13 Department of Health. Tackling health inequalities: a programme for action. London: DoH, 2003. (www.dh.gov.uk/assetRoot/04/01/93/62/04019362.pdf)

14 Exworthy M, Stuart M, Blane D, Marmot M. *Tackling health inequalities since the Acheson inquiry*. Bristol: Policy Press, 2003.

15 Department of Health. Health inequalities—national targets on infant mortality and life expectancy—technical briefing (revised 2002) www.dh.gov.uk/assetRoot/04/07/78/96/04077896.pdf (accessed 11 Apr 2005).

16 Social Exclusion Unit. *Bringing Britain together: a national strategy for neighbourhood renewal*. London: Stationery Office, 1998. (Cm 4045.)

17 Office for National Statistics. Life expectancy at birth by health and local authorities in the United Kingdom 1991-1993 to 2001-2003, including revised results for England and Wales 1991-1993 to 2000-2002, 2004. www.statistics.gov.uk/statbase/Product.asp?vlnk=8841 (accessed 11 Apr 2005).

18 Office for National Statistics. Population estimates for England and Wales. www.statistics.gov.uk/statbase/Product.asp?vlnk=601&More=N (accessed 11 Apr 2005). (Includes link to data for Scotland and Northern Ireland.)

19 Bajekal, M. Healthy life expectancy by area deprivation: magnitude and trends in England, 1994–1999. *Health Stat Q* 2005;25:18-27.

20 Low A, Low A. Measuring the gap: quantifying and comparing local health inequalities. *J Public Health* 2004;26:388-95.

21 Gordon D. Census based deprivation indices: their weighting and validation. *J Epidemiol Community Health* 1995;49(suppl 2):S39-44.

22 Dorling D, Thomas B. Appendix. In: *People and places: a 2001 census atlas of the UK*. Bristol: Policy Press, 2004.

23 Dorling D, Mitchell R, Shaw M, Orford S, Davey Smith G. The Ghost of Christmas Past: the health effects of poverty in London in 1896 and 1991. *BMJ* 2000;321:1547-51.

24 Wagstaff A, Paci P, van Doorslaer E. On the measurement of inequalities in health, *Soc Sci Med* 1991;33:545-57.

25 Dorling D. *Death in Britain: how local mortality rates have changed: 1950s-1990s*. York: Joseph Rowntree Foundation, 1997.

26 Szreter S, Mooney G. Urbanization, mortality, and the standard of living debate: new estimates of the expectation of life at birth in nineteenth century British cities. *Economic History Review* 1998;51:84-112.

27 Goodman A, Shephard A. *Inequality and living standards in Great Britain: some facts*. London: Institute for Fiscal Studies, 2002.

28 Cabinet Office. *Strategic audit: progress and challenges for the UK*. London: Prime Minister's Strategy Unit, 2005.

29 Dorling D, Thomas B. Know your place: housing wealth and inequality in Great Britain 1980-2003 and beyond. The Shelter policy library, 26 November 2004. http://england.shelter.org.uk/files/docs/7970/Knowyourplace.pdf (accessed 11 Apr 2005).

30 Hills J. Policy challenges and dilemmas for the next 20 years. In: Glennerster H, Hills J, Piachaud D, Webb J, eds. *One hundred years of poverty and policy*. York: Joseph Rowntree Foundation, 2004.

31 Morris J, Donkin A, Wonderling D, Wilkinson P, Dowler E. A minimum income for healthy living. *J Epidemiol Community Health* 2000;54:885-9.

32 Morris J, Deeming C. Minimum incomes for health living (MIHL): next thrust in UK social policy? *Policy Politics* 2004;32:441-54.

33 Bromley C. Has Britain become immune to inequality? In: Park A, Curtice J, Thomson K, Jarvis L, Bromley C, eds. *British social attitudes: continuity and change over two decades*. London: Sage, 2003.

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 series[1][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

Stratification of the cohort helps minimise bias

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.

P+ Results of propensity score analysis are on bmj.com

Kunin-Lunenfeld
Applied Research
Unit, Baycrest
Centre for Geriatric
Care, Toronto
Paula A Rochon
senior scientist

Department of
Health Policy,
Management and
Evaluation, Faculty
of Medicine,
University of
Toronto, Toronto
Geoffrey M
Anderson
*chair in health
management
strategies*

Correspondence to:
G M Anderson,
Institute for Clinical
Evaluative Sciences,
2075 Bayview
Avenue, Toronto,
ON, Canada
M4N 3M5
geoff.anderson@
utoronto.ca

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.[5] 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.[1][2] 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

Table 1 Unadjusted and regression adjusted odds ratio for hip fracture comparing atypical antipsychotic drugs with no antipsychotic in all older people and with stratification for age and sex

| | No of participants | | | |
	Atypical antipsychotic	No antipsychotic	Unadjusted odds ratio (95% CI)	Regression adjusted odds ratio (95% CI)
All participants	34 960	1 251 435	10.72 (10.18 to 11.30)	2.22 (2.09 to 2.36)
Age 66-75:				
Men	4 417	355 755	23.14 (18.92 to 28.31)	3.93 (2.69 to 5.74)
Women	5 345	418 235	15.48 (13.31 to 18.00)	4.11 (3.17 to 5.33)
Age ≥76:				
Men	8 823	180 851	7.92 (7.03 to 8.93)	2.53 (2.16 to 2.97)
Women	16 375	296 594	5.19 (4.86 to 5.54)	1.95 (1.78 to 2.13)

Table 2 Unadjusted and regression adjusted odds ratios for hip fracture comparing atypical antipsychotic drugs with typical antipsychotic drugs in patients with dementia and with stratification for age and sex

| | No of participants | | | |
	Atypical antipsychotic	Typical antipsychotic	Unadjusted odds ratio (95% CI)	Regression adjusted odds ratio (95% CI)
All participants	21 427	33 263	0.46 (0.44 to 0.50)	0.46 (0.43 to 0.49)
Age 66-75:				
Men	2 107	3 220	0.48 (0.36 to 0.63)	0.51 (0.35 to 0.73)
Women	2 297	3 374	0.42 (0.34 to 0.57)	0.45 (0.35 to 0.56)
Age ≥76:				
Men	5 914	9 892	0.46 (0.40 to 0.59)	0.45 (0.39 to 0.52)
Women	11 109	16 777	0.47 (0.43 to 0.51)	0.47 (0.43 to 0.51)

Table 3 Advantages and disadvantages of analytical strategies

	Regression	Stratification
Advantages	Familiar to researchers	Can focus on key confounders
	Uses all the data	Can be easily used to assess presence of effect modification
	Standard software available	Standard software available
Disadvantages	Comparability of treatment groups difficult to assess	Imbalance may still be present within strata
	Involves extrapolation	Can reduce power

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 v 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 relation 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.[6 7] 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 trial. In the example study, previous studies of psychoactive drugs and hip fracture have shown similar sized effects.[8]

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.

1 Rochon PA, Gurwitz JH, Sykora K, Mamdani M, Streiner DL, Garfinkel S, et al. Readers guide to the critical appraisal of cohort studies: 1. Role and design. *BMJ* 2005;330:895-7.
2 Mamdani, M, Sykora K, Li P, Normand SLT, Streiner DL, Austin PC, et al. Reader's guide to the critical appraisal of cohort studies: 2. Assessing potential for confounding. *BMJ* 2005;330:960-2.
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.
4 Greenlander S, Morgenstern H. Matching and efficiency in cohort studies. *Am J Epidemiol* 1990;131:151-9.
5 Rosner B. *Fundamentals of biostatistics.* 5th ed. Pacific Grove, CA: Duxbury Press, 2000.
6 Rosenbaum PR. Sensitivity analyses for certain permutation inferences in matched observational studies. *Biometrika* 1987;74:13-26.
7 Schneeweiss S, Wang PS. Association between SSRI use and hip fractures and the effect of residual confounding bias in claims database studies. *J Clin Psychopharmacol* 2004;24:632-8.
8 Ensrud KE, Blackwell T, Mangione CM, Bowman PJ, Bauer DC, Schwartz A, et al. Central nervous system active medications and risk for fractures in older women. *Arch Intern Med* 2003;163:949-57.

(Accepted 18 February 2005)

Key questions

Are the analytical strategies clearly described?

Do different analytical strategies used yield consistent results?

Are the results plausible?