fertility) and the predicting variables. In order to explore if low fertility can partially explain why ageing, GDP, $I_{bs}$, obesity and urbanization are correlated with OC56, the enter multiple linear regression was performed to calculate the correlations between OC56 incidence and the risk factors when fertility was incorporated and excluded as a predicting variable respectively.

Subsequently, standard multiple linear regression (Stepwise) was performed to select the predicting variable(s) which have the greatest influence on OC56 when fertility was incorporated and excluded as a predicting variable respectively.

5. The equations of the best fitting trendlines displayed in the scatter plots analysis of relationships between OC56 incidence and fertility ($y = 0.006x^2 - 0.504x + 14.816$, $R^2 = 0.485$) and GDP PPP ($y = 0.7167x + 0.2225$, $R^2 = 0.2571$) were used to calculate and remove the contributing effects of fertility and GDP PPP on OC56 incidence rate respectively. This allowed us to create of two new dependent variables, “Residual of OC56 standardised on fertility” and “Residual of OC56 standardised on GDP PPP”.

Means of the OC56 incidence rate, the “Residuals of OC56 standardised on fertility” and “Residual of OC56 standardised on GDP PPP” of all the countries were calculated for mean difference comparison.

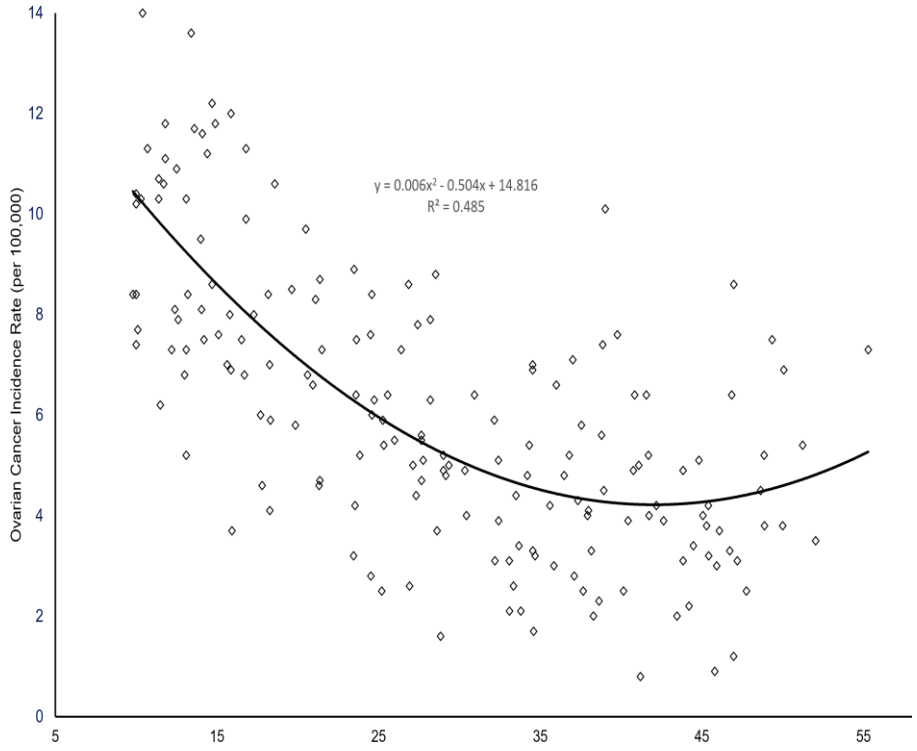
Analysis of variance (ANOVA) was conducted to detect the significant differences among the means of OC56 incidence rate, “Residual of OC56 standardised on fertility” and “Residual of OC56 standardised on GDP PPP” between the six WHO regions [45]. Further post-hoc (Bonferroni) tests were performed to identify the source of the significant difference and to compare the magnitude of the difference.

Bivariate correlations, multiple linear regression analysis (Enter and Stepwise) and ANOVA post hoc (Bonferroni) for mean calculation and comparison were conducted with SPSS v. 24. The raw data were used for mean calculation of OC56 incidence rate and “Residual of OC56 standardised on fertility” and “Residual of OC56 standardised on GDP PPP”. The variables were log transformed to increase homoscedasticity for the correlation analyses. The significance was kept at the 0.05 level, but 0.01 and 0.001 levels were also reported. Standard multiple linear regression analysis criteria were set at probability of F to enter ≤ 0.05 and probability of F to remove ≥ 0.10.

## Results

The relationship identified in the scatterplots between fertility and OC56 was noted to be polynomial with a strong, but inverse correlation ($R^2=0.485$, $p<0.001$, $n=179$, Figure 1).

Figure 1. The relationship between parity and ovarian cancer incidence rate




The strong relationship between fertility and OC56 identified in the scatterplots was confirmed by the subsequent Pearson r nonparametric analyses based on the log-transformed data.

Globally, fertility was significantly, but inversely correlated to OC56 incidence (r=-0.632 and rho= -0.655, $p<0.001$ respectively in Pearson and non-parametric analyses) (Table 1).

It is also revealed that ageing, GDP, $I_{bs}$, obesity and urbanization were also in strong and significant correlations to OC56 incidence in both Pearson and non-parametric analyses respectively (Table 1).

Table 1: Pearson r (above the diagonal) and nonparametric (below the diagonal) correlation between all variables

| | OC56 | Ageing | Birth rate | GDP | Ibs | Obesity | Urbanization |
|---|---|---|---|---|---|---|---|
| OC56 | 1 | 0.394*** | -0.632*** | 0.507*** | 0.455*** | 0.189* | 0.280*** |
| Ageing | 0.428*** | 1 | -0.737*** | 0.748*** | 0.766*** | 0.322*** | 0.570*** |
| Fertility | -0.655*** | -0.769*** | 1 | -0.772*** | -0.712*** | -0.338*** | -0.557*** |
| GDP PPP | 0.531*** | 0.759*** | -0.813*** | 1 | 0.742** | 0.485*** | 0.713*** |
| $I_{bs}$ | 0.602*** | 0.849*** | -0.883*** | 0.858*** | 1 | 0.457*** | 0.551*** |
| Obesity | 0.169* | 0.350*** | -0.377*** | 0.453*** | 0.409*** | 1 | 0.484*** |
| Urbanization | 0.345*** | 0.657*** | -0.628*** | 0.781*** | 0.711*** | 0.506*** | 1 |

The table describes the bivariate correlation between all the variables. *** p<0.001; Country number: 167-182.

Ovary cancer (OC56) incidence rate is from the International Agency for Research on Cancer. Birth rate indexing parity, GDP PPP and urbanization are from the World Bank. Ageing expressed as life expectancy ($e_{65}$) is from the United Nations. Obesity prevalence is the World Health Organization.

Biological State Index ($I_{bs}$) is downloaded from previous publication which were calculated with the data of the world fertility and life tables.

The relationship between OC56 and each independent variable (ageing, fertility, GDP, Ibs, obesity and urbanization) was tested by keeping the other five variables statistically constant in partial correlation analysis. Fertility was the only predictor showing a significant correlation (r= -0.448, p< 0.001) with OC56 independent of the other five variables (Table 2). Ageing showed significant, but weak correlation to OC56 (r= -0.178, p<0.05). The Fisher r-to-z transformation revealed that fertility correlated to OC56 significantly stronger than ageing (z=2.68, p<0.01). GDP, $I_{bs}$, obesity and urbanization showed the strong and significant correlation to OC56 in the bivariate correlation analyses respectively. However, none of them presented a strong or significant correlation with OC56 independent of the other five predictors. This indicates that fertility is the only strong and significant predictor of OC56 independent of the secondary association between OC56 incidence and $I_{bs}$ (magnitude of OC56 accumulation) and environmental factors (ageing, fertility, GDP, obesity and urbanization).

Table 2 Comparison of partial correlation coefficients between ovary cancer incidence and each variable when the other five variables are controlled for

| | Fertility | | | Ageing | | | GDP | | | $I_{bs}$ | | | Obesity | | | Urbanization | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variables | r | P | df | R | p | df | r | p | df | r | p | df | r | p | df | r | p | df |
| Fertility | -0.448 | <0.001 | 160 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| Ageing | - | - | - | -0.178 | 0.023 | 160 | - | - | - | - | - | - | - | - | - | - | - | - |
| GDP | - | - | - | - | - | - | 0.148 | 0.060 | 160 | - | - | - | - | - | - | - | - | - |
| $I_{bs}$ | - | - | - | - | - | - | - | - | - | 0.079 | 0.315 | 160 | - | - | - | - | - | - |
| Obesity | - | - | - | - | - | - | - | - | - | - | - | - | -0.048 | 0.544 | 160 | - | - | - |
| Urbanization | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | -0.131 | 0.095 | 160 |

The table describes the partial correlation between Ovary cancer (OC56) incidence between each variable while the other four variables are controlled for. - Controlled variable

Ovary cancer (OC56) incidence rate is from the International Agency for Research on Cancer. Fertility indexed by birth rate, GDP PPP and urbanization are from the World Bank. Ageing expressed as life expectancy ($e_{65}$) is from the United Nations. Obesity prevalence is the World Health Organization.

Biological State Index ($I_{bs}$) is downloaded from previous publication which were calculated with the data of the world fertility and life tables.

Standard multiple linear regression (enter) analysis was applied to further predict OC61 incidence when ageing, parity, GDP, obesity and urbanization were included as the independent predicting variables.

When fertility was excluded as one of the independent variables, GDP PPP (β=0.471, p<0.001) and $I_{bs}$ (β=0.250, p<0.05) were the two significant variables of OC61 incidence. However, when fertility was included as an independent predictor, only the correlation between fertility and OC61 incidence was strong and significant. None of the other five predictors showed strong and significant correlation to OC56 (Table 3). Similarly, in a stepwise linear regression model, when fertility was not included as one of the independent predictors, GDP and $I_{bs}$ were selected as the most influential variables on OC56. However, when fertility was included together with the other five independent variables, only fertility was selected as the most influential predictor of OC56 with the $R^2$ increase from 0.278 to 0.434. This suggested that GDP and $I_{bs}$ did not appear to account for the major part of the impact of OC56 incidence. This finding supports our previous suggestion that fertility is the significant predictor of OC56 incidence in partial correlation analysis.

Table 3 Independent predictors of ovarian cancer incidence rate based on multiple linear regression modelling

| Variable | β | Std. Error | Sig. | β | Std. Error | Sig. |
|---|---|---|---|---|---|---|
| Fertility | - | - | - | -0.694 | 0.111 | <0.001 |
| Ageing | -0.037 | 0.341 | 0.752 | -0.207 | 0.309 | 0.052 |
| GDP | 0.471 | 0.055 | <0.001 | 0.163 | 0.052 | 0.174 |
| $I_{bs}$ | 0.250 | 0.658 | 0.032 | 0.100 | 0.589 | 0.342 |
| Overweight | -0.056 | 0.069 | 0.496 | -0.020 | 0.060 | 0.778 |
| Urbanization | -0.122 | 0.105 | 0.211 | -0.125 | 0.092 | 0.146 |

The table describes the multiple linear regression analysis (Enter) results including and excluding birth rate as a predictor of breast cancer. *df* = 164; - excluded variable

Ovary cancer (OC56) incidence rate is from the International Agency for Research on Cancer. Fertility indexed by birth rate, GDP PPP and urbanization are from the World Bank. Ageing expressed as life expectancy ($e_{65}$) is from the United Nations. Obesity prevalence is the World Health Organization.

Biological State Index ($I_{bs}$) is downloaded from previous publication which were calculated with the data of the world fertility and life tables.

Table 4 showed that the mean OC56 incidence rate was lowest in Africa (4.19) and highest in Europe (8.70). The means of OC56 in the other four regions are Americas (5.89), Eastern Mediterranean (5.19), South East Asia (5.90) and Western Pacific (6.63). A post hoc Bonferroni analysis conducted on the multiple mean comparisons revealed that there were a number of significant mean differences in OC56 incidence rates between different WHO regions (Table 4). Mean of OC56 incidence in Europe was significantly greater than in Africa, Americas, East Mediterranean, South East Asia and West Pacific. Mean of OC56 in Americas was significantly greater than in Africa. The regions with greater means of fertility had lower means of OC56 incidence rates ($r=0.985$, $p<0.001$, $n=6$).

A subsequent ANOVA with post hoc Bonferroni procedure performed on the means of "Residual of OC56 standardised on fertility" in different WHO regions showed there was no significant difference among and between regions (Table 4). Whilst the same procedure was performed on the means of "Residual of OC56 standardised on GDP PPP", the developed region, Europe still had the significantly higher "Residual of OC56 standardised on GDP PPP" than Africa, Americas and East Mediterranean (Table 4). The results from the post hoc Bonferroni tests conducted on mean comparisons between the WHO regions suggested that regional variations of OC56 incidence may only reach statistically significant levels if the contributing effect of their respective fertility was included. In other words, except for fertility, the total contribution of the other OC56 risk factors to OC56 incidence may not be sufficient for the difference in mean rates to reach significance level. This result was supported by the findings identified in our previous partial correlation (Table 2) and multiple linear regression (Table 3) that fertility is the critical risk factor of OC56.

Table 4 Comparison of mean difference of fertility, Residuals of OC56 standardised on fertility and GDP PPP respectively between WHO regions

| OC56 incidence rate | | | Residual of OC56 standardised on fertility | | | Residual of OC56 standardised on GDP | | |
|---|---|---|---|---|---|---|---|---|
| I n Mean | J | Mean difference (I-J) | I n Mean | J | Mean difference (I-J) | I n Mean | J | Mean difference (I-J) |
| AF | AM | -1.70* | AF | AM | -0.16 | AF | AM | -0.03 |
| n=46 | EM | -0.99 | n=45 | EM | -0.29 | n=44 | EM | 1.03 |
| Mean=4.19 | EU | -4.50*** | Mean= -0.29 | EU | -0.30 | Mean= -0.15 | EU | -2.15*** |
| | SEA | -1.71 | | SEA | -0.67 | | SEA | -1.02 |
| | WP | -2.44** | | WP | -0.81 | | WP | -0.98 |
| AM, | AF | 1.69* | AM | AF | 0.16 | AM | AF | 0.03 |
| n=31 | EM | 0.70 | n=31 | EM | -0.13 | n=29 | EM | 1.06 |
| Mean= 5.89 | EU | -2.81*** | Mean= -0.13 | EU | -0.15 | Mean=-0.12 | EU | -2.12** |
| | SEA | -0.01 | | SEA | -0.52 | | SEA | -0.99 |
| | WP | -0.74 | | WP | -0.65 | | WP | -0.95 |
| EM | AF | 0.99 | EM | AF | 0.29 | EM | AF | -1.03 |
| n=22 | AM | -0.70 | n=21 | AM | 0.13 | n=18 | AM | -1.06 |
| Mean= 5.19 | EU | -3.51*** | Mean= 0.001 | EU | -0.01 | Mean= -1.18 | EU | -3.18*** |
| | SEA | -0.71 | | SEA | -0.38 | | SEA | -2.05 |
| | WP | -1.45 | | WP | -0.52 | | WP | -2.01 |
| EU | AF | 4.50*** | EU | AF | 0.30 | EU | AF | 2.15*** |
| n=50 | AM | 2.81*** | n=49 | AM | 0.15 | n=50 | AM | 2.12** |
| Mean=8.70 | EM | 3.51*** | Mean=0.14 | EM | 0.01 | Mean=2.00 | EM | 3.18*** |
| | SEA | 2.80* | | SEA | -0.37 | | SEA | 1.13 |
| | WP | 2.06* | | WP | -0.50 | | WP | 1.17 |
| SEA | AF | 1.71 | SEA | AF | 0.67 | SEA | AF | 1.02 |
| n=11 | AM | 0.01 | n=11 | AM | 0.52 | n=10 | AM | 0.99 |
| Mean= 5.90 | EM | 0.71 | Mean= 0.38 | EM | 0.38 | Mean= 0.87 | EM | 2.05 |

| OC56 incidence rate | | | Residual of OC56 standardised on fertility | | | Residual of OC56 standardised on GDP | | |
|---|---|---|---|---|---|---|---|---|
| In Mean | J | Mean difference (I-J) | In Mean | J | Mean difference (I-J) | In Mean | J | Mean difference (I-J) |
| | EU | -2.80* | | EU | 0.37 | | EU | -1.13 |
| | WP | -0.73 | | WP | -0.13 | | WP | 0.04 |
| WP | AF | 2.44** | WP | AF | 0.81 | WP | AF | 0.98 |
| n= 22 | AM | 0.74 | n=21 | AM | 0.65 | n= 19 | AM | 0.95 |
| Mean=6.63 | EM | 1.45 | Mean=-0.01 | EM | 0.52 | Mean=0.83 | EM | 2.01 |
| | EU | -2.06* | | EU | 0.50 | | EU | -1.17 |
| | SEA | 0.73 | | SEA | 0.13 | | SEA | -0.04 |

The mean difference comparison results conducted with One-way ANOVA Post hoc Bonferroni are reported. * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Ovary cancer (OC56) incidence rate is from the International Agency for Research on Cancer. Fertility indexed by birth rate, GDP PPP and urbanization are from the World Bank. Ageing expressed as life expectancy ($e_{65}$) is from the United Nations. Obesity prevalence is the World Health Organization.

Biological State Index ($I_{bs}$) is downloaded from previous publication which were calculated with the data of the world fertility and life tables.

## Discussion

The present ecological study suggests that:

1. Low fertility, not only ageing, GDP, $I_{bs}$, obesity and urbanization, may be a significant determinant of OC56 incidence.

2. The low fertility has a statistically significant effect on OC56 that is independent of the effect of ageing, GDP, $I_{bs}$, obesity and urbanization.

3. Once the effects of low fertility are taken into account, the correlations between OC56 and ageing, GDP, $I_{bs}$, obesity and urbanization statistically disappear, indicating that low fertility may be an explanation for why ageing, GDP, $I_{bs}$, obesity and urbanization have been correlated with OC56.

4. Statistically, outscoring ageing, GDP, $I_{bs}$, obesity and urbanization, low fertility may be a significant risk factor of OC56.

The relationship between female reproductive behaviour and gynecological cancers has been under research for over 300 years [25, 47]. In general, extensive studies of the strong protective effect on OC56 in Asian, African, American and European populations have concluded that nulliparous women may have a 30%-60% higher risk than parous women [48] [49] . Studies also reported that each additional full-term pregnancy lowers OC56 risk by approximately 15% [49] [50].

Full-term pregnancy and the subsequent lactation cause anovulation and suppresses secretion of pituitary gonadotropins, which may make women produce less oestrogen or less menstrual cycles [51]. This formed the three prevailing hypotheses to explain the relationship between fertility and OC56 risk: 1) The incessant-ovulation hypothesis postulates that pregnancies reduce the number of times a woman ovulates in her life and thus the chance for mutation to occur during the repair of ruptured epithelial tissue is reduced [52] [53] [54]. 2) The inflammation hypothesis implicates that epithelial cells may be exposed to less chronic inflammation and mutation due to less ovulations [55]. 3) The pituitary/gonadotropin hypothesis considers pregnancy may prevent gonadotropins from being overstimulated, which may reduce the proliferation of malignant transformation in the inclusion cysts and clefts invaginated and formed in the ovarian epithelium [56].

A self-reinforcing cycle may be formed between more and positive family member interactions and oxytocin production, which may protect female from OC56 initiation. Greater fertility not only provides the direct physiological protective effects from developing OC56, but also offers the heathy benefit to females as it protects females through reacting to positive psychological well-being. Oxytocin is a peptide hormone and neuropeptide. Its production is associated with good feelings and emotions [57]. Recent

studies have reported that oxytocin not only helps with birth, bonding with the baby, and milk production, but also inhibits the progression of human ovarian carcinoma cells [23, 24] and shows the therapeutic effects on other cancers [23]. Researches constantly reported that family related activities are the major promotors of oxytocin production. Bigger family size due to greater fertility may offer more positive psychological feeling through more daily interactions between family members [58, 59], such as spouses [60-62], mother and children [63], and father and children [64]. Positive psychological feeling from greater family size may also be able to bond couples to keep monogamous [65, 66].

Studies also reported that positive psychological well-being may make the functions of neuroendocrine and immune systems more efficient, which may also reduce the risk of developing OC56 [67-70].

Family member from bigger family may be reminded and/or recommended more by other members to have necessary medical examination and have a healthy lifestyle [71]. A number of studies have shown that greater family size may decrease breast cancer cells, from initiating and proliferating [72-74].

Our finding showed that fertility was negatively correlated to OC56 incidence significantly stronger than other factors (ageing, GDP, $I_{bs}$, obesity and urbanization). This may be because fertility affects OC56 risk from both physiological and psychological perspectives, but ageing, GDP, $I_{bs}$, obesity and urbanization may not be able to offer that much protection.

The strength of this study is that it suggests that fertility is a significant determinant of OC56 risk, despite that we used a new approach to test the fertility has the protective role in female OC56 initiation at population level. This finding is in agreement with three studies conducted by Hankinson et al. [27], Vachon et al [26] and Cramer et al [48] respectively which concluded that, outscoring other risk factors, fertility is a significant predictor of OC56.

Our study indicates that when the contributing effect of the fertility is not incorporated as the risk factor of OC56, the difference of regional variations of OC56 between the six WHO regions does not reach significant level. This supports our hypothesis that low fertility is a significant risk factor of OC56, but it does not support the WHO and IARC's statement that OC56 incidence rate is associated with regions with different socioeconomic level [2, 29] [7]. Although GDP and fertility, the significant risk predictor of OC56, are highly correlated, it is still debated whether industrialization and higher incomes lead to lower fertility, or whether lower fertility leads to industrialization and higher incomes [77] [78] [79]. Interestingly, studies have shown that fertility increases when the socioeconomic development is beyond some level [77].

This study has several limitations: 1) Each country was considered as a whole subject for the ecological study. The country-specific data included in this study may be different from the data collected from individual participants. Therefore, the correlations identified from the data analysis may not hold true for all the individuals to have the risk in OC56 development. 2) There may be some random errors when the United Nations and its agencies collected and aggregated data at country level. Data from developed countries may be more complete than those from developing countries. 4) There are different categories of OC56, but we could not differentiate them for the correlation analysis due to the unavailability of such data.

## Conclusion

Low fertility may be a significant and strong determinant of OC56 risk independent of ageing, GDP, $I_{bs}$, obesity and urbanization. The effects of low fertility on predicting OC56 are outscoring ageing, GDP, $I_{bs}$, obesity and urbanization. Our finding may be helpful for governments, policy-makers, funders and researchers to track and investigate the priorities of OC56 research when they allocate the funds and resources to meet the research priorities. For example, while ageing, and socioeconomic factors (urbanization and GDP) and its associated factors (obesity and $I_{bs}$), and stable fertility (birth rate) worldwide cannot be modified, can the health authorities prioritize the research to reduce the number of ovulation cycles of females [27, 75,76]? Considering the main contributing effects of fertility on ovarian cancer have not been fully investigated, currently, can the studies of gynaecological cancers prioritize the improvement and diversification of those ovulation-inhibiting contraceptive approaches for females who choose to be nulliparous or not to have more children?

### Authors' contributions

WY conceived the hypothesis and IS and MH consolidated the hypothesis. WY and MH conducted data analysis. All authors interpreted the data analysis results. IS and MH provided suggestions for WY to draft the manuscript. All authors reviewed, edited and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

The authors express appreciation to Jacques Ferlay from the International Agency for Research on Cancer of World Health Organization for his assistance in locating and defining the data.

## References


1. International Agency for Research on Cancer. *Cancer*. 2017 22 December 2017]; Available from: http://globocan.iarc.fr/Pages/cancer.aspx.
2. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* Int J Cancer, 2015. **136**(5): p. E359-86.
3. Ferlay, J., et al., *Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008.* Int J Cancer, 2010. **127**(12): p. 2893-917.
4. Stewart, B.W., *World Cancer Report 2014*. 2014, Lyon: Lyon, FRA: International Agency for Research on Cancer.
5. Vos, T., et al., *Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015.* The Lancet, 2016. **388**(10053): p. 1545-1602.
6. Seiden MV, *Gynecologic Malignancies*, in *Principles of Internal Medicine (18th ed.)*, K.D. Longo DL, Jameson JL, Fauci AS, Hauser SL, Loscalzo J. Harrison, Editor. 2012, McGraw-Hill.
7. Reid, B.M., J.B. Permuth, and T.A. Sellers, *Epidemiology of ovarian cancer: a review.* Cancer Biology & Medicine, 2017. **14**(1): p. 9.
8. Vargas, A.N., *Natural history of ovarian cancer.* ecancermedicalscience, 2014. **8**.
9. The American Cancer Society. *Do We Know What Causes Ovarian Cancer?* 2017; Available from: https://www.cancer.org/cancer/ovarian-cancer/causes-risks-prevention/what-causes.html.
10. Australian Government- Cancer Australia. *What are the risk factors for ovarian cancer?* 2017; Available from: https://ovarian-cancer.canceraustralia.gov.au/risk-factors.
11. Rodriguez, C., et al., *Body mass index, height, and the risk of ovarian cancer mortality in a prospective cohort of postmenopausal women.* Cancer Epidemiology and Prevention Biomarkers, 2002. **11**(9): p. 822-828.
12. Cauley, J.A., et al., *The epidemiology of serum sex hormones in postmenopausal women.* American journal of epidemiology, 1989. **129**(6): p. 1120-1131.
13. Allender, S., et al., *Quantification of urbanization in relation to chronic diseases in developing countries: a systematic review.* J Urban Health, 2008. **85**(6): p. 938-951.
14. Huang, Z., et al., *Incidence and mortality of gynaecological cancers: Secular trends in urban Shanghai, China over 40 years.* European Journal of Cancer, 2016. **63**: p. 1-10.
15. Zhang, X., S.V. Nicosia, and W. Bai, *Vitamin D receptor is a novel drug target for ovarian cancer treatment.* Current cancer drug targets, 2006. **6**(3): p. 229-244.
16. Jordan, S.J., et al., *Does smoking increase risk of ovarian cancer? A systematic review.* Gynecologic oncology, 2006. **103**(3): p. 1122-1129.
17. You, W. and H. M, *Cancer incidence increasing globally: The role of relaxed natural selection.* Evol Appl., 2017. **00:1–13**.
18. You, W.-P. and M. Henneberg, *Type 1 diabetes prevalence increasing globally and regionally: the role of natural selection and life expectancy at birth.* BMJ Open Diabetes Research & Care, 2016. **4**(1): p. e000161.
19. Budnik, A. and M. Henneberg, *Worldwide increase of obesity is related to the reduced opportunity for natural selection.* PloS one, 2017. **12**(1): p. e0170098.

20. Cancer Research UK. *Ovarian cancer risks and causes*. [Document] 2016 28/07/2016 13:29 20.08.2016]; Available from: http://www.cancerresearchuk.org.

21. McLemore, M.R., et al., *Epidemiologic and genetic factors associated with ovarian cancer.* Cancer nursing, 2009. **32**(4): p. 281.

22. Fraumeni, J.F., et al., *Cancer mortality among nuns: role of marital status in etiology of neoplastic disease in women.* Journal of the National Cancer Institute, 1969. **42**(3): p. 455-468.

23. Imanieh, M.H., et al., *Oxytocin has therapeutic effects on cancer, a hypothesis.* European Journal of Pharmacology, 2014. **741**: p. 112-123.

24. Morita, T., et al., *Oxytocin inhibits the progression of human ovarian carcinoma cells in vitro and in vivo.* International journal of cancer, 2004. **109**(4): p. 525-532.

25. Britt, K. and R. Short, *The plight of nuns: hazards of nulliparity.* The Lancet, 2012. **379**(9834): p. 2322-2323.

26. Vachon, C.M., et al., *Association of parity and ovarian cancer risk by family history of breast or ovarian cancer in a population-based study of postmenopausal women.* Epidemiology, 2002. **13**(1): p. 66-71.

27. Hankinson, S.E., et al., *A prospective study of reproductive factors and risk of epithelial ovarian cancer.* Cancer, 1995. **76**(2): p. 284-290.

28. Jayson, G.C., et al., *Ovarian cancer.* The Lancet, 2014. **384**(9951): p. 1376-1388.

29. Ferlay, J., et al. *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]*. 2013 28.05.2016]; Available from: http://globocan.iarc.fr.

30. Jemal, A., et al., *Global cancer statistics.* CA Cancer J Clin, 2011. **61**(2): p. 69-90.

31. Wei, K., et al., *Ovary cancer incidence and mortality in China, 2011.* Chinese Journal of Cancer Research, 2015. **27**(1): p. 38.

32. Horner, M., et al., *SEER Cancer Statistics Review, 1975-2006, National Cancer Institute. Bethesda, MD*. 2009.

33. The World Bank Group. *World Bank Open Data*. 2016 12.07.2016]; Available from: http://data.worldbank.org/.

34. Ness, R.B., et al., *Racial differences in ovarian cancer risk.* Journal of the National Medical Association, 2000. **92**(4): p. 176.

35. Bertone-Johnson, E.R., *Epidemiology of ovarian cancer: a status report.* The Lancet, 2005. **365**(9454): p. 101-102.

36. Jin, F., et al., *Incidence trends for cancers of the breast, ovary, and corpus uteri in urban Shanghai, 1972–89.* Cancer Causes & Control, 1993. **4**(4): p. 355-360.

37. LEFKOWITZ, E.S. and C.F. Garland, *Sunlight, vitamin D, and ovarian cancer mortality rates in US women.* International Journal of Epidemiology, 1994. **23**(6): p. 1133-1136.

38. Moore, M., P. Gould, and B.S. Keary, *Global urbanization and impact on health.* International journal of hygiene and environmental health, 2003. **206**(4): p. 269-278.

39. WHO. *Urbanization and health*. WHO 2010 2010-12-07 15:20:05 2 November 2016]; Available from: http://www.who.int/bulletin/volumes/88/4/10-010410/en/.

40. United Nations-Department of Economic and Social Affairs-Population Division. *Abridged life table, for females, by major area, region and country, 1950-2100*. 2013 18.07.2016].

41. John, E.M., et al., *Characteristics relating to ovarian cancer risk: collaborative analysis of seven US case-control studies. Epithelial ovarian cancer in black women.* Journal of the National Cancer Institute, 1993. **85**(2): p. 142-147.

42. Russo, A., et al., *Hereditary ovarian cancer.* Critical reviews in oncology/hematology, 2009. **69**(1): p. 28-44.

43. Engeland, A., S. Tretli, and T. Bjørge, *Height, body mass index, and ovarian cancer: a follow-up of 1.1 million Norwegian women.* Journal of the National Cancer Institute, 2003. **95**(16): p. 1244-1248.

44. WHO. *Global Health Observatory, the data repository*. WHO 2015 [11.26.2015]; Available from: http://www.who.int/gho/database/en/.

45. O'brien, R.M., *A caution regarding rules of thumb for variance inflation factors.* Quality & Quantity, 2007. **41**(5): p. 673-690.

46. WHO, *Global Health Risks Mortality and Burden of Disease Attributable to Selected Major Risks*. 2009, Geneva: Geneva : World Health Organization.

47. Ramazzini, B.a., *De morbis artificum Bernardini Ramazzini diatriba = Diseases of workers : the Latin text of 1713 revised, with translation and notes / by Wilmer Cave Wright*. Special edition / privately printed for the members of the Classics of Medicine Library. ed. Diseases of workers, ed. W.C.e.t. Wright and i.b. Classics of Medicine Library. 1983: Birmingham, Alabama : The Classics of Medicine Library.

48. Cramer, D.W., et al., *Determinants of Ovarian Cancer Risk. I. Reproductive Experiences and Family History23.* JNCI, Journal of the National Cancer Institute, 1983. **71**: p. 711.

49. Moorman, P.G., et al., *Reproductive factors and ovarian cancer risk in African-American women.* Annals of epidemiology, 2016. **26**(9): p. 654-662.

50. Adami, H.-O., et al., *Parity, age at first childbirth, and risk of ovarian cancer.* The Lancet, 1994. **344**(8932): p. 1250-1254.

51. WHO, *Bulletin of the World Health Organization: The breast cancer conundrum*, in *Bull World Health Organ*. 2013. p. 626–627.

52. Riman, T., S. Nilsson, and I.R. Persson, *Review of epidemiological evidence for reproductive and hormonal factors in relation to the risk of epithelial ovarian malignancies.* Acta obstetricia et gynecologica Scandinavica, 2004. **83**(9): p. 783-795.

53. Casagrande, J., et al., *" Incessant ovulation" and ovarian cancer.* The Lancet, 1979. **314**(8135): p. 170-173.

54. Fathalla, M., *Incessant ovulation and ovarian cancer–a hypothesis re-visited.* Facts, views & vision in ObGyn, 2013. **5**(4): p. 292.

55. Ness, R.B., et al., *Factors related to inflammation of the ovarian epithelium and risk of ovarian cancer.* Epidemiology, 2000. **11**(2): p. 111-117.

56. Choi, J.-H., et al., *Gonadotropins and ovarian cancer.* Endocrine reviews, 2007. **28**(4): p. 440-461.

57. Magon, N. and S. Kalra, *The orgasmic history of oxytocin: Love, lust, and labor.* Indian journal of endocrinology and metabolism, 2011. **15**(Suppl3): p. S156.

58. Angeles, L., *Children and Life Satisfaction.* Journal of Happiness Studies, 2009. **11**(4): p. 523-538.

59. Nan, H., et al., *Psychometric evaluation of the Chinese version of the Subjective Happiness Scale: evidence from the Hong Kong FAMILY Cohort.* Int J Behav Med, 2014. **21**(4): p. 646-52.

60. Carmichael, M.S., et al., *Plasma oxytocin increases in the human sexual response.* The Journal of Clinical Endocrinology & Metabolism, 1987. **64**(1): p. 27-31.

61. Carmichael, M.S., et al., *Relationships among cardiovascular, muscular, and oxytocin responses during human sexual activity.* Archives of sexual behavior, 1994. **23**(1): p. 59-79.

62. Gordon Jr, G., R.L. Burch, and S.M. Platek, *Does semen have antidepressant properties?* Archives of Sexual Behavior, 2002. **31**(3): p. 289-293.

63. Kendrick, K.M., *The neurobiology of social bonds.* Journal of neuroendocrinology, 2004. **16**(12): p. 1007-1008.

64. Weisman, O., O. Zagoory-Sharon, and R. Feldman, *Oxytocin administration to parent enhances infant physiological and behavioral readiness for social engagement.* Biological psychiatry, 2012. **72**(12): p. 982-989.

65. Insel, T.R. and T.J. Hulihan, *A gender-specific mechanism for pair bonding: oxytocin and partner preference formation in monogamous voles.* Behavioral neuroscience, 1995. **109**(4): p. 782.

66. Young, L.J., A.Z. Murphy Young, and E.A. Hammock, *Anatomy and neurochemistry of the pair bond.* Journal of Comparative Neurology, 2005. **493**(1): p. 51-57.

67. Antonova, L. and C.R. Mueller, *Hydrocortisone down-regulates the tumor suppressor gene BRCA1 in mammary cells: a possible molecular link between stress and breast cancer.* Genes Chromosomes Cancer, 2008. **47**(4): p. 341-52.

68. Cohen, S. and M.S. Rodriguez, *Pathways Linking Affective Disturbances and Physical Disorders.* Health Psychology, 1995. **14**(5): p. 374-380.

69. Diener, E. and M.Y. Chan, *Happy People Live Longer: Subjective Well-Being Contributes to Health and Longevity.* Applied Psychology: Health and Well-Being, 2011. **3**(1): p. 1-43.

70. Williams, R.B. and N. Schneiderman, *Resolved: Psychosocial interventions can improve clinical outcomes in organic disease (pro).* Psychosomatic Medicine, 2002. **64**: p. 552–557.

71. Aizer, A.A., et al., *Marital status and survival in patients with cancer.* J Clin Oncol, 2013. **31**(31): p. 3869-76.

72. Kim, Y., et al., *Psychological distress among healthy women with family histories of breast cancer: effects of recent life events.* Psychooncology, 2005. **14**(7): p. 555-63.

73. Peled, R., et al., *Breast cancer, psychological distress and life events among young women.* BMC Cancer, 2008. **8**: p. 245.

74. Bai, A., et al., *A survey of overall life satisfaction and its association with breast diseases in Chinese women.* Cancer Med, 2016. **5**(1): p. 111-9.

75. Whitaker, L., *The plight of nuns: hazards of nulliparity.* Journal of Family Planning and Reproductive Health Care, 2012. **38**(2): p. 116-116.

76. Kent, A., *Nuns and Contraceptives.* Reviews in Obstetrics and Gynecology, 2012. **5**(3-4): p. e166.

77. Myrskylä, M., H.-P. Kohler, and F.C. Billari, *Advances in development reverse fertility declines.* Nature, 2009. **460**(7256): p. 741.

78. Galor, O., *The demographic transition: causes and consequences.* Cliometrica, 2012. **6**(1): p. 1-28.

79. Sinding, S.W., *Population, poverty and economic development.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2009. **364**(1532): p. 3023-3030.

## Supplemental File

Supplemental File 1

Table AF1 Collinearity among the variables

Coefficients[a]

| Model | | Collinearity Statistics | |
|---|---|---|---|
| | | Tolerance | VIF |
| 1 | Ageing e(65) UN LN | .305 | 3.276 |
| | Birth rate 1992 LN | .337 | 2.967 |
| | GDP PPP 2012 LN | .240 | 4.169 |
| | Ibs LN | .313 | 3.192 |
| | Obesity asr (%) ln | .654 | 1.529 |
| | Urban 2012 LN | .465 | 2.149 |

a. Dependent Variable: OCa rate asr LN

Coefficients[a]

| Model | | Collinearity Statistics | |
|---|---|---|---|
| | | Tolerance | VIF |
| 1 | Birth rate 1992 LN | .283 | 3.530 |
| | GDP PPP 2012 LN | .251 | 3.991 |
| | Ibs LN | .388 | 2.578 |
| | Obesity asr (%) ln | .677 | 1.477 |
| | Urban 2012 LN | .461 | 2.167 |
| | OCa rate asr LN | .553 | 1.810 |

a. Dependent Variable: Ageing e(65) UN LN

Coefficients[a]

| Model | | Collinearity Statistics | |
|---|---|---|---|
| | | Tolerance | VIF |
| 1 | GDP PPP 2012 LN | .257 | 3.892 |
| | Ibs LN | .317 | 3.157 |
| | Obesity asr (%) ln | .656 | 1.526 |
| | Urban 2012 LN | .461 | 2.171 |
| | OCa rate asr LN | .702 | 1.425 |
| | Ageing e(65) UN LN | .326 | 3.069 |

a. Dependent Variable: Birth rate 1992 LN

Coefficients[a]

| Model | | Collinearity Statistics | |
|---|---|---|---|
| | | Tolerance | VIF |
| 1 | lbs LN | .321 | 3.112 |
| | Obesity asr (%) ln | .676 | 1.479 |
| | Urban 2012 LN | .544 | 1.839 |
| | OCa rate asr LN | .546 | 1.832 |
| | Ageing e(65) UN LN | .315 | 3.174 |
| | Birth rate 1992 LN | .281 | 3.561 |

a. Dependent Variable: GDP PPP 2012 LN

Coefficients[a]

| Model | | Collinearity Statistics | |
|---|---|---|---|
| | | Tolerance | VIF |
| 1 | Obesity asr (%) ln | .693 | 1.442 |
| | Urban 2012 LN | .459 | 2.177 |
| | OCa rate asr LN | .542 | 1.843 |
| | Ageing e(65) UN LN | .371 | 2.695 |
| | Birth rate 1992 LN | .263 | 3.795 |
| | GDP PPP 2012 LN | .245 | 4.089 |

a. Dependent Variable: lbs LN

Coefficients[a]

| Model | | Collinearity Statistics | |
|---|---|---|---|
| | | Tolerance | VIF |
| 1 | Urban 2012 LN | .493 | 2.030 |
| | OCa rate asr LN | .540 | 1.853 |
| | Ageing e(65) UN LN | .309 | 3.240 |
| | Birth rate 1992 LN | .260 | 3.849 |
| | GDP PPP 2012 LN | .245 | 4.080 |
| | lbs LN | .330 | 3.028 |

a. Dependent Variable: Obesity asr (%) ln

Coefficients[a]

| Model | | Collinearity Statistics | |
|---|---|---|---|
| | | Tolerance | VIF |
| 1 | OCa rate asr LN | .547 | 1.829 |
| | Ageing e(65) UN LN | .300 | 3.338 |
| | Birth rate 1992 LN | .260 | 3.846 |
| | GDP PPP 2012 LN | .281 | 3.561 |
| | lbs LN | .312 | 3.207 |
| | Obesity asr (%) ln | .702 | 1.425 |

a. Dependent Variable: Urban 2012 LN