Limitations of analyses of effectiveness using observational data

Barnaby C. Reeves*

Bristol Heart Institute, Clinical Science South Bristol, Level 7, Bristol Royal Infirmary, Marlborough Street, Bristol, UK

Online publish-ahead-of-print 19 June 2006

This editorial refers to 'Quality of life after coronary revascularization in the elderly'¹ by M.M. Graham et al., on page 1690.

Graham and colleagues report health-related quality of life (HRQoL) in a cohort of patients 1 and 3 years after undergoing cardiac catheterization.¹ The paper focuses on differences in HRQoL between groups of elderly patients (all >70 years of age) who were subsequently revascularized and those who received medication only. The former had consistently better HRQoL at both times, with a tendency for patients who had bypass surgery to report better health than those who had percutaneous intervention. The paper contributes to the growing literature suggesting that coronary revascularization is under-used,²–⁴ although it remains uncertain whether under-use arises because revascularization is withheld or refused.

Increasing priority is being placed on estimating the effectiveness of interventions using outcomes that directly reflect patients’ perceptions of the value of treatment (i.e. the extent to which benefits exceed harms), especially for interventions designed primarily to reduce the symptoms and physical impairments associated with chronic disease. Physicians are often suspicious about HRQoL, which they may perceive to be ‘soft’ or subjective. This suspicion is not warranted when properly validated instruments such as the Seattle Angina Questionnaire (and, more recently, the Coronary Revascularization Outcome Questionnaire),⁵ are used in participant-blinded randomized controlled trials (RCTs), where the interpretation of differences between groups is straightforward. Rather, physicians should see these instruments as measuring directly aspects of patients’ lives, which they aim to influence by their clinical management.

Interpretation of differences in HRQoL is more complex in observational studies, such as Alberta Provincial Project for Outcomes Assessment in Coronary Heart Disease (APPROACH).¹ The authors analysed HRQoL data collected routinely, which are susceptible to three main sources of bias.

- Selection bias from confounding, as allocation to treatment was not randomized.
- Attrition bias from patients in the original cohort who did not complete the Seattle Angina Questionnaire.
- Information bias, because assessment of HRQoL was not blinded (a problem that also affects open RCTs).

In this context, it is important to examine carefully the potential consequences of these biases for interpretation of the findings. This is especially true when the effects of interest (i.e. differences in HRQoL) are small (2–3 points on a scale of 100) and of the same order of magnitude as would plausibly arise from biases, as in the analyses reported by Graham et al.¹ It is also important to distinguish estimation of effect size from the statistical significance of the effects; if biases are present, then the increased statistical precision afforded by a large sample size will simply yield a more precise estimate of the biases. As a measure of the uncertainty about the ‘true’ effect, conventional confidence intervals calculated from observational data are inevitably misleading.⁶,⁷ This is a major problem with the proliferation of analyses of large observational databases.

Selection bias is the most serious potential source of bias in the APPROACH database. This phenomenon is intuitive. Physicians try to ‘tailor’ the treatment they recommend to their patients, leading to ‘confounding by indication’. Patients also influence treatment decisions, not least because they value the varying profiles of benefits and harms associated with different treatments (e.g. higher short-term risk and greater longer-term benefit vs. lower short-term risk but greater risk of continuing or recurrence symptoms) in different ways. The important point is that confounding from selection bias cannot be eliminated. Statistical analyses that adjust for imbalances in known risk factors between groups may reduce bias, but cannot remove it,⁸ not least because some factors that are important in determining treatment are likely to be unknown or difficult to characterize. RCTs, in which concealment of allocation was inadequate or uncertain, provide a vivid illustration of the magnitude of bias that can be introduced by selection, i.e. overestimation of treatment effects by 30%.⁹ Confounding undermines statistical inference because differences in prognostic factors between groups cannot be assumed to have arisen by chance. Confounding may or may not cause systematic bias, but inevitably increases uncertainty about the observed effect.⁶,⁷

The opinions expressed in this article are not necessarily those of the Editors of the European Heart Journal or of the European Society of Cardiology.

*Corresponding author. Tel: +44 117 928 3143; fax: +44 117 929 9737. E-mail address: barney.reeves@bristol.ac.uk

¹ doi:10.1093/eurheartj/ehl038

© The European Society of Cardiology 2006. All rights reserved. For Permissions, please e-mail: journals.permissions@oxfordjournals.org
The complexity of potential attrition bias is well illustrated by the APPROACH study.\textsuperscript{1} One year after catheterization, the response rate increased across the three age groups (from 65–78%). A possible interpretation of this observation is that non-responders, who were on average younger, tended to be patients who had better health. In contrast, the response rate at 3 years among those who responded at 1 year decreased with age (from 90–78%). An interpretation of this observation is that non-responders after 3 years, who were on average older, tended to have poorer health. What is almost certain is that patients who did not respond did not constitute a random sample of the original cohort, at any point of time. Of course, if non-response is indeed associated with health, we also cannot know the extent to which patients' current health states are the consequence of (i) their health states when they presented or (ii) changes in their health states since presentation which are contingent on the treatment received.

In a participant-blinded RCT, participants' HRQoL responses cannot be affected by information bias, provided they remain successfully blinded. In observational studies, however, participants' HRQoL responses are likely to be influenced to some extent by their satisfaction with the treatment they have received. Patients receiving medical treatment may view their health state differently depending on whether they chose this treatment or were advised by their physicians not to have coronary revascularization.

Returning to the findings of the APPROACH study, the critical question is whether the differences observed could plausibly be explained by confounding or other biases. As pointed out above, this depends on the size of the differences and the likely susceptibility of the analyses to bias, not their statistical significance. If we try to weigh up the likely biases, then:

- confounding is most likely to cause the difference to be overestimated (because participants who received medical treatment tended to have more severe disease or greater morbidity at baseline);
- non-response at 1 year is most likely to cause the difference to be underestimated (because participants with better health may be selectively missing from the groups with better health, i.e. those who underwent coronary revascularization);
- non-response at 3 years is most likely to cause the difference to be underestimated (because participants with poorer health may be selectively missing from the groups with poorer health, i.e. those who received medical treatment only);
- it is difficult to predict the likely direction of information bias but this fact alone should make us less certain about the estimated effects.

Although one can make educated guesses about the likely directions of some biases, one cannot quantify their relative contributions to the overall effect, which should again make the reader feel more uncertain about the findings. Set against this uncertainty is the observation that the HRQoL findings\textsuperscript{1} are entirely consistent with the findings of a small RCT that reported conventional clinical outcomes.\textsuperscript{3}

Inferring causality from observed associations is the bread and butter of aetiological epidemiologists and ‘consistency of findings’ is just one of nine established criteria that should be considered.\textsuperscript{10} However, epidemiologists studying aetiology often have the luxury of studying risk factors which have much larger effects than the effects typically seen when comparing alternative treatments. Also, exposure to risk factors can often be graded, allowing quantification of the dose–response relationship. When investigating effectiveness, researchers have the supreme advantage of being able to allocate exposure (i.e. treatment allocation) randomly. For all of the ethical responsibility that this carries, it is an advantage that should not be foregone lightly. Physicians seeking to inform their practice should also value the evidence they read by the same principles.

\textbf{Conflict of interest:} none declared.

\textbf{References}