Essay

Preventing Bias in Cluster Randomised Trials

Bruno Giraudeau[1,2,3,4]*, Philippe Ravaud[4,5,6]

1 INSERM, CIC 202, Tours, France, 2 Université François Rabelais, Tours, France, 3 CHRU de Tours, Tours, France, 4 INSERM, U738, Paris, France, 5 Assistance Publique–Hôpitaux de Paris, Hôpital Bichat, Département d'Epidémiologie, Biostatistique et Recherche Clinique, Paris, France, 6 Université Paris 7–Denis Diderot, Paris, France

In randomised trials, internal validity is defined as "the extent to which the design and conduct of a study are likely to have prevented bias" [1,2]. In conducting such trials, trialists try to prevent selection bias through randomisation and allocation concealment (defined as "the process used to ensure that the person deciding to enter a participant into a randomised controlled trial does not know the comparison group into which that individual will be allocated" [2]) and attrition bias through an intention-to-treat (ITT) analysis. ITT analysis has indeed been defined as one of the cornerstones of the analysis strategy of randomised trials because it allows for preserving the benefits of randomisation [3]. With an ITT analysis, data for all randomised participants are analysed in the groups to which they were originally randomly allocated "regardless of their adherence with the entry criteria, regardless of the treatment they actually received, and regardless of subsequent withdrawal from treatment or deviation from the protocol" [4]. ITT analysis entails the use of ad hoc statistical methods to handle missing outcome data when participants withdraw from the trial or are lost to follow-up [4,5]. The ITT principle is widely used in analysing data from individually randomised trials but is much more difficult in cluster randomised trials (CRTs), and this issue is not clearly covered in the main methodological textbooks on the topic [6,7].

Here, we describe the difficulties in preventing selection bias and applying ITT analysis in CRTs (other biases are discussed in Puffer et al. [8]) and propose some solutions to deal with these issues in this trial design.

CRTs: Randomising Clusters Rather Than Individuals

In CRTs, "intact social units, or clusters of individuals, rather than individuals themselves, are randomised" [6]. For example, hospitals, wards, or physicians

Summary Points

- In CRTs, the comparability of groups is challenged because groups of trial participants rather than the participants themselves are randomised.

- The specific chronology of such trials compromises allocation concealment (i.e., clusters are recruited and randomised and then participants are recruited), which can induce differential recruitment and thus quantitative and qualitative imbalance between groups.

- The principle of intention to treat is challenged in CRTs because of the lack of any statistical method to handle non-recruited participants.

- Empty clusters (i.e., clusters with no data for participants), which are randomised units, are discarded from the analysis—a violation of the very principle of intention to treat.

- Some CRTs may be better analysed as observational studies, with some form of adjustment used such as propensity-score methods.

can be randomised, as well as schools or geographical areas. Such a design is well adapted to assess organisational and behavioural interventions and health promotion programmes—interventions that are usually implemented at the level of health organisational units or geographical areas

[9]. As well, randomisation of clusters rather than individuals could prevent contamination. As an example, to assess the impact of a screening programme, widespread publicity is needed to encourage participants to undergo screening. With individual randomisation, communication among participants would induce contamination, but with cluster randomisation, it may strengthen participant compliance.

Use of the CRT has greatly increased over the past 15 years [10] and has motivated the publication of an extension of the CONSORT Statement for this design [11] because of its particular methodological issues. The main issue is that observations from the same cluster are more similar than observations from two different clusters. This situation requires the use of both an inflated sample size and adapted statistical analysis to take into account this concern [6,7]. Otherwise, for the recruitment process, we can distinguish two different designs: a whole CRT or a CRT with active recruitment. In the first design, once the person in charge of the cluster (the cluster guardian [12]) has agreed to participate, every individual belonging to the cluster is automatically recruited for the trial; in the second design, the guardian (or someone other than the guardian but depending on the guardian) must select and include participants for the trial. Also, Murray distinguishes between cohort designs, whereby a sample of participants is included, followed up, and assessed repeatedly (as is usually done in individually randomised trials), and repeated cross-sectional designs, whereby a distinct sample of participants is assessed

The Essay section contains opinion pieces on topics of broad interest to a general medical audience.

A   **Whole cluster randomised trials (cohort design)**

First clusters are randomised, then any participant belonging to the cluster is automatically included in the trial. In cohort designs, participants are identified and followed up (e.g. Campbell *et al* [29]).
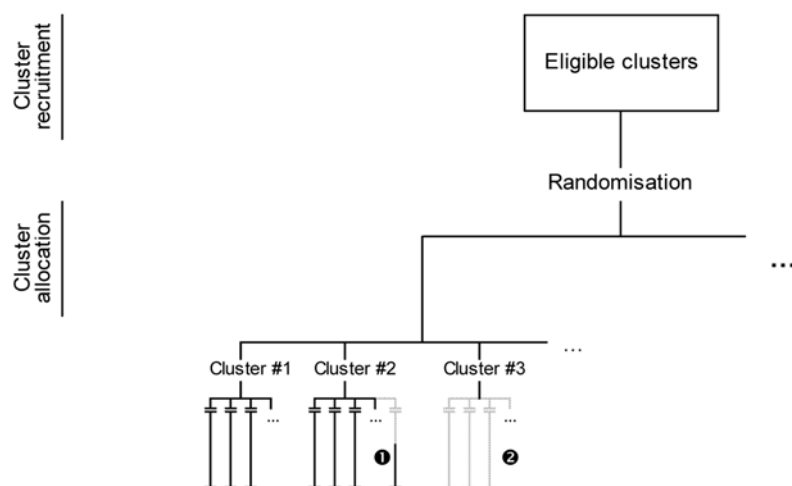
❶ In-migrant participant.
❷ Cluster withdrawal after randomisation and prior to any collection of data on participants.
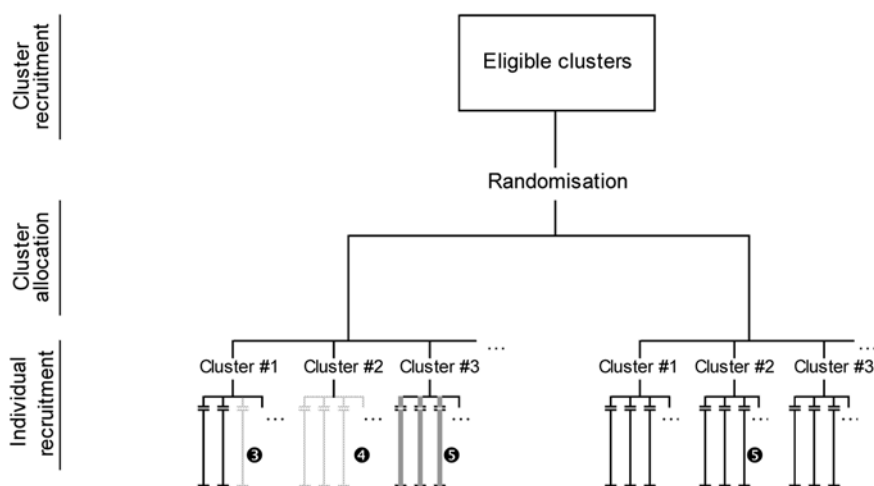
[29] To assess the effectiveness of a peer-led intervention that aimed to prevent smoking uptake in secondary schools, 59 schools were randomised. Any student was included and assessed at baseline, immediately after the intervention and at 1 and 2 years.

B   **Cluster randomised trials with active recruitment**

First clusters are recruited, then randomised, and finally, participants are recruited (e.g. Bellary *et al* [30]).

❸ Eligible but nonincluded participant.
❹ Inactive cluster guardian: no eligible participant recruited.
❺ Nonsimilarity of recruited participants between intervention groups.

[30] To assess the effectiveness of a culturally sensitive, enhanced care package for improving of cardiovascular risk factors in patients of South Asian origin with type 2 diabetes, 21 UK general practices were randomised. In each practice, eligible patients were included by the physicians in charge of the cluster (i.e. the guardian) and followed-up during a two-year period

Figure 1. Recruitment issues in cluster randomised trials.
doi:10.1371/journal.pmed.1000065.g001

❶ Participant withdrawal or lost to follow up.
❷ Cluster withdrawal or lost to follow up.
❸ Participant transfer from one cluster to another. This may lead to a switch between intervention groups.
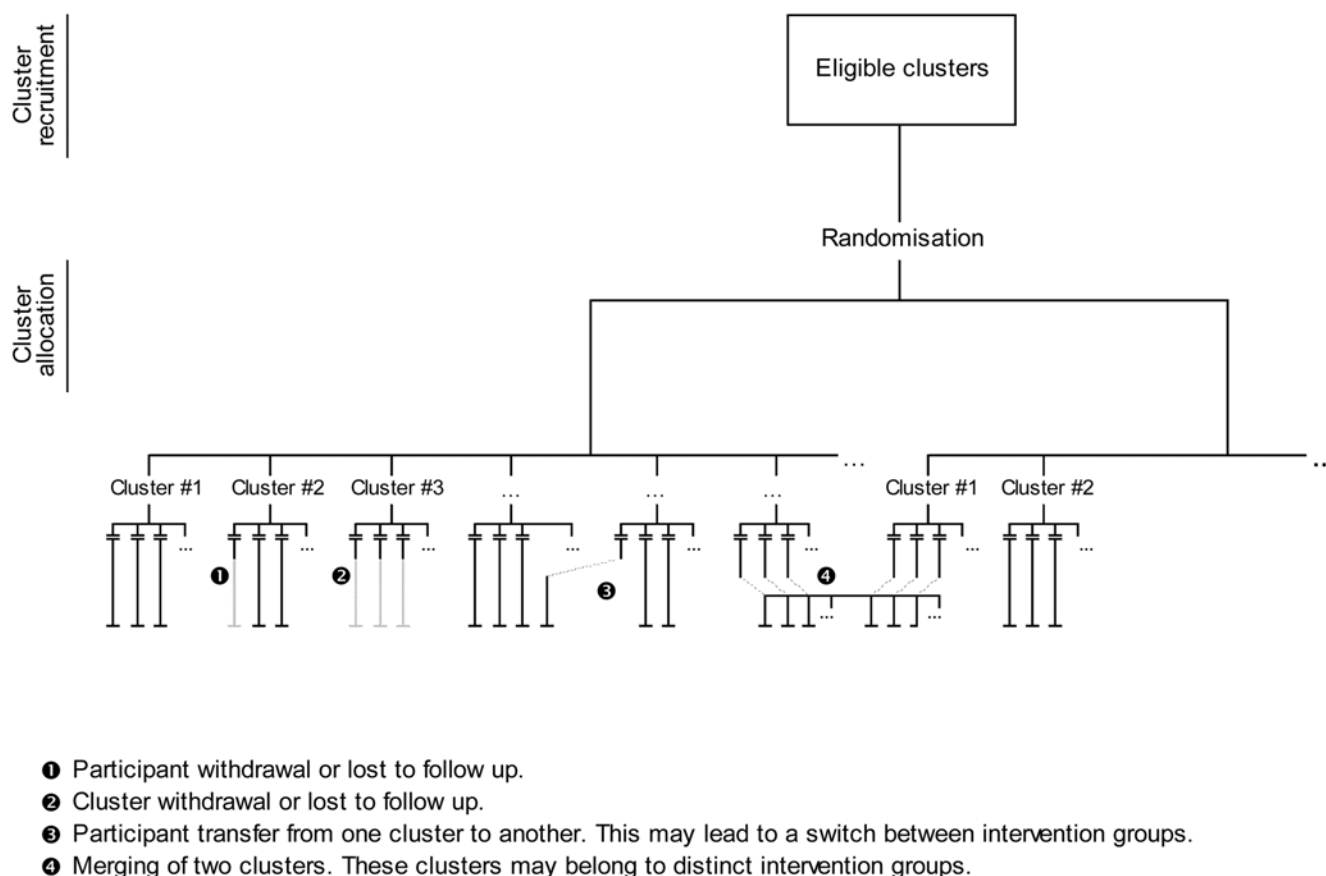❹ Merging of two clusters. These clusters may belong to distinct intervention groups.

Figure 2. Follow-up issues in cluster randomised trials.
doi:10.1371/journal.pmed.1000065.g002

at each assessment time [7]. Such a distinction makes sense only in whole CRTs. The design is always a cohort one when it involves active recruitment.

The Challenge of Comparability of Groups in CRTs

In individually randomised trials, randomisation and allocation concealment aim to prevent selection bias and allow for the comparability of groups at the beginning of the study. The aim of ITT analysis is to maintain this comparability during analysis. In CRTs, the situation is more complex because (i) the design entails a hierarchical structure (cluster and individual levels) and (ii) recruitment and follow-up processes differ from those in individually randomised trials.

First, the hierarchical structure of CRTs requires the consideration of both the cluster and individual levels. When El-bourne et al. [13] appealed for an extension of the CONSORT Statement to CRTs, they explicitly stated that authors must report whether "[the ITT] principle applies to clusters, individuals or both". Campbell

defined ITT analysis in CRTs as taking into account all participants, regardless of whether non-adherence to the protocol occurred at the individual or cluster level [14]. Indeed, in CRTs, participants lost to follow-up and protocol deviation can occur at the level of the cluster (cluster withdrawal or lost to follow-up, inactive cluster, merging of clusters) or the individual (participant withdrawal or lost to follow-up, or transfer from one cluster to another). Moreover, results of a CRT can be analysed by considering the individual or cluster as the statistical unit. Even in the latter case, a summary statistic is estimated from individual responses within each cluster. Therefore, the comparability of groups must be achieved at both the cluster (i.e., for randomised units) and individual levels.

Second, in a CRT, both recruitment and follow-up raise specific issues that may compromise the comparability of groups (Figures 1 and 2). The usual definition of an ITT analysis in CRTs is restrictive because it focuses only on follow-up issues. However, during the recruitment process, clusters may withdraw, or cluster guardians may not actively recruit participants; these situations

lead to empty clusters, which should be taken into account to apply the ITT principle. Another important issue is differential recruitment between active clusters, which could be both quantitative (i.e., different number of participants recruited) and qualitative (i.e., participants with different characteristics recruited in both groups). As an example, in the UK BEAM trial, 71.4% of recruited participants were in the intervention group, even though the randomisation was 1:1, and groups were not comparable on several clinically important criteria [15]. Eldridge et al. concluded that "about a quarter [of CRTs] were potentially biased because of procedures surrounding recruitment and identification of patients" [16], as previously acknowledged by Puffer et al. [8].

The Challenge of Applying ITT Analysis in CRTs

In CRTs, cluster withdrawal or loss to follow-up and participant withdrawal or loss to follow-up must be handled as is usually done in individually randomised trials, by use of ad hoc missing data methods [17,18]. More specific to CRTs

Box 1: Recommendations for the Planning of and Analysis of Data in CRTs To Improve the Comparability of Groups

Handling Recruitment Issues
Cluster level:

- Monitor cluster guardians' adherence to the protocol before the randomisation of the cluster they are in charge of.
  - For whole CRTs, this monitoring will help prevent clusters withdrawing after randomisation and before the collection of data for participants.
  - For CRTs with active recruitment, this monitoring will help prevent empty clusters due to inactivity of guardians.

Individual level:

Whole CRTs

- Ensure the inclusion of all participants belonging to the randomised clusters.
- For a cohort design (i.e., participants are included, followed up, and assessed):
  - Consider participants who migrate out of clusters (out-migrants) as being withdrawn or lost to follow-up.
  - Discard data for participants who migrate into a cluster (in-migrants).

CRTs with active recruitment

- Whenever possible, identify and completely include participants before randomising a cluster, to maintain the usual chronology of a randomised trial (recruitment followed by randomisation) and help prevent selection bias.

If complete inclusion of participants is not possible, the following two solutions can be used in combination:

- If possible, randomise clusters only when the first participant is included (index case concept) to prevent empty clusters.
- If possible, have blinded independent recruiters include participants, so that the inclusion process can be independent from the allocation process.

Analysis: Handling Follow-Up Issues
Cluster level:

- Take into account any cluster that withdrew or was lost to follow-up by using ad hoc missing data methods for participants included in these clusters.
- If two or more clusters merge during the trial, split the merged clusters so that the clusters analysed are those that were initially recruited.

Individual level:

- Take into account any participant who withdrew or was lost to follow-up by using ad hoc missing data methods.
- Keep any participant who transferred from one cluster to another in the original cluster.

Analysis: Handling Quantitative and Qualitative Imbalance between Intervention Groups
Whole CRTs:

- Use adjustment methods to cope with potential imbalances in both cluster and individual characteristics, which are frequent when the number of randomised clusters is small.

CRTs with active recruitment:

- Calibrate estimation by using weighting adjustments (e.g., propensity-score methods) to deal with quantitative and qualitative imbalance due to potential differential recruitment induced by lack of allocation concealment when including participants.

---

are participant switches from one cluster to another (e.g., patients who change physicians) and cluster merging (e.g., two medical practices that merge). These follow-up issues are not problematic per se and can be easily handled by keeping clusters as they were randomised. Thus, participants who switch from one cluster to another must be kept in their original cluster, and merged clusters must be split.

The major difference between individually randomised trials and CRTs in applying the ITT principle is that in the former, we know which participants were randomised and must be taken into account in the ITT analysis, whereas in CRTs, we know which clusters have been randomised but we frequently do not know exactly which participants within clusters should have been included in the trial. Whole CRTs and CRTs with active recruitment raise distinct issues, but for both, this dilemma is the real challenge of applying ITT analysis and preserving the comparability of groups.

Whole CRTs

In whole CRTs, empty clusters occur when cluster guardians withdraw just after the randomisation and before any data on participants are collected. The guardians make these decisions once they know the allocation result, which could influence the likelihood of withdrawal. So excluding such empty clusters from analysis (as is always done) is in essence post-randomisation exclusion. Furthermore, in CRTs, the number of randomised units is much lower than that in individually randomised trials, so any randomised unit has greater influence in a CRT than in an individually randomised trial. Also, in CRTs with empty clusters, although we know some cluster characteristics, often we do not know the number or characteristics of participants who should have been included in those clusters if they had been active. This scenario prevents the use of any statistical solution: knowing cluster characteristics does not allow in any way the derivation of participant characteristics. Such randomised clusters that lack any data on participants are then removed before analysis, which violates the very definition of the ITT (i.e., exclusion of randomised units). In CRTs, therefore, extreme vigilance is needed to ensure cluster guardians' adherence to the trial protocol, before the randomisation of the clusters they are in charge of, because currently, we lack statistical methods to limit the induced bias.

CRTs with Active Recruitment

As emphasised, CRTs with active recruitment "first recruit the clusters, then

randomise and finally recruit the participants: such an approach invites bias" [19]. In the extreme, some cluster guardians may be inactive (i.e., they do not recruit any participants), and these inactive guardians may moreover be more numerous in one group than in another. This situation is similar to "passive withdrawal": guardians are inactive recruiters because they do not adhere to the allocation result (i.e., the cluster they are in charge of has not been randomised to the group they expected to be in). Actually, any time recruitment of participants occurs after randomisation, (i) participants are selected by someone who could be aware of the group to which the participants will be allocated, and (ii) participants consent not to random allocation but rather to participation in and allocation to a pre-determined group, which could induce differential recruitment or consent and selection bias [8,20,21]. The only way to prevent differential recruitment is to ensure some form of allocation concealment during recruitment. Puffer et al. [8] proposed the identification and complete inclusion of participants before the randomisation of a cluster, which allows for maintenance of the usual chronology of a randomised trial. Such a strategy thus prevents both empty randomised clusters and selection bias but cannot be systematically implemented for logistical reasons (e.g., in a trial including incident patients rather than prevalent cases). Two comple-

mentary solutions are (i) changing the time of randomisation by randomising clusters only when the first participant is identified (the index case concept) and (ii) blinding independent recruiters to the allocation group. The index case concept could prevent empty clusters but does not prevent differential recruitment.

Actually, some similarity exists between this potential differential recruitment in CRTs and survey non-response. Non-response is a source of bias because non-responders are not a random sample of surveyed people. In CRTs, participants included without allocation concealment are not a random sample of eligible participants. In the context of survey non-response, much work has been done to calibrate estimation, and methods such as weighting adjustments or imputations are classically used [22,23]. In CRTs, the situation is more complex because contrary to surveys, the number of eligible participants is not known. However, the question remains as to whether all CRTs really benefit from randomisation and whether some trial results would be better analysed as observational studies with some form of adjustment used, such as propensity scores [24,25]. Some methodologists have resorted to such a solution [26,27], which raises another issue: "how the levels of the individual subject and the cluster should be considered in the

estimation and application of propensity scores" [28]. Further statistical research to transpose such methods specifically to CRTs is therefore warranted.

We provide some recommendations for the planning and analysis of data in CRTs to improve the comparability of groups in such trials in light of the issues of recruitment and ITT analysis (Box 1): some issues may be easily handled, whereas others remain without a solution.

Conclusion

The application of the ITT principle in analysis of data is much more complex in CRTs than in individually randomised trials because the principle must be applied at the level of both the cluster and the individual, and because of challenging issues surrounding the recruitment process in CRTs. These issues raise concerns regarding the internal validity of such trials.

Acknowledgments

Author Contributions

ICMJE criteria for authorship read and met: BG PR. Contributed to the writing of the paper: BG PR.

References

1. Campbell DT (1957) Factors relevant to the validity of experiments in social settings. Psychol Bull 54: 297–312.
2. Green S, Higgins J, eds (2005) Glossary. Cochrane handbook for systematic reviews of interventions 4.2.5. Available: http://www.cochrane.org/resources/glossary.htm. Accessed 23 March 2009.
3. Fisher LD, Dixon DO, Herson J, Frankowski RK, Hearon MS, et al. (1990) Intention to treat in clinical trials. In: Pearce KE, ed. Statistical issues in drug research and development. New York: Marcel Dekker. pp 331–350.
4. Hollis S, Campbell F (1999) What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ 319: 670–674.
5. Lachin JM (2000) Statistical considerations in the intent-to-treat principle. Control Clin Trials 21: 167–189.
6. Donner A, Klar N (2000) Design and analysis of cluster randomization trials. London: Arnold. 178 p.
7. Murray DM (1998) Design and analysis of group-randomized trials. New York: Oxford University Press. 467 p.
8. Puffer S, Torgerson D, Watson J (2003) Evidence for risk of bias in cluster randomised trials: Review of recent trials published in three general medical journals. BMJ 327: 785–789.
9. Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG (1999) Methods for evaluating area-wide and organisation-based interventions in health and health care: A systematic review. Health Technol Assess 3: iii-92.

10. Bland JM (2004) Cluster randomised trials in the medical literature: Two bibliometric surveys. BMC Med Res Methodol 4: 21.
11. Campbell MK, Elbourne DR, Altman DG (2004) CONSORT statement: Extension to cluster randomised trials. BMJ 328: 702–708.
12. Edwards SJ, Braunholtz DA, Lilford RJ, Stevens AJ (1999) Ethical issues in the design and conduct of cluster randomised controlled trials. BMJ 318: 1407–1409.
13. Elbourne DR, Campbell MK (2001) Extending the CONSORT statement to cluster randomized trials: For discussion. Stat Med 20: 489–496.
14. Campbell MJ (2000) Cluster randomized trials in general (family) practice research. Stat Methods Med Res 9: 81–94.
15. Farrin A, Russell I, Torgerson D, Underwood M (2005) Differential recruitment in a cluster randomized trial in primary care: The experience of the UK back pain, exercise, active management and manipulation (UK BEAM) feasibility study. Clin Trials 2: 119–124.
16. Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G (2008) Internal and external validity of cluster randomised trials: Systematic review of recent trials. BMJ 336: 876–880.
17. Taljaard M, Donner A, Klar N (2008) Imputation strategies for missing continuous outcomes in cluster randomized trials. Biom J 50: 329–345.
18. Hunsberger S, Murray D, Davis CE, Fabsitz RR (2001) Imputation strategies for missing data in a school-based multi-centre study: The Pathways study. Stat Med 20: 305–316.

19. Torgerson DJ, Torgerson CJ (2008) Designing randomised trials in health, education and the social sciences. An introduction. New York: Palgrave Macmillan. 256 p.
20. Hahn S, Puffer S, Torgerson DJ, Watson J (2005) Methodological bias in cluster randomised trials. BMC Med Res Methodol 5: 10.
21. Jordhoy MS, Fayers PM, Ahlner-Elmqvist M, Kaasa S (2002) Lack of concealment may lead to selection bias in cluster randomized trials of palliative care. Palliat Med 16: 43–49.
22. Groves RM, Dillman DA, Eltinge JL, Little RJA (2002) Survey nonresponse. New York: John Wiley & Sons. 520 p.
23. Kessler RC, Little RJ, Groves RM (1995) Advances in strategies for minimizing and adjusting for survey nonresponse. Epidemiol Rev 17: 192–204.
24. Rosenbaum P, Rubin D (1983) The central role of the propensity score in observational studies for causal effects. Biometrika 70: 41–55.
25. Rosenbaum PR, Rubin DB (1984) Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc 79: 516–524.
26. Ravaud P, Flipo RM, Boutron I, Roy C, Giraudeau B, et al. (2009) ARTIST (osteoarthritis intervention standardized) study of standardised consultation versus usual care for patients with osteoarthritis of the knee in primary care in France: Pragmatic randomised controlled trial. BMJ 338: b421. doi:10.1136/bmj.b421.
27. van Marwijk HW, Ader H, de Haan M, Beekman A (2008) Primary care management of major depression in patients aged > or = 55

years: Outcome of a randomised clinical trial. Br J Gen Pract 58: 680–686, I–II; discussion 687.

28. Nicholas J, Gulliford MC (2008) Commentary. What is a propensity score? Br J Gen Pract 58: 687.

29. Campbell R, Starkey F, Holliday J, Audrey S, Bloor M, et al. (2008) An informal school-based peer-led intervention for smoking prevention in adolescence (ASSIST): A cluster randomised trial. Lancet 371: 1595–1602.

30. Bellary S, O'Hare JP, Raymond NT, Gumber A, Mughal S, et al. (2008) Enhanced diabetes care to patients of south Asian ethnic origin (the United Kingdom Asian Diabetes Study): A cluster randomised controlled trial. Lancet 371: 1769–1776.