Readers guide to critical appraisal of cohort studies: 3. Analytical strategies to reduce confounding

Sharon-Lise T Normand, Kathy Sykora, Ping Li, Muhammad Mamdani, Paula A Rochon and Geoffrey M Anderson

BMJ 2005;330:1021-1023
doi:10.1136/bmj.330.7498.1021

Updated information and services can be found at:
http://bmj.com/cgi/content/full/330/7498/1021

These include:

Data supplement
"Results of propensity score analysis"
http://bmj.com/cgi/content/full/330/7498/1021/DC1

References
This article cites 7 articles, 4 of which can be accessed free at:
http://bmj.com/cgi/content/full/330/7498/1021#BIBL

Rapid responses
You can respond to this article at:
http://bmj.com/cgi/eletter-submit/330/7498/1021

Email alerting service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Topic collections
Articles on similar topics can be found in the following collections
• Other Statistics and Research Methods: descriptions (525 articles)

Notes

To order reprints of this article go to:
http://www.bmjournals.com/cgi/reprintform

To subscribe to BMJ go to:
http://bmj.bmjjournals.com/subscriptions/subscribe.shtml


Readers guide to critical appraisal of cohort studies: 3. Analytical strategies to reduce confounding
Sharon-Lise T Normand, Kathy Sykora, Ping Li, Muhammad Mamdani, Paula A Rochon, Geoffrey M Anderson

Analytical strategies can help deal with potential confounding but readers need to know which strategy is appropriate.

The previous articles in this series1 2 argued that cohort studies are exposed to selection bias and confounding, and that critical appraisal requires a careful assessment of the study design and the identification of potential confounders. This article describes two analytical strategies—regression and stratification—that can be used to assess and reduce confounding. Some cohort studies match individual participants in the intervention and comparison groups on the basis of confounders, but because matching may be viewed as a special case of stratification we have not discussed it specifically and details are available elsewhere.3 4 Neither of these techniques can eliminate bias related to unmeasured or unknown confounders. Furthermore, both have their own assumptions, advantages, and limitations.

Regression Regression uses the data to estimate how confounders are related to the outcome and produces an adjusted estimate of the intervention effect. It is the most commonly used method for reducing confounding in cohort studies. The outcome of interest is the dependent variable, and the measures of baseline characteristics (such as age and sex) and the intervention are independent variables. The choice of method of regression analysis (linear, logistic, proportional hazards, etc) is dictated by the type of dependent variable. For example, if the outcome is binary (such as occurrence of hip fracture), a logistic regression model would be appropriate; in contrast, if the outcome is time to an event (such as time to hip fracture) a proportional hazards model is appropriate.

Regression analyses estimate the association of each independent variable with the dependent variable after adjusting for the effects of all the other variables. Because the estimated association between the intervention and outcome variables adjusts for the effects of all the measured baseline characteristics, the resulting estimate is called the adjusted effect. For example, regression could be used to control for differences in age and sex between two groups and to estimate the intervention effect adjusted for age and sex differences.

Results of propensity score analysis are on bmj.com

The main advantage of regression techniques is that they use data from all the participants. In addition, most researchers are familiar with these techniques and the analysis can be done using readily available software.

The validity of results from regression techniques rests on specific assumptions. A detailed discussion of these assumptions is beyond the scope of this article, but two are particularly relevant when estimating an intervention effect. Firstly, commonly used regression models assume that the intervention effect will be constant across subgroups defined by baseline characteristics. If the intervention effect differs—for example, between men and women—an interaction or effect modification is said to occur between the intervention and sex. When the effects are different across groups, separate effect estimates should be calculated through inclusion of interaction terms.

Secondly, the regression based estimate of an intervention effect involves some extrapolation. Extrapolation means that the estimate involves prediction of the effect across combinations of baseline variables that may not be observed in the data. The greater the degree of overlap in baseline characteristics between the intervention and comparison groups, the less extrapolation there is. However, the extent of this extrapolation, and the fact that it may put the analysis on shaky ground, is not always clear to the reader.

### Stratification

Stratification is a process in which the sample is divided into subgroups or strata on the basis of characteristics that are believed to confound the analysis. The effects of the intervention are then measured within each subgroup. The goal of stratification is to create subgroups that are more balanced in terms of confounders. If age and sex were confounders, then strata based on age and sex could be used to control for confounding. The intervention effect is calculated by working out the difference in average outcomes between the intervention and comparison groups within each stratum. It is important to determine whether the relation between the intervention and outcome differs across strata. If the effect estimates are the same across strata, a summary estimate can be calculated by pooling the individual estimates. However, substantial differences in estimates across strata suggest effect modification, and a summary estimate should not be calculated.

Stratification has the advantage of creating subgroups that are more similar in terms of the baseline characteristics than the entire population, and this can result in less biased estimates of the intervention effect. However, stratification may reduce the power of the study to detect intervention effects because the total number of participants in each stratum will be reduced. Another limitation is that subgroups may not be balanced with respect to baseline risk factors, in which case the estimates of the intervention effect could still be biased. For this reason, stratification is often combined with regression techniques.

Tables 1 and 2 present estimates of the association between antipsychotic use and hip fracture obtained in two comparisons in the Ontario cohort used in the earlier articles in this series. The results for both comparisons were estimated by regression and stratification strategies.

### Assessing analytical strategies

Critical appraisal of observational cohort studies requires a basic understanding of regression and stratification methods, the assumptions they rely on, and their advantages and limitations (table 3). The strategies described here may reduce confounding but cannot eliminate it entirely. Readers should ask three questions when assessing the results of a cohort study.

- **Are the analytical strategies clearly described?** The methods section should be clear enough for readers to determine which analytical strategy (such as regression or stratification) was used and how specific confounders were incorporated. For example, if regression is used, it is important to know which variables were included in the model and how these variables were related to the outcome. If stratification is used, it is important to know the variables that were included to define the strata. It is also important to assess the appropriateness of the analytical strategy in terms of the assumptions associated with the approach.

- **Do different analytical strategies give consistent results?** Both analytical strategies are designed to identify and reduce confounding but they use different techniques and are based on different assumptions. Use of more than one analytical strategy can be useful. Although obtaining similar results with different analytical strategies does not guarantee that confounding has been reduced, it does provide some support for the results. In contrast, when different analytical strategies give different results, it may be useful to review the limitations, advantages, and assumptions of each strategy.

An important step in assessing results of regression analyses is to compare adjusted and unadjusted estimates.
estimates of the effect. If the adjusted and unadjusted intervention estimates differ greatly, it implies that differences in baseline characteristics have had a substantial effect on the outcome. Table 1 shows a large difference between the unadjusted and adjusted odds ratio estimates for hip fracture in the total population (10.7 to 2.2). This suggests that the large differences in the distribution of baseline characteristics were a source of confounding. In contrast, the comparison restricted to patients with dementia in table 2 produces similar unadjusted and adjusted odds ratio estimates.

Most regression models assume a constant relationship between the outcome and intervention across all baseline characteristics, and stratification provides a technique for examining this assumption. In table 1, the odds ratios for hip fracture differ greatly across the four age-sex strata (unadjusted odds ratio from 23.14 to 5.19 and adjusted odds ratio from 1.95 to 4.11). These differences suggest an effect modification between use of atypical antipsychotics and age and sex. Stratified analyses using propensity score methods show similar results (see bmj.com).

**Are the results plausible?**

Because cohort studies are subject to confounding from unmeasured or unknown confounders, it is always unclear whether efforts to control confounding through design (such as a randomised controlled design) or through more complete or accurate measurement and adjustment of confounders would give a different result. One approach to answering this question is to determine the sensitivity of the results to unmeasured confounders. This type of sensitivity analysis is informed by a review of the literature to determine the size of the effects of known potential confounders, the size of the effects measured in the study, and the prevalence of potential confounders. The sensitivity analysis uses simulations that provide direct estimates of the size and degree of imbalance of the “unmeasured” confounder needed to negate the results of the study. If the study results are sensitive to a small amount of bias, it is important to consider the extent to which confounders were taken into account in the analysis at the design or analysis stage.

The biological plausibility of the results is also an important consideration. This is a complex question, and the issues will vary from study to study. In the study of the relation between antipsychotic use and hip fracture, the drugs could alter the risk of falls (and therefore the risk of hip fracture) through several mechanisms. These include sedation, changes in muscle rigidity, changes in balance, and cardiac effects such as hypotension and arrhythmia.

The results of any study should also be placed in the context of other similar studies including previous observational studies or randomised controlled trials. In the example study, previous studies of psychoactive drugs and hip fracture have shown similar sized effects.

**Concluding remarks**

Randomised controlled trials and cohort studies are both subject to problems related to the consistent definition of interventions and outcomes. However, only cohort studies are subject to selection bias and confounding due to differences in baseline characteristics between the intervention and comparison groups. The questions defined in this series provide a systematic approach that a reader can use to critically appraise the design, content, and analysis of a cohort study.

We thank Jennifer Gold, Monica Lee, and Michelle Laxer for help in preparing this manuscript.

Contributors and sources: The series is based on discussions that took place at regular meetings of the Canadian Institutes for Health Research chronic disease new emerging team. SLTN is a senior biostatistician with extensive experience in theoretical and practical issues related to the design, analysis, and interpretation of cohort studies who wrote the first draft of this paper and is the guarantor. PAR and MM commented on drafts of this paper. KS and PL programmed and conducted analyses. PAR and GMA conceived of the idea for the series, worked on drafts of this paper, and coordinated the development of the series.

Funding: This work was supported by a Canadian Institutes for Health Research (CIHR) operating grant (CIHR No. MOP 53124) and a CIHR chronic disease new emerging team programme (NEI-54010).

Competing interests: None declared.

3 Evans S. Matched cohorts can be useful [commentary to Helms M et al. Short and long term mortality associated with foodborne bacterial gastrointestinal infections: registry based study]. BMJ 2003;326:360.