UMass Chan Medical School

eScholarship@UMassChan

2013-01-01

The incident user design in comparative effectiveness research

Eric S. Johnson
Kaiser Permanente

Et al.

Let us know how access to this document benefits you.

Repository Citation

REVIEW

The incident user design in comparative effectiveness research

Eric S. Johnson[1]*, Barbara A. Bartman[2], Becky A. Briesacher[3], Neil S. Fleming[4], Tobias Gerhard[5], Cynthia J. Kornegay[6], Parivash Nourjah[2], Brian Sauer[7], Glen T. Schumock[8], Art Sedrakyan[9], Til Stürmer[10], Suzanne L. West[11] and Sebastian Schneeweiss[12]

[1] *The Center for Health Research, Kaiser Permanente, Portland, Oregon, USA*
[2] *Center for Outcomes and Evidence, Agency for Healthcare Research and Quality, Rockville, Maryland, USA*
[3] *Division of Geriatric Medicine, University of Massachusetts Medical School, Worcester, Massachusetts, USA*
[4] *Center for Health Care Research, Baylor Health Care System, Dallas, Texas, USA*
[5] *Institute for Health, Health Care Policy, and Aging Research, Rutgers University, New Brunswick, New Jersey, USA*
[6] *Office of Surveillance and Epidemiology, Food and Drug Administration, Silver Spring, Maryland, USA*
[7] *Salt Lake City IDEAS CenterVeterans Affairs Salt Lake City Health Care System, Salt Lake City, Utah, USA*
[8] *Department of Pharmacy Practice, University of Illinois at Chicago School of Pharmacy, Chicago, Illinois, USA*
[9] *Division of Outcomes and Effectiveness Research, Weill Cornell Medical College, New York, New York, USA*
[10] *Department of Epidemiology, University of North Carolina, Gillings School of Global Public Health, Chapel Hill, North Carolina, USA*
[11] *Health, Social and Economics Research, Research Triangle Institute International, Research Triangle Park, North Carolina, USA*
[12] *Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA*

ABSTRACT

Comparative effectiveness research includes cohort studies and registries of interventions. When investigators design such studies, how important is it to follow patients from the day they initiated treatment with the study interventions? Our article considers this question and related issues to start a dialogue on the value of the incident user design in comparative effectiveness research. By incident user design, we mean a study that sets the cohort's inception date according to patients' new use of an intervention. In contrast, most epidemiologic studies enroll patients who were currently or recently using an intervention when follow-up began. We take the incident user design as a reasonable default strategy because it reduces biases that can impact non-randomized studies, especially when investigators use healthcare databases. We review case studies where investigators have explored the consequences of designing a cohort study by restricting to incident users, but most of the discussion has been informed by expert opinion, not by systematic evidence. Published 2012. This article is a U.S. Government work and is in the public domain in the USA.

KEY WORDS—inception cohort; incident interventions; study design; comparative effectiveness; pharmacoepidemiology

INTRODUCTION

Comparative effectiveness research includes cohort studies and registries of interventions. How important is it to follow patients from the day they initiated treatment with the study interventions? How well can non-randomized studies approximate randomized controlled trials if they follow continuing or prevalent users? What trade-offs do investigators face when deciding where their study will fall on the continuum from restricting cohort enrollment to patients who are naive to the entire class of an intervention (most restrictive design) to expanding enrollment to patients regardless of their past use?

Our article considers the questions outlined earlier and related issues to start a dialogue on the value of incident user designs in comparative effectiveness research. We take the incident user design as a reasonable default strategy because it reduces biases, especially when investigators use secondary data sources, such as healthcare databases. Although the incident user design is preferable on theoretical grounds, there may be exceptions where that eligibility criterion does not matter and a less restrictive study design may provide a valid answer that is more timely, more affordable, and more applicable to routine care. We review case studies where investigators have

Correspondence to: Eric S. Johnson, The Center for Health Research, Kaiser Permanente, Portland, Oregon, USA. E-mail: Eric.S.Johnson@kpchr.org

explored the consequences of designing a cohort study by restricting to incident users, but most of the discussion has been informed by expert opinion, not by systematic evidence. Because we only reviewed published literature, human subjects' approval was not required.

The objectives of this article were to consider the incident user design as a default or first-line study design for comparative effectiveness research and to provide guidance with discussion on the advantages and limitations of the approach. We organized the article as follows: (i) issues that motivate the incident user design; and (ii) definition of incident users in retrospective studies: trade-offs for internal validity and applicability (with further discussion on internal validity, trade-offs, and related design and analysis choices in the Supporting Information). We end with a brief set of recommendations for reporting.

ISSUES THAT MOTIVATE THE INCIDENT USER DESIGN

Investigators struggle to design non-randomized studies that obtain findings as credible as those from randomized controlled trials. To achieve that credibility, they restrict the question, the design, and the analysis.[1–3] However, investigators designing non-randomized studies typically cannot achieve the additional credibility attained through blinding or masking patients and providers to the interventions.[4]

Without the incident user design, investigators may struggle to replicate the eligibility criteria of a randomized controlled trial. For example, a patient who appears eligible today as a current user of an intervention may have been too young to be eligible at the time he or she started the intervention years earlier. It could also be that his or her laboratory and blood pressure findings may have been modified by years of intervention so that he or she now appears ineligible, although he or she would have met criteria at the time he or she started the intervention. Consequently, alternatives to the incident user design can generate uncertainty about the population's eligibility and complicate comparisons across studies.

Sometimes, the findings from non-randomized studies of comparative effectiveness disagree with the findings from randomized controlled trials. Discrepancies between the cardiovascular findings from the Women's Health Initiative (WHI) clinical trial and its cohort study of hormone therapy—as well as other studies on the topic—may be the most discussed example.[5–8] Ray suggested that the discrepancy may be explained, in part, by restricting non-randomized studies to new (or incident) users of hormone therapy because the incident user design can reduce biases that occur when comparisons include patients who were already using the drug at the start of the study.[5]

Yet, other investigators have approximated the findings from randomized controlled trials by comparing current users of drug therapy and ignoring the duration of drug use. For example, Psaty and colleagues reported that current users of angiotensin-converting enzyme inhibitors and current users of diuretics experienced a similar rate of myocardial infarction: Their case–control study findings (for that comparison and outcome) agreed closely with similar head-to-head comparisons in the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT).[9,10] Ray and colleagues found that the excess rate of coronary heart disease events among patients using COX-2 selective non-steroidal anti-inflammatory drugs (NSAIDs) was consistent for current users and the subgroup of incident users—a rare comparison of strategies for defining cohorts within a cohort study.[11] Given the inconsistency in these examples, what is the theoretical motivation to prefer the incident user design over other choices that enroll continuing or prevalent users? Below, we consider two theoretical motivations. Although it is beyond the scope of this article, there are additional theoretical motivations for the incident user design including the clear articulation of a causal question about the benefits or harms of starting interventions; in contrast, the counterfactual may be harder to articulate for studies that enroll continuing or prevalent users.

Avoiding adjustment of intermediate covariables

One reason to prefer the incident user design is that it avoids the problem of adjusting for characteristics that may be in the causal pathway.[5] For example, in the case–control study noted earlier, Psaty and colleagues went to exceptional effort to obtain patients' pretreatment blood pressure values, which were documented in paper charts at the health maintenance organization (HMO) an average of 11 years before the date that determined current drug use.[9] Even with that effort, pretreatment blood pressure values were missing in one third of patients who started treatment before joining the HMO. Had Psaty and colleagues adjusted for patients' most recent blood pressure values, they would have biased the comparisons between classes of antihypertensive drugs because the most recent blood pressure value is in the causal path between treatment and myocardial infarction and shares common causes with myocardial infarction. Because most non-randomized studies lack detailed historical data on pretreatment characteristics, it is often more

credible to restrict the design to incident users—for whom such pretreatment characteristics can be collected more completely and reliably.

A fair representation of early and late events

The second reason to prefer the incident user design is that it captures all events that occurred after the start of therapy.[5,12,13] Molride and Abenhaim explained that some patients are susceptible to harm from a drug, and those events may occur earlier in the course of therapy. Once these susceptible patients have suffered events early in the course of therapy, only less susceptible patients remain. Mixing incident and prevalent users may obscure excess harm because the effect measure is weighted toward prevalent users who provide the majority of person-time and were less susceptible to the harm.[13,14] Similar concerns apply to any outcome, including benefits, when the population is required to survive during the intervention.

Guess described a related idea by noting that the hazard ratio for harm changed over time since the start of drug use.[15] Such changes in effect size as a function of the duration of drug use may result from shifts in cohort composition—as described earlier— or biologic effects or both. Similarly, an intervention's benefits may require an induction period of months or years to reduce clinical event rates: Newly diagnosed patients with diabetes who started intensive therapy for glucose control took up to 10 years to achieve the clinical benefits in the United Kingdom Prospective Diabetes Study.[16] When Prentice and colleagues considered patients' duration of hormone therapy—before the start of the study—in the WHI cohort study, their findings approximated more closely the findings from the WHI's controlled trial—at least for some endpoints.[6]

Newly marketed drugs may have a disproportionate share of incident users compared with drugs that were marketed years earlier. For example, a cohort study that was conducted shortly after the marketing of celecoxib that compared patients currently using celecoxib versus those currently using naproxen could distort estimates of comparative effectiveness because the celecoxib users would be more likely to be incident users at higher risk of any early harms related to NSAIDs. Restricting enrollment to incident users enables comparisons at a comparable time in the natural history of their treatment.

Ray traced the idea of the incident user design back to Feinstein's article on chronology bias, which appeared in 1968.[17] Ray's own review of new user design for non-randomized studies remains the most comprehensive account of its value in reducing bias.[5]

McMahon and MacDonald provided an earlier and thoughtful consideration of the new user design.[12] A task force on research practices for retrospective databases organized by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) has also addressed the incident user design and its alternatives.[14] Our article builds on their efforts by considering how the incident user design applies to comparative effectiveness research. We expand on their previous discussions to consider other interventions, such as medical devices, although most of the examples concern medications. This article asks that investigators consider the incident user design as a default strategy but recognizes that there will be exceptions where current user designs may be preferred. This article encourages more transparent reporting of design choices—in the spirit of STROBE (Strengthening the Reporting of Observational Studies in Epidemiology)—and an appreciation of the trade-offs that those choices may entail for validity, applicability, timeliness, and feasibility.[18]

DEFINING INCIDENT USERS IN RETROSPECTIVE STUDIES: TRADE-OFFS FOR INTERNAL VALIDITY AND APPLICABILITY

No consensus exists for defining an incident user of a drug or other intervention using secondary data sources, such as computerized pharmacy fill records or electronic health record prescription orders (see Supporting Information for a discussion of related ideas as they apply to prospective studies). Ray and colleagues defined incident use of NSAIDs as patients who had not filled those medications in the 365 days preceding time zero.[11] In theory, the physician's order date, or prescription, recorded in some electronic health records, could serve as a more meaningful time zero for approximating a cohort study analogue of the intention-to-treat analysis used in randomized controlled trials.[19] For example, some physicians may dispense product samples along with a prescription; if the patient tolerates the sample, then he or she may fill the prescription, which would appear later as a pharmacy claim. Consequently, the prescription date could reflect the start of therapy more accurately.[20] However, some patients will not fill the prescription, sometimes called primary non-adherence, which would attenuate the comparative effect.

Recurrent use

Because some health plans will allow patients to fill a 180-day supply of medication, submitted as a single pharmacy claim, briefer definitions may misclassify patients who actually used their medication during

the baseline period. For example, a patient may have filled his prescription 200 days before the date selected to define incident use. If the study used 180 days to define incident users, he or she would appear to be an incident user but would have been taking medication during the months when baseline characteristics were measured. The challenge is more complicated than it may seem: Some patients take long drug holidays and then restart their medication.[21] Others adhere so poorly—say, every other day—that a prescription intended as a 90-day supply can persist for 180 days. Recognizing the range of possibilities brings a sense of humility about classifying patients as incident users; only a proportion of apparent incident users are truly treatment naive as of the date that investigators assign for time zero. Some investigators may prefer a data-driven method for identifying the number of days that define incident use in their population by using pharmacy databases and the waiting time distribution.[22]

When investigators know that a patient has used the intervention in the past, one option is to stratify comparative estimates so that each group has its own baseline hazard for calculating the hazard ratio that captures comparative effectiveness. For example, patients with a known history of NSAID use more than 365 days before time zero could be compared with other patients who had the same known history. If the investigators know that medication is used episodically (e.g., cycles of chemotherapy) and they wish to capture the totality of benefits and harms across those episodes, then they should consider more complex structural models that address time-varying drug exposures and confounders; otherwise, it is preferable to evaluate the benefits and harms for the first observed episode only.[23]

Recently marketed drugs

For a recently marketed drug, defining incident use as no fills in the preceding 365 days may identify patients for whom the first observed prescription fill in a given database represents their first ever use (i.e., truly treatment naive). For older drugs, defining incident use as no fills in the preceding 365 days may identify patients starting a new episode of therapy, but not necessarily their first ever use. For example, when Ray and colleagues conducted their cohort study using pharmacy data from 1999 through 2001, naproxen had been marketed in the USA since 1980; consequently, apparent incident users during 1999 through 2001 may have had past episodes of naproxen use and survived any harms that they experienced during earlier episodes.[2,11] In contrast, apparent incident users

of the more recently marketed COX-2 selective NSAIDs, celecoxib and rofecoxib, were more likely to be treatment naive—at least to those products (but not all NSAIDs). If investigators wish to analyze recurrent episodes of therapy, it is important to adjust standard errors for the correlation of episodes within patients to obtain the correct confidence intervals.[24]

Stricter definitions of incident use

Stricter definitions of incident use may improve the internal validity of comparative effectiveness estimates for the reasons outlined in the previous section. However, those improvements in validity entail trade-offs for applicability and precision. For example, patients starting therapy after 365 days without therapy may be at an earlier point in the natural history of their illness (or may be experiencing a milder severity) and therefore at a lower absolute risk of clinical events than that of patients who would be eligible according to a 180-day or 90-day *wash-out period*, a term borrowed from randomized controlled trial protocols.[25] The extent of any difference in the absolute risk probably depends on the indication and the duration that defined incident use. When Schneeweiss and colleagues calculated the rate of suicide and suicide attempts for all incident users of serotonin reuptake inhibitors (SSRIs), they found that the 1-year rate was slightly higher in incident users defined by 1 year without an antidepressant medication (6.03 per 1000 person years; 95% CI, 5.54–6.55) than in incident users defined by 3 years without an antidepressant medication (5.18 per 1000 person years; 95% CI, 4.65–5.75).[26]

For comparisons of first-line therapies, validity will be improved by insisting that patients have not used any therapies in the entire class or related classes (e.g., no selective SSRIs or serotonin–norepinephrine reuptake inhibitors). However, requiring patients to have no use of any therapies in the entire class—versus requiring patients to have no use of a specific product or intervention—could reduce applicability by excluding patients with more severe symptoms or patients at a later stage in their natural history. Consider the problem of comparing second-line therapies for rheumatoid arthritis, such as the initiation of TNF alpha antagonists. Most of these patients used other disease-modifying antirheumatic drugs and have now switched to a second-line therapy. In this situation, or similar scenarios, stepped-up therapy correlates with progression of the condition. Consequently, investigators should differentiate between comparative effectiveness in first-line therapy versus comparative effectiveness in second-line therapy.[27] For second-line therapy, the cohort would be

defined by using a common first-line therapy, say methotrexate in the arthritis example, and the study intervention would be the addition of or switch to a product in the class of TNF alpha antagonists. Such a comparison of incident second-line users would improve the comparability of patients' arthritis severity and progression because they required stepped-up therapy. However, the duration and intensity of the first-line therapy may be a confounder for the second-line drug analysis.

Reduced study size as a consequence of increasing restrictions

An incident user cohort based on the 365-day definition may produce a less biased estimate of comparative effectiveness, but that finding (e.g., the risk difference) may not apply to as many patients with the condition. Restricting enrollment to incident users can dramatically reduce the size of the cohort and the precision of the comparative effectiveness estimates. For example, when Schneeweiss and colleagues identified elderly patients who filled statin therapies according to a pharmacy claims database, approximately 61 000 met the definition of current use (as of the index date), but approximately 21 000 patients met the 365-day definition for incident use.[2] Investigators with secondary data sources may find that definitions requiring more than 365 days exclude another 20% or more of patients: In US insurance plans, 20% of members typically discontinue insurance coverage annually, often switching to a new insurance plan.[28] Stricter definitions of incident use may reduce the precision of comparative effectiveness estimates to a point where the confidence intervals can no longer rule out clinically important levels of harm or benefit. In some instances, stricter definitions of incident use may require multicenter studies to achieve adequate precision—especially for testing equivalence and non-inferiority hypotheses.[29]

RECOMMENDATIONS FOR REPORTING

Based on our review, we propose the following recommendations for reporting.

(1) Investigators should report whether they designed the study to compare incident users of the intervention or whether they enrolled patients according to different eligibility criteria. Those eligibility criteria and their effect on the numbers of patients should be documented in a CONSORT-style participant flow diagram that will allow readers to assess the applicability of the findings to their populations and settings.[30]

(2) Investigators should report how they defined the dates of intervention use and, for studies that followed incident users, the time used to classify patients as incident users along with the clinical rationale for that time.

(3) Investigators should report whether the baseline characteristics were measured before incident drug use or whether those characteristics may reflect the effects of the study defining interventions. For characteristics measured before incident drug use, investigators should report the timing of those measurements in relation to the start of the intervention.

(4) Investigators should conduct sensitivity analyses with varying definitions of incident use to illustrate the stability of findings with respect to validity and precision.

CONFLICTS OF INTEREST

KEY POINTS
- We take the incident user design as a reasonable default strategy because it reduces confounding and adjustment for intermediate characteristics in the causal path.
- The incident user design results in more effective propensity scores because the baseline characteristics that contribute to the score predict incident use of interventions.
- The incident user design allows investigators to identify all of the outcomes that may be related to the intervention; prevalent user designs may miss some early events.
- Although the incident user design offers many advantages, the most obvious trade-off is a loss of precision in estimates of comparative effectiveness; a multisite study may be required to obtain a sufficient number of incident users.

ACKNOWLEDGEMENTS

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

A1.0. Defining incident users in prospective studies: trade-offs for internal validity and applicability

A2.0. Advantages of the incident user design for improving internal validity

A3.0. Trade-offs of the incident user design

A4.0. Incident user designs in the context of other design and analysis choices

A5.0. Priority research gaps

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

REFERENCES

1. Vandenbroucke JP. When are observational studies as credible as randomized trials? *Lancet* 2004; **363**: 1728–1731.
2. Schneeweiss S, Patrick AR, Stürmer T, *et al*. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Medical Care* 2007; **45**(Suppl 2): S131–S142.
3. Perrio M, Waller PC, Shakir SAW. An analysis of the exclusion criteria used in observational pharmacoepidemiological studies. *Pharmacoepidemiol Drug Safety* 2007; **16**: 329–336.
4. Friedman LM, Furberg CD, DeMets DL. Blindness (chapter 7). In *Fundamentals of Clinical Trials* (4th edn). Springer: New York, 2010; 119–132.
5. Ray WA. Evaluating medication effects outside of clinical trials: new user designs. *Am J Epidemiol* 2003; **158**: 915–920.
6. Prentice RL, Langer R, Stefanick ML, *et al*. Combined postmenopausal hormone therapy and cardiovascular disease: toward resolving the discrepancy between observational studies and the Women's Health Initiative clinical trial. *Am J Epidemiol* 2005; **162**: 404–414.
7. Petitti DB, Freedman DA. Invited commentary: How far can epidemiologists get with statistical adjustment? *Am J Epidemiol* 2005; **162**: 415–418.
8. Hernán MA, Alonso A, Logan R, *et al*. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiol* 2008; **19**: 766–779.
9. Psaty BM, Heckbert SR, Koepsell TD, *et al*. The risk of myocardial infarction associated with antihypertensive drug therapies. *JAMA* 1995; **274**: 620–625.
10. ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs. diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA* 2002; **288**: 2981–2997.
11. Ray WA, Stein CM, Daughtery JR, Hall K, Arbogast PG, Griffin MR. COX-2 selective non-steroidal anti-inflammatory drugs and risk of serious coronary heart disease. *Lancet* 2002; **360**: 1071–1073.
12. McMahon AD, MacDonald TM. Design issues for drug epidemiology. *Br J Clin Pharmacol* 2000; **50**: 419–425.
13. Moride Y, Abenhaim L. Evidence of the depletion of susceptibles effect in non-experimental pharmacoepidemiologic research. *J Clin Epidemiol* 1994; **47**: 731–737.
14. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. Good research practices for comparative effectiveness research: Approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: The International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force—Part II. *Value Health* 2009; **12**: 1053–1061.
15. Guess HA. Exposure-time-varying hazard function ratios in case–control studies of drug effects. *Pharmacoepidemiol Drug Saf* 2006; **15**: 81–92.
16. UKPDS Group. Intensive blood-glucose control with sulfonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 1998; **352**: 837–853.
17. Feinstein AR. Clinical epidemiology III: The clinical design of statistics in therapy. *Ann Intern Med* 1968; **69**: 1287–1312.
18. Vandenbroucke JP, von Elm E, Altman DG, *et al*. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *PLoS Med* 2007; **4**: e297. DOI: 10.1371/journal.pmed.0040297.
19. Jacobus S, Schneeweiss S, Chan KA. Exposure misclassification as a result of free sample drug utilization in automated claims databases and its effect on pharmacoepidemiologic studies of selective COX-2 inhibitors. *Pharmacoepidemiol Drug Saf* 2004; **13**: 695–702.
20. Hippisley-Cox J, Coupland C. Unintended effects of statins in men and women in England and Wales: population-based cohort study using the QResearch database. *BMJ* 2010; **340**: c2197. DOI: 10.1136/bmj.c2197.
21. Vrijens B, Vincze G, Kristanto P, Urquhart J, Burnier M. Adherence to prescribed antihypertensive drug treatments: longitudinal study of electronically compiled dosing histories. *BMJ* 2008; **336**: 1114–1117.
22. Hallas J, Gaist D, Bjerrum L. The waiting time distribution as a graphical approach to epidemiologic measures of drug utilization. *Epidemiology* 1997; **8**: 666–670.
23. Hernán MA, Cole SR, Margolick J, Cohen M, Robins JM. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiol Drug Saf* 2005; **14**: 477–491.
24. Stürmer T, Glynn RJ, Kliebsch U, Brenner H. Analytic strategies for recurrent events in epidemiologic studies: background and application to hospitalization risk in the elderly. *J Clin Epidemiol* 2000; **53**: 57–64.
25. Knipschild P, Leffers P, Feinstein AR. The qualification period. *J Clin Epidemiol* 1991; **44**: 461–464.
26. Schneeweiss S, Patrick AR, Solomon DH, *et al*. Variation in the risk of suicide attempts and completed suicides by antidepressant agents in adults: a propensity score-adjusted analysis of 9 years' data. *Arch Gen Psych* 2010; **67**: 497–506.
27. Solomon DH, Lunt M, Schneeweiss S. The risk of infection associated with TNF antagonists: making sense of epidemiologic evidence. *Arthritis Rheum* 2008; **58**: 919–928.
28. Short PF, Graefe DR, Schoen C. Churn, churn, churn: how instability of health insurance shapes America's uninsured problem. Issue brief, The Commonwealth Fund. New York, NY, 2003.
29. Fleming TR. Current issues in non-inferiority trials. *Statist Med* 2008; **27**: 317–332.
30. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; **340**: c332. DOI: 10.1136/bmj.c332.