Reader's guide to critical appraisal of cohort studies: 1. Role and design.

Paula A. Rochon  
Baycrest Centre for Geriatric Care

Jerry H. Gurwitz  
University of Massachusetts Medical School, Jerry.Gurwitz@umassmed.edu

Kathy Sykora  
University of Toronto

See next page for additional authors
Reader's guide to critical appraisal of cohort studies: 1. Role and design.

Authors
Paula A. Rochon, Jerry H. Gurwitz, Kathy Sykora, Muhammad M. Mamdani, David L. Streiner, Susan Garfinkel, Sharon-Lise T. Normand, and Geoffrey M. Anderson

Rights and Permissions
Citation: BMJ. 2005 Apr 16;330(7496):895-7. Link to article on publisher's website

This article is available at eScholarship@UMMS: http://escholarship.umassmed.edu/meyers_pp/62
Valid evidence on the benefits and risks of healthcare interventions is essential to rational decision making. Randomised controlled trials are considered the best method for providing evidence on efficacy. However, they face important ethical and logistical constraints and have been criticised for focusing on highly selected populations and outcomes. Some of these problems can be overcome by cohort studies. Cohort studies can be thought of as natural experiments in which outcomes are measured in real world rather than experimental settings. They can evaluate large groups of diverse individuals, follow them for long periods, and provide information on a range of outcomes, including rare adverse events. However, the promise of cohort studies as a useful source of evidence needs to be balanced against concerns about the validity of that evidence.

In this three paper series we will provide an approach to the critical appraisal of cohort studies. This article describes the role and design of cohort studies and explains how selection bias can confound the relation between the intervention and the outcome. The second article will outline strategies for identification and assessment of the potential for confounding, and the third article describes statistical techniques that can be used to deal with confounding. Each paper defines a set of questions that, taken together, can provide readers with a systematic approach to critically assessing evidence from cohort studies.

Randomised trial or cohort study?

Cohort studies are similar to randomised controlled trials in that they compare outcomes in groups that did and did not receive an intervention. The main difference is that allocation of individuals is not by chance. Table 1 gives some important similarities and differences between the two types of study. Because they are expensive and recruiting patients can be difficult, randomised controlled trials are generally short term and used to determine efficacy in selected populations under strict conditions. Cohort studies can be used to determine if the efficacy observed in randomised trials translates into effectiveness in broader populations and more realistic settings and to provide information on adverse events and risks.

Selection bias as a threat to validity

The internal validity of a study is defined as the extent to which the observed difference in outcomes between the two comparison groups can be attributed to the intervention rather than other factors. The biggest advantage of randomised controlled trials compared with cohort studies is that the random allocation process enhances the internal validity of a study by minimising selection bias and confounding. This paper relies on the definitions provided by CONSORT (box 1).

Allocation by chance in a randomised controlled trial should mean that the groups being compared are similar in terms of both measured and unmeasured baseline factors. This is not so in cohort studies, and therefore cohort studies are vulnerable to selection bias. In cohort studies, factors that determined whether a person received the intervention could result in the groups differing in factors related to the outcome, either because people were preferentially selected to receive one treatment or because of choices that they made. These baseline differences in prognosis could confound the assessment of the effect of the intervention.

In cohort studies care must be taken to minimise, assess, and deal with selection bias. A comprehensive

Cohort studies can use diverse populations

This is the first of three articles on appraising cohort studies

Kunin-Lunenfeld Applied Research Unit, Baycrest Centre for Geriatric Care, Toronto, ON, Canada
Paula A Rochon senior scientist
Jerry H Gurwitz executive director Institute for Clinical Evaluative Sciences, Toronto, ON, Canada
Kathy Sykora senior biostatistician
Muhammad Mamdani senior scientist
Susan Garfinkel research coordinator
Department of Psychiatry, University of Toronto, Toronto, ON, Canada
David L Streiner professor
Department of Health Care Policy, Harvard Medical School, Boston, USA
Sharon-Lise T Normand professor of health care policy (biostatistics)
Table 1 Comparison of cohort studies and randomised controlled trials

<table>
<thead>
<tr>
<th>Item</th>
<th>Cohort studies</th>
<th>Randomised controlled trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Populations studied</td>
<td>Diverse populations of patients who are observed in a range of settings</td>
<td>Highly selected populations recruited on the basis of detailed criteria and treated at selected sites</td>
</tr>
<tr>
<td>Allocation to the intervention</td>
<td>Based on decisions made by providers or patients</td>
<td>Based on chance and controlled by investigators</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Can be defined after the intervention and can include rare or unexpected events</td>
<td>Primary outcomes are determined before patients are entered into study and are focused on predicted benefits and risks</td>
</tr>
<tr>
<td>Follow-up</td>
<td>Many cohort studies rely on existing experience (retrospective studies) and can provide an opportunity for long follow-up</td>
<td>Prospective studies often have shorter follow-up because of costs and pressure to produce timely evidence</td>
</tr>
<tr>
<td>Analysis</td>
<td>Sophisticated multivariate techniques may be required to deal with confounding</td>
<td>Analysis is straightforward</td>
</tr>
</tbody>
</table>

Does the comparison make clinical sense?
The clinical relevance of comparisons needs to be assessed for each case. In the analysis of antipsychotic use and hip fracture, for instance, all four types of comparison might be relevant. However, this might not be true in other analyses. For example, although it would be possible for a cohort study to compare HIV positive patients receiving antiretroviral therapy with those receiving no intervention, this comparison would be irrelevant to many clinicians. A more relevant cohort study would compare patients receiving one antiretroviral therapy with patients receiving another intervention. In contrast, a clinically relevant study of the adverse effects of a commonly used treatment such as a non-steroidal anti-inflammatory drug might include a comparison with a no intervention population since no drug treatment could be a realistic option for some people.

Cohort studies should not only describe the populations being compared but also include a discussion of the clinical context for that comparison and provide a justification for the comparison. Readers of these studies should determine if the study makes a comparison that is realistic and relevant to their decision needs.

Box 1: CONSORT definitions of selection bias and confounding

**Selection bias**—a systematic error in creating intervention groups, causing them to differ with respect to prognosis. The groups differ in measured or unmeasured baseline characteristics because of the way in which participants were selected for the study or assigned to their study groups.

**Confounding**—a situation in which the estimated intervention effect is biased because of some difference between the comparison groups apart from the planned interventions such as baseline characteristics, prognostic factors, or concomitant interventions. For a factor to be a confounder, it must differ between the comparison groups and predict the outcome of interest.

Box 2: Possible types of comparisons in cohort study

**General population**
1. Intervention vs alternative intervention
2. Intervention vs no intervention

**Restricted population**
1. Intervention vs alternative intervention
2. Intervention vs no intervention
Selection bias are clear—for example, if access to atypical bias when evaluating a cohort study. Some sources of comparison groups will have on potential selection the results.

An inevitable consequence of restriction is reduced sample size. In the example, the sample decreased from 1.3 million to about 80 000 when the dementia restriction was applied. When smaller databases are being used, restriction can greatly limit the power of the study. Restriction on the basis of clinical character-

What are the potential selection biases?

Selection bias occurs when there is something inherently different between the groups being compared that could explain differences in the observed outcomes. One powerful strategy to minimise selection bias is to restrict inclusion in the study to those with a defined diagnosis or specific characteristics. Restricting the groups to a specific characteristic removes the potential for bias related to that characteristic and can reduce differences in related characteristics. Table 2 presents data from a cohort of older adults given atypical antipsychotics and a no intervention comparison group. Patients taking atypical antipsychotics were over 12 times more likely (63.1% v 4.7%) to have dementia. Dementia is related to the risk of hip fracture, and this imbalance may be an important source of confound-

What are the potential selection biases?

Relevant is the extent and nature of the restriction on the basis of clinical characteristics. The more restrictive the population, the less generalisable the results.

It is important to keep in mind the effect the choice of comparison groups will have on potential selection bias when evaluating a cohort study. Some sources of selection bias are clear—for example, if access to atypical antipsychotics was limited to patients of specialists this could result in patients who received these drugs being different from those who did not. Some sources of bias may be more subtle. For example, if doctors thought that atypical antipsychotics had fewer side effects than typical antipsychotics, they might preferentially use the atypical antipsychotics in frailter patients. This form of selection bias, referred to as channelling bias or confounding by indication, occurs when patients are assigned to one intervention or another on the basis of prognostic factors and is key issue in cohort studies.

Readers should recognise the potential for selection bias in all cohort studies and carefully consider possible sources of bias. In the next article we will outline the link between selection bias and confounding and describe a strategy for identifying and assessing the potential for confounding.

We thank Andreas Lanpas for his comments and Jennifer Gold, Michelle Laxer, and Monica Lee for help in preparing the manuscript.

Contributors and sources: The series is based on discussions that took place at regular meetings of the Canadian Institute for Health Research chronic disease new emerging team. PAR is a geriatrician with extensive research experience in cohort studies of prescription drugs who wrote the first draft of this article and is the guarantor. JHG and MM are clinicians and researchers and SELIN and DLS are statisticians who commented on drafts of this paper. KS programmed and conducted analyses and SG conducted literature searches and reviews. PAR and GMA conceived the idea for the series and GMA worked on drafts of this article and coordinated the development of the series.

Funding: This work was supported by a CIHR operating grant (CIHR No. MOP 533124) and a CIHR chronic disease new emerging team programme (NET-54010).

Competing interests: None declared.

Key questions

- What comparison is being made?
- Does the comparison make clinical sense?
- What are the potential selection biases?

Table 2 Effect on age distribution and sample size of restricting comparison of atypical antipsychotic with no intervention to individuals with dementia

<table>
<thead>
<tr>
<th>All older people</th>
<th>Older people with dementia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Atypical antipsychotic (n=34 960)</td>
</tr>
<tr>
<td>Mean (SD) age</td>
<td>80.46 (7.63)</td>
</tr>
<tr>
<td>No (%) with dementia</td>
<td>21 427 (61.3)</td>
</tr>
</tbody>
</table>