PUBLISHED VERSION

Plutzer, Kamila; Mejia, Gloria Cecilia; Spencer, Andrew John; Keirse, Marc J. N. C.
Dealing with missing outcomes: Lessons from a randomized trial of a prenatal intervention to prevent early childhood caries, The Open Dentistry Journal, 2010; 4:55-60

PERMISSIONS

http://benthamscience.com/open/toaddj/MSandI.htm

21st December 2012

http://hdl.handle.net/2440/63259

Dealing with Missing Outcomes: Lessons from a Randomized Trial of a Prenatal Intervention to Prevent Early Childhood Caries

Kamila Plutzer[a,*], Gloria C. Mejia[a,b], A. John Spencer[a] and Marc J. N. C. Keirse[c]

[a]*Australian Research Centre for Population Oral Health, School of Dentistry, The University of Adelaide, Adelaide, South Australia*

[b]*Department of Preventive and Restorative Dental Sciences, School of Dentistry, The University of California, San Francisco, USA*

[c]*Department of Obstetrics, Gynaecology and Reproductive Medicine, Flinders University, Adelaide, South Australia*

Abstract: Severe early childhood caries (S-ECC) affects 17% of 2-3 year old children in South Australia impacting on their general health and well-being. S-ECC is largely preventable by providing mothers with anticipatory guidance. Randomised controlled trials (RCTs) are the most decisive way to test this, but that approach suffers from near inevitable loss to follow-up that occurs with preventative strategies and distant outcome assessment.

We re-examined the results of an RCT to prevent S-ECC using sensitivity analyses and multiple imputation to test different assumptions about violation of random allocation (1%) and major loss to follow-up (32%). Irrespective of any assumptions about missing outcomes, providing expectant mothers with anticipatory guidance during pregnancy and in the child's first year of life, significantly reduced the incidence of S-ECC at 20 months of age. However, the relative risk of S-ECC varied from 0.18 (95% confidence interval (CI): 0.06 – 0.52) to 0.70 (95% CI: 0.56 – 0.88). Also the 'number needed to treat' (NNT) to prevent one case of S-ECC varied 2.5-fold: from 8 to 20 women given anticipatory guidance. Multiple imputation provided a best estimate of 0.25 (95% CI: 0.11 – 0.56) for the relative risk and of 14 (95% CI: 10 – 33) for the number needed to treat.

Avoiding loss to follow-up is crucial in any RCT, but is difficult with preventative health care strategies. Instead of abandoning randomisation in such circumstances, sensitivity analyses and multiple imputation can consolidate the findings of an RCT and add extra value to the conclusions derived from it.

INTRODUCTION

Severe early childhood caries (S-ECC) is an important public health problem. The US National Institutes of Health defined S-ECC as the presence of one or more missing, filled or decayed tooth surfaces, whether cavitated or not, in a child below 3 years of age [1]. If left untreated, it severely affects a child's general health, learning ability and quality of life and can have a life-long impact on oral health [2, 3]. Children who experience caries as infants or toddlers have a much greater probability of further caries in their primary and secondary dentitions [3]. In the 1990s, the prevalence of S-ECC among 2-3 years olds in South Australia was 16.7% [4]. As such children are entirely dependent on the care provided by others, mothers and other childcarers need oral health programs to support them in preventing S-ECC, while communities need to know what does and does not work in S-ECC prevention [5]. To address this issue we conducted a randomised controlled trial (RCT) of anticipatory guidance [6] targeting first time mothers [7].

While RCTs are generally considered as the "gold standard" for obtaining solid evidence on the effect of health care interventions, some have questioned their applicability to complex interventions [8] and to health promotion strategies in particular [9-11]. The latter tend to be complex and require a long follow-up that may exceed the attention span of participants, who do not have a problem in the first place. Attrition bias (literally referring to 'wearing out') as people loose interest or are lost for one reason or another, may defeat the whole object of random allocation. This becomes especially problematic when retention rates and loss to follow-up differ substantially between the randomised groups [12-14].

Another, more general problem arises from people's perception that what is new must be better, thereby limiting their willingness to be randomly assigned to what is perceived as the lesser of the strategies compared. Limited participation will inevitably affect the generalisability of the findings to people outside the actual participants. Zelen [15] described an approach to address this. In it, potential participants are randomised to the proposed study arms before consent is sought and consent is then sought only from those assigned to the experimental group. This design has been

*Address correspondence to this author at the Australian Research Centre for Population Oral Health, School of Dentistry, The University of Adelaide, SA 5005, Australia; Tel: +61 8 8303 3292; Fax: +61 8 8303 4858; E-mail: kamila.plutzer@adelaide.edu.au

criticized by ethicists on several grounds, mostly relating to people's right to know and to the fallacy of assuming that people in the control group would also have consented, if they had been aware of the alternative option [16-18].

These objections are largely overcome in a Zelen design with double consent [19]. In the double consent design, potential participants are still randomised before consent. Thereafter, they are informed about their group allocation, given full information on the alternative interventions and offered the choice of either accepting or rejecting the random allocation and of moving to the opposite group, if they choose to do so. As the most unbiased assessment is between the groups as allocated [19], irrespective of whether they received the intervention or not, based on the "intention-to-treat principle" (ITT) [20], there are two major drawbacks to this approach. The first is a loss of statistical power, if many participants refuse the allocation, and the second is a loss of difference between the groups as randomised, when participants move from one group to the other.

In this paper, we address the consequences of the double consent form, but especially the attrition bias inherent in studies with prolonged follow-up, using sensitivity analysis and multiple imputation. To do so, we have used the data of our RCT "Cavity free children" in which the intervention targeted pregnant women expecting their first child, while the outcome was the oral health of their child at 20 months of age [7]. We address two questions that are relevant to similar studies conducted elsewhere. First, to what extent did the change of subjects from their allocated group influence the results? Second, how would the results change, if all enrolled subjects had participated until the end of the study without loss to follow-up?

MATERIALS AND METHODS

In 2002, we started an RCT to assess an oral health promotion intervention "Cavity free children", targeting pregnant women expecting their first child. Recruitment occurred with ethical approval at all five public maternity units in Adelaide during routine antenatal visits [7]. Zelen's design with double consent [19] was used to allocate women to the intervention and control groups. The intervention consisted of providing expectant mothers with printed information on oral health and early childhood caries, which was reinforced with additional information, sent by mail when the child was 6 months and 12 months old, focusing on the child's needs at that age [6]. Outcome assessment was the presence of any carious lesion on the upper incisors, assessed by dental examination at 20 months of age. The trial resulted in a significant difference in the occurrence of S-ECC between the intervention and control infants (1.7% versus 9.6%; p <0.01) [7].

However, of 649 women allocated to the intervention and control groups, only 441 (68%) brought their child for a dental examination to assess outcome at 20 months of age [7]. Twenty-four had been legitimately excluded for pre-specified reasons, unrelated to their allocation and all due to legitimate absence of a mother-infant pair because of miscarriage, stillbirth, infant death, congenital malformations, child in the custody of Child and Youth Services, or mother in custody. Fifteen were in the intervention group and nine in

the control group. A further 184 participants were lost to follow-up: 32 had moved to a distant locality, 66 were not traceable, and 86 failed to attend for dental examination for reasons, such as a new pregnancy, conditions of employment, problems with the child's health or with transport, or declining further participation.

The comparability of the groups was evaluated at baseline and again for those attending dental examinations to assess whether loss to follow-up had distorted the comparability of groups. Baseline characteristics assessed were various demographic factors (e.g. age, ethnicity, marital status, occupation, level of education) and women's perception of their general and dental health and oral hygiene [7]. The same characteristics and the child's age at tooth eruption were compared for those completing follow-up. There were no statistically significant differences between the intervention and control groups at baseline and at follow-up. While this suggests that the difference in S-ECC between the intervention and control groups is entirely attributable to the effect of the intervention, many would consider a loss to follow-up of 32% unacceptably high for an RCT [14]. We, therefore, conducted sensitivity analyses and multiple imputation as an approach to consolidate its findings.

Sensitivity analyses basically assess how robust the results are to changes in how the study was done. In other words, to what extent do different assumptions about the people, who were randomised but not included in the outcome assessment, affect the outcome. To do so, we tested the assumptions that those lost to follow-up would all have had S-ECC, no S-ECC, or the same incidence of S-ECC as those retained in the intervention group, the control group or overall.

We also conducted multivariable imputation, a simulation-based approach originally designed for complex surveys, in which several important data may be missing for variable proportions of participants [21, 22]. It addresses the fact that analyses including only participants with complete data may not yield valid inferences about the entire study population [22]. It is based on the statistical principle that every subject in a random sample can be replaced by another subject, randomly drawn from the same sample as the original subject, without compromising the conclusions. Missing values for each subject are filled in (imputed) with values predicted from the rest of the subject's known characteristics. A random component is added to the imputed value to account for the uncertainty due to the imputed value not being observed, but estimated. This is done by creating not one but several (multiple) imputed data sets, which are analysed separately and later combined to provide estimates and covariance matrices [23]. Multivariable techniques, employing a sequence of regression models, are used for imputing missing values [23].

In our analysis, the estimates were calculated using IVEware [23]. Outcome data were imputed for each participant lost to follow-up, conditional on the known variables for that participant, considering imputation on a variable by variable basis [23]. The basic strategy was to create imputations through sequences of multiple regressions, each time overwriting previously drawn values, building interdependence among imputed values [23, 24]. In total, a series of five im-

Table 1. Effect of Cross-Over Between the Intervention and the Control Group (Zelen's Double Consent Design) and Loss to Follow-Up on the Effectiveness of an Early Childhood Caries Prevention Programme

Type of Analysis	S-ECC Observed (n/N) *		Outcome Parameters (95% Confidence Interval)					
	Intervention	Control	Relative Risk		Risk Difference		NNT **	
Analysis for all participants recruited irrespective of their pregnancy outcome and presence or absence of a mother-infant pair								
Primary intention to treat analysis	4/322	20/327	0.20	(0.07 - 0.59)	0.05	(0.02 - 0.08)	20	(13 - 50)
Analysis for all mother-infant pairs in the study								
Intention to treat analysis	4/307	20/318	0.21	(0.07 - 0.60)	0.05	(0.02 - 0.08)	20	(13 - 50)
Analysis per treatment received	4/312	20/313	0.20	(0.07 - 0.58)	0.05	(0.02 - 0.08)	20	(13 - 50)
Analysis for mother-infant pairs with outcome assessment								
Analyses per group allocation	4/227	20/214	0.19	(0.36 - 0.52)	0.09	(0.03 - 0.12)	11	(8 - 33)
Analysis per treatment received	4/232	20/209	0.18	(0.06 - 0.52)	0.08	(0.04 - 0.12)	13	(8 - 25)

* S-ECC = severe early childhood caries
** Number Needed to Treat to prevent one case of S-ECC. It is calculated as (1/Risk Difference)

Table 2. Effect of Different Assumptions About Missing Outcome Data on the Effectiveness of an Early Childhood Caries Prevention Programme

Assumptions About Missing Outcomes	Number with S-ECC*		Outcome Parameters (95% Confidence Interval)					
	Intervention	Control	Relative Risk		Risk Difference		NNT **	
All lost to follow-up have S-ECC	84	124	0.70	(0.56 - 0.88)	0.12	(0.04 - 0.19)	8	(5 - 25)
None lost to follow-up have S-ECC	4	20	0.21	(0.07 - 0.60)	0.05	(0.02 - 0.08)	20	(13 - 50)
All have S-ECC frequency as intervention group	5	22	0.24	(0.09 - 0.61)	0.05	(0.02 - 0.08)	20	(13 - 50)
All have S-ECC frequency as control group	12	30	0.41	(0.22 - 0.79)	0.06	(0.02 - 0.09)	17	(11 - 50)
All have S-ECC frequency as entire population	8	26	0.32	(0.15 - 0.69)	0.06	(0.02 - 0.09)	17	(11 - 50)
Multiple imputation of missing outcomes	7	29	0.25	(0.11 - 0.56)	0.07	(0.03 - 0.10)	14	(10 - 33)

* Cases of severe early childhood caries (S-ECC) that would have been observed in the intervention (n = 307) and control group (n = 318) with mother-infant pair as randomised, if the assumption about missing outcomes would be correct.
** Number Needed to Treat to prevent one case of S-ECC.

putations were used to create the data set. The independent variables used in the regression model to predict the incidence of S-ECC were maternal age, gestational age at enrolment and at the time of birth, the mother's body mass index and her self-reported decayed, missing or filled teeth (DMFT). Information on each of these was available for more than 97.5% of enrolled women (missing data: range: 0-14, mean: 5.2; 0.8%).

SPSS 15.0 for Windows, IVEware and Revman 4.2 were used for data analysis.

RESULTS

Effects of the Zelen Double Consent Design

Bias, resulting from non-acceptance of the allocation, applied to only five mothers who opted for the intervention group after having been assigned to the control group. As none of their infants experienced S-ECC, their small number marginally affected only the denominators and not the numerators without changing the number needed to treat (NNT) to prevent one case of S-ECC (Table **1**).

Effects of Loss to Follow-up

The assumptions that infants without follow-up would all have S-ECC or would all not have S-ECC, both confirmed the statistically significant effect of the anticipatory guidance (Table **2**). However, the relative risk and absolute risk reduction (i.e. risk difference) changed markedly, resulting in a 2.5 times difference in the NNT from 8 to 20. The more realistic approach of considering that those lost to follow-up would follow the overall pattern or that of their own or their opposite group resulted in marked differences in relative risk, but had little effect on the NNT, ranging between 17 and 20 (Table **2**).

The multiple imputation approach (Table **2**) affected the relative risk compared with both the intention to treat analysis and the analysis based on infants available for outcome

assessment (Table **1**), but the risk difference and, consequently, the NNT of 14 was nearly identical to that calculated on the basis of the available follow-up data (NNT = 13).

DISCUSSION

Over the years RCTs have attracted friends and foes. This is likely to continue as not everything can be evaluated in this way [11]. RCTs are formally planned, prospective studies to compare the effects of an intervention with those obtained by alternative approaches. Randomisation, analogous to a fair coin toss, is the key word in this type of study as it is the only known means of ensuring that every participant, irrespective of her or his prior characteristics, has an equal chance of being assigned to one intervention or another [13]. Proper conduct of an RCT is governed by several further requirements, including a clearly formulated protocol with stated conditions for inclusion and exclusion of subjects and prior calculation of the number needed to minimize the likelihood that the results will reflect the play of chance [13, 20]. While these elements are all essential to the interpretation and meaning of the findings, the real benefit of an RCT arises from whether its findings are likely to be applicable to others, both patients and clinicians, outside the trial.

The two crucial issues of 'are the results an unbiased assessment of the intervention studied' and 'are the results applicable to other people than those in the trial' are commonly referred to in the literature respectively as internal and external validity [25]. The latter is often referred too as 'generalisability' [25]. In the field of health promotion, these two issues are more intertwined than in any other field. If the results are seriously biased, there is no need to know about them. If they are unbiased, but cannot be generalised to a substantial part of the population, their usefulness for enhancing population health is likely to be minimal.

In our RCT, we dealt with these issues using a strict protocol with inclusion and exclusion criteria, validated outcome measures and prior power analysis [7]. We recruited women at all public maternity units in Adelaide. By comparing the characteristics of our participants with those of all women giving birth in Adelaide [26], we could ascertain that these women were a 'random sample' of first time mothers attending for public antenatal care. The results may not be generalisable to the 35% of women seeking private maternity care, as these tend to be older, of higher socio-economic status and more likely to be Caucasian than women attending public hospitals [26]. Our concern, however, was to ensure that the results would be applicable to the majority of first time mothers and their infants in the public health system. Our exclusion criteria, based on the fact that not every pregnant woman will have a child nearly two years down the track, were also explicit and stringent. We could not control, however, for women opting out, choosing the alternative approach, lost to follow-up, or declining further participation for one reason or another. While our choice for a Zelen design was based on the need to enhance participation and to obtain a representative sample of the population, the current in-depth analysis was prompted by the substantial and unavoidable loss to follow-up.

Zelen's original design [15], intended to address issues of validity, is rightly rejected by most ethicists [16, 17], as it results from an era in which people's need for full information about alternative options and responsibility for their own health were not given the same attention as they are today. In its double consent form, Zelen's design [19] largely overcomes such objections and substantially enhances the generalisability of the conclusions that can be derived from a single trial. The findings are more likely to be applicable to the population concerned than those of a trial in which few eligible persons have sufficient equipoise to accept chance allocation. This would seem to be particularly important with preventative strategies. Firstly, there is often some refusal to participate in preventative actions, which the Zelen design can account for. Secondly, contamination of the control group, whose members wittingly or unwittingly acquire some of the exposure that should have been restricted only to those in the intervention group, can be a major problem in such trials. This problem, also referred to as 'contamination bias' or 'dilution bias' [27], is largely avoided by offering participants a genuine choice of moving from one group to the other, if they desire to do so.

Cross-over from one group to another is a major concern, however, as groups that were comparable at the time of randomisation may no longer be comparable at the time of outcome assessment based on the actual treatment received. Analyses based on the 'intention to treat' (ITT) principle, the accepted standard for all RCTs [12, 20], cannot resolve this either when treatments have become effectively 'homogenised' [17, 19]. Adamson et al. [27], who identified 58 trials that had used a Zelen design between 1990 and 2005, found a mean cross-over rate of 13.8%. Even with the single consent design, a cross-over rate of only 10% requires a 20% increase in sample size to maintain power [19]. With a 10% cross-over in the double consent design, sample size needs to increase with 60% in order to achieve the power of a conventional RCT [19, 27]. Fortunately, cross-over was a minor issue in our trial, occurring in less than 1% and being unidirectional, only from the control group to the intervention group. As no cases of S-ECC were observed in the few participants who changed group allocation, adjustment of the outcome statistics affected only the denominators and not the numerators. This may not be so in other trials, however, and emphasises the need for sensitivity analyses when interpreting the results of such trials.

Loss to follow-up is a major issue in any RCT addressing outcomes that are not immediate. After all, its major strength is in the ITT principle, analysing results irrespective of what happens after randomisation [13, 20] or, as others have put it, 'once randomised, always analysed' [14]. In health promotion interventions, which target healthy people and distant outcomes, this is a genuine nightmare. Efforts to track all participants for as long as is needed, however intensive, are bound to fail as people pursue other objectives in life. Sensitivity analyses in general and multiple imputation techniques in particular, imperfect as they are, provide an answer to this. They have not been given much attention, however [12, 28]. For example, the extensive glossary of the Cochrane Collaboration on terms and definitions related to RCTs (http://www.cochrane.org/resources/glossary.htm) does not even refer to multiple imputation.

Frequent as loss to follow-up is, there is no consensus on how to deal with it in an RCT, other than that every effort should be made to reduce losses to the absolute minimum [12-14, 28]. This is more readily achieved when the interval between intervention and outcome is short, when targeting people dependent on ongoing care, and when there is ongoing contact between investigators and participants. None of this applied to our trial, as the protocol stipulated absence of any contact with the control group between informed consent and outcome assessment, about two years later, in order to mimic the real life situation.

Some consider that the results of RCTs with substantial loss to follow-up should be disregarded, but it is arbitrary what percentage should be considered as substantial [14]. When only a minority of people experience the outcome of interest, any loss to follow-up that exceeds the frequency of that outcome has the potential of creating a two-fold difference between what was observed and what could have been observed, if there had been no loss to follow-up. To avoid this issue, some have chosen intermediate outcomes, that are more readily available, as surrogates for what really matters [29, 30]. Not infrequently, intermediate outcomes have led to conclusions that could not be upheld by more substantive assessments later on [30, 31]. It has also led to different interpretations of what is meant by the ITT principle [12, 32]. Generally, little attention has been given to other means of testing the robustness of the findings.

Our most simple sensitivity analyses, assuming all possible outcomes for the missing mothers and infants randomised, as shown in Table 2, confirmed and added strength to the conclusion that anticipatory guidance can reduce the prevalence of S-ECC [7]. In its least favourable or "worst case" scenario, providing anticipatory guidance would still benefit 5% of the population, as one case of S-ECC would be prevented for every 20 pregnant women provided with such guidance. More often than not, however, the "worst case" scenario will provide results that are of little relevance to practice. Even if there are only a moderate number of missing outcomes, assuming all missing values to be good or bad is too strong an assumption to give a reasonable estimate of the effect of an intervention [33]. Moreover, such analyses can easily provide inconsistent or even contradictory results [12], leaving little option but to disregard the data.

While, for our trial, these simple analyses demonstrate the robustness of the findings to any assumptions made about missing outcomes, this is still of little value when people want to know not only whether something works, but also how much it works. Where resources are limited, numbers needed to treat that differ 2.5-fold are not particularly helpful. Multiple imputation, whilst not replacing the need for complete data, provides a far more reasonable approach to the problem. Apart from a more balanced appraisal of the validity of a trial, when there are differences between people randomised and those with outcome assessment, it also provides a more realistic estimate of what can be expected when the intervention is applied to a similar population elsewhere. Obviously, it will be helpful mainly if it incorporates variables that are shown to be predictive of attrition bias and loss to follow-up [33].

In conclusion, although evaluating preventative strategies and oral health promotion programmes in particular by means of RCTs is fraught with difficulties [9, 10], this is not a reason to abandon RCTs when seeking solutions to important health issues. Despite their difficulties, RCTs remain at the top of all known methods for evaluating whether something works or not. Neither do we advocate that sensitivity analyses and multiple imputation techniques provide an adequate alternative to complete follow-up of all subjects randomised. However, RCTs and multiple imputation are not goals by themselves. They are merely tools in the quest of what is and what is not effective in health care. Every tool has its limitations and, as every dentist knows, few complicated tasks are ever accomplished with only a single tool.

ACKNOWLEDGMENTS

REFERENCES

[1]   Drury TF, Horowitz AM, Ismail AI, Maertens MP, Rozier RG, Selwitz RH. Diagnosis and reporting early childhood caries for research purposes: a report of a workshop by the NCDR. J Public Health Dent 1999; 59: 192-7.

[2]   Horowitz HS. Research issues in early childhood caries. Commu Dent Oral Epidemiol 1998; 26(Suppl 1): 67-81.

[3]   Peretz B, Ram D, Azo E, Efrat Y. Preschool caries as an indicator of future caries: a longitudinal study. Pediatr Dent 2003; 25: 114-8.

[4]   Wyne AH. Prevalence and risk factors of nursing caries in Adelaide South Australia. Pediatr Dent J 1999; 9: 31-6.

[5]   Gussy MG, Waters EG, Walsh O, Kilpatrick NM. Early childhood caries: current evidence for aetiology and prevention. J Paediatr Child Health 2006; 42: 37-43.

[6]   Nowak AJ, Casamassimimo PS. Using anticipatory guidance to provide early dental intervention. J Am Dent Assoc 1995; 126: 1156-63.

[7]   Plutzer K, Spencer AJ. Efficacy of an oral health promotion intervention in the prevention of early childhood caries. Commun Dent Oral Epidemiol 2008; 6: 335-46.

[8]   Kotaska A. Inappropriate use of randomized trials to evaluate complex phenomena: case study of vaginal breech delivery. BMJ 2004; 329: 1039-42.

[9]   Watt R, Fuler S, Harnett R, Treasure E, Stillman-Lowe C. Oral health promotion evaluation – time for development. Commun Dent Oral Epidemiol 2001; 29: 161-6.

[10]   Petersen PE, Kwan S. Evaluation of community-based oral health promotion and oral disease prevention – WHO recommendations for improved evidence in public health practice. Commun Dent Health 2004; 21(Suppl): 319-29.

[11]   Freeman R. Health promotion and the randomised controlled trial: a square peg in a round hole? BMC Oral Health 2009; 9: 1 (doi:10.1186/1472-6831-9-1).

[12]   Hollis S, Campbell F. What is meant by intention to treat analysis? survey of published randomised controlled trials. BMJ 1999; 319: 670-4.

[13]   Keirse MJNC, Hanssens M. Control of error in randomised clinical trials. Eur J Obstet Gynecol Reprod Biol 2000; 92: 67-74.

[14]   Schultz KF, Grimes DA. Sample size slippages in randomised controlled trials: exclusions and the lost and wayward. Lancet 2002; 359: 781-5.

[15]   Zelen M. A new design for randomized controlled trials. New Engl J Med 1979; 300: 1242-5.

[16]   Ellenberg SS. Randomisation designs in comparative clinical trials. New Engl J Med 1984; 310: 1404-8.

[17]   Altman DG, Whitehead J, Parmar MKB, Stenning.SP, Fayers PM, Machin D. Randomised consent designs in cancer clinical trials. Eur J Cancer 1995; 31A: 1934-44.

[18]   Homer CS. Using the Zelen design in randomized controlled trials: debates and controversies. J Adv Nurs 2002; 38: 200-7.

[19]   Zelen M. Randomized consent design for clinical trials: an update. Stat Med 1990; 9: 645-56.

[20]   Altman DG, Schultz KF, Moher D, *et al.* For the CONSORT group. the revised CONSORT statement for reporting randomised controlled trials: explanation and elaboration. Ann Intern Med 2001; 134: 663-94.

[21]   Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol 2006; 59: 1087-91.

[22]   Harel O, Zhou XH. Multiple imputation: review of theory implementation and soft ware. Stat Med 2007; 26: 3057-77.

[23]   Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. Surv Methodol 2001; 27: 85-95.

[24]   Raghunathan TE, Solenberger PW, Van Hoewyk J. IVEware: Imputation and Variance Estimation Software 2002; [http://www.isr.umich.edu.src/smp/ive/]

[25]   Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G. Internal and external validity of cluster randomised trials: systematic review of recent trials. BMJ 2008; 336: 876-80.

[26]   Chan A, Scott J, Nguyen AM, Sage L. Pregnancy outcome in South Australia 2007. Adelaide: Pregnancy outcome unit, SA Health, Government of South Australia 2008.

[27]   Adamson J, Cockayne S, Puffer S, Torgerson DJ. Review of randomised trials using the post-randomised consent (Zelen's) design. Contemp Clin Trials 2006; 27: 305-19.

[28]   Wright CC, Sim J. Intention-to-treat approach to data from randomized controlled trials: a sensitivity analysis. J Clin Epidemiol 2003; 56: 833-42.

[29]   Hujoel PP, DeRouen TA. A survey of endpoint characteristics in periodontal clinical trials published 1988-1992 and implications for future studies. J Clin Periodontol 1995; 22: 397-407.

[30]   Grimes DA, Schultz KF. Surrogate outcomes in clinical research: hazardous to your health. Obstet Gynecol 2005; 105: 1114-8.

[31]   Glezerman M. Five years to the term breech trial: the rise and fall of a randomized controlled trial. Am J Obstet Gynecol 2006; 194: 20-5.

[32]   Gravel J, Opatrny L, Shapiro S. The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? Clin Trials 2007; 4: 350-6.

[33]   Sterne JA, White IR, Carlin JB, *et al.* Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009; 338: 157-60.