

**SPH Student Poster Session, March 18, 2002
University of California, Berkeley**

**APPROACHES TO CONFOUNDING AND HETEROGENEITY IN SYSTEMATIC REVIEWS OF
OBSERVATIONAL STUDIES: A SURVEY OF HIV/AIDS LITERATURE**

Madhukar Pai, Michael McCulloch, John M. Colford Jr., *Lisa A. Bero.

Division of Epidemiology, School of Public Health, University of California, Berkeley CA 94720

*Institute for Health Policy Studies, and Department of Clinical Pharmacy, University of California, San Francisco, CA 94143

Presenter: Michael McCulloch, Doctoral Student, Division of Epidemiology

Introduction

While the summary estimate of crude (unadjusted) effect measures in a meta-analysis of randomized controlled trials (RCT) is assumed to be unconfounded (because of randomization), the same assumption cannot be made in the meta-analysis of observational studies. Frequently, crude effect measures may provide confounded estimates in observational studies. Statistical combination of crude effect measures, therefore, may lead to fairly precise but inaccurate summary estimates.

Evaluating heterogeneity is a critical component of systematic reviews (SRs). Heterogeneity is of particular importance in SRs of observational studies. It can be argued that the statistical combination of studies should not be the central focus of SRs of observational studies; rather, more can be gained by exploring the sources for heterogeneity. Empiric work has shown that evaluation of heterogeneity is not universally done and, even when done, not always reported.

We surveyed published systematic reviews (SRs) on HIV/AIDS to ascertain how these reviews detect and explore heterogeneity and account for confounding.

Methods

We identified all SRs of observational studies in HIV/AIDS published during 2001, and indexed in MEDLINE. We used an efficient search strategy developed to identify SRs using PubMed (Shojania & Bero 2001). We classified a review as a SR if it was either tagged “meta-analysis” by MEDLINE or self-identified as meta-analysis or SR by the authors. 25 met our criteria for systematic reviews. Among these, 8 used only observational studies. Two reviewers (MP & MM) evaluated these articles independently and abstracted data onto a questionnaire, based on Petitti (2001). Disagreements were resolved by consensus.

Results: Confounding

All eight SRs involved meta-analysis, and covered a range of subspecialties and journals. Three of eight reviews (Rapezzi *et al*, 2001; Cruciani *et al*, 2001; John *et al*, 2001) did not address confounding at all: no mention was made of confounding; and crude measures were combined with no justification for combining crude measures. Three reviews partially addressed confounding: one combined crude and adjusted measures with some justification (Crepaz *et al*, 2001); another combined crude measures without justification but discussed Simpson’s paradox (Ciesla *et al*, 2001); and the third review separately pooled rates across studies, with some justification (Weller *et al*, 2001). Two reviews addressed confounding comprehensively. Graham *et al*. (2001) stratified effect measures as crude or adjusted and combined them separately. They discussed the implications of using summary measures from adjusted rather than crude effects, justified their decision to summarize the data in separate estimates, and presented data about covariate adjustment in each study. Rottingen *et al*. (2001) performed separate analyses for crude and adjusted estimates and identified a form of reporting bias in studies where the associations between HIV and more than one sexually transmitted disease (STD) were evaluated. When a study identified an association between one STD and HIV, the study tended to report only an adjusted effect for this one STD and either an unadjusted or else no effect estimates for the other STDs evaluated. Combining only adjusted estimates resulted in biased summary estimates.

Results: Heterogeneity

All eight SRs involved meta-analysis and covered a range of subspecialties and journals. The table shows the varied approaches used for evaluating heterogeneity.

Study	Tested for heterogeneity?	Reported results of the tests?	Found significant heterogeneity for at least some comparisons?	Model used for combining effect measures	Justified the use of a particular model?	Formally explored reasons for heterogeneity?	Method used for exploring heterogeneity
Rapezzi <i>et al.</i> , 2001	No	NA	NA	NA	NA	NA	NA
Ciesla <i>et al.</i> , 2001	No	NA	NA	NA	NA	NA	NA
John <i>et al.</i> , 2001	Yes	No	No	Fixed effects model (FEM)	Yes	NA	NA
Crepaz <i>et al.</i> , 2001	Yes	Yes	Yes	Rosenthal's model	No	Yes	Subgroup analysis
Cruciani <i>et al.</i> , 2001	Yes	Yes	Yes	Random effects model (REM), and FEM when it differed from REM	Yes	No	NA
Rottingen <i>et al.</i> , 2001	Yes	Yes	Yes	FEM	Yes	No	NA
Weller <i>et al.</i> , 2001	Yes	Yes	Yes	Pooled results across studies	Yes	Yes	Subgroup analysis
Graham <i>et al.</i> , 2001	Yes	Yes	Yes	FEM and REM; presented REM results	Yes	Yes	Subgroup analysis
Summary proportions	6/8 (75%)	5/8 (63%)	5/8 (63%)		5/8 (63%)	3/8 (38%)	

Discussion

Confounding in the meta-analysis of observational studies presents methodological challenges. There is wide variation in how confounding is addressed in primary studies. For example, some of the studies may not perform multivariate analysis, others may perform multivariate analysis but adjust for different sets of covariates, and the primary authors may selectively report crude or adjusted measures based on statistical significance. These issues influence the data available for meta-analysis. In our survey, most reviews either did not address confounding at all or addressed it partially. Reviewers of observational studies need to be aware of confounding and need to systematically explore crude and adjusted measures separately, assess the differences, and present data on covariates adjusted for in the individual studies. They also need to justify their decisions to base their conclusions on one method over the other.

Our results confirm some of the findings of earlier empiric work (Petitti, 2001). Testing for heterogeneity was not universal. Not all SRs with significant heterogeneity explored reasons for it. None of the reviews with heterogeneity refrained from meta-analyses. No review stated the rationale for using REM or FEM methods based on consideration of the question addressed. These results suggest that authors need to be more aware of the importance of exploring heterogeneity and need to be consistent in reporting of heterogeneity. In the presence of heterogeneity, authors should not force the results into a single summary estimate. If they do choose to present summary estimates despite heterogeneity, readers must be reminded about the difficulty of interpreting any summary results.