Journal of Clinical Epidemiology 68 (2015) 35–43

Journal of Clinical Epidemiology

Cross-linked survey analysis is an approach for separating cause and effect in survey research

Donald A. Redelmeier[a,b,c,d,e,*], Deva Thiruchelvam[c], Andrew J. Lustig[c,f]

[a]*Department of Medicine, University of Toronto, 2075 Bayview Ave, Toronto, Ontario, Canada M4N3M5*
[b]*Evaluative Clinical Science, Sunnybrook Research Institute, 2075 Bayview Ave, Toronto, Ontario, Canada M4N3M5*
[c]*Institute for Clinical Evaluative Sciences in Ontario, 2075 Bayview Ave, Toronto, Ontario, Canada M4N3M5*
[d]*Division of General Internal Medicine, Sunnybrook Health Sciences Centre, G-151, 2075 Bayview Ave, Toronto, Ontario, Canada M4N3M5*
[e]*Center for Leading Injury Prevention Practice Education & Research, 2075 Bayview Ave, Toronto, Ontario, Canada M4N3M5*
[f]*Department of Psychiatry, University of Toronto, 2075 Bayview Ave, Toronto, Ontario, Canada M4N3M5*

Accepted 4 September 2014; Published online 11 October 2014

Abstract

Objectives: We developed a new research approach, called cross-linked survey analysis, to explore how an acute exposure might lead to changes in survey responses. The goal was to identify associations between exposures and outcomes while reducing some ambiguities related to interpreting cause and effect in survey responses from a population-based community questionnaire.

Study Design and Setting: Cross-linked survey analysis differs from a cross-sectional, longitudinal, and panel survey analysis by individualizing the timeline to the unique history of each respondent. Cross-linked survey analysis, unlike a repeated-measures self-matching design, does not track changes in a repeated survey question given to the same respondent at multiple time points.

Results: Pilot data from three analyses (n = 1,177 respondents) illustrate how a cross-linked survey analysis can control for population shifts, temporal trends, and reverse causality. Accompanying graphs provide an intuitive display to readers, summarize results, and show differences in response distributions. Population-based individual-level linkages also reduce selection bias and increase statistical power compared with a single-center cross-sectional survey. Cross-linked survey analysis has limitations related to unmeasured confounding, pragmatics, survivor bias, statistical models, and the underlying artifacts in survey responses.

Conclusion: We suggest that a cross-linked survey analysis may help in epidemiology science using survey data.

Keywords: Self-matched studies; Epidemiologic research designs; Evaluation studies as topic; Intervention studies; Survey analysis; Outcome and process assessment

1. Introduction

Epidemiology science helps people understand the consequences of decisions by discovering connections between exposures and outcomes. Survey research is a popular method in epidemiology science because it can provide a quick, inexpensive, practical, and safe approach for addressing important questions. Survey research, however, is often problematic for causal inference because the exposures and outcomes are measured in a simultaneous manner. Interpretation can be a major challenge because of possible reverse causality, a lack of blinding, and other confounding (Box 1). Survey research sometimes fails to convince skeptical readers about the meaning of an observed correlation and sometimes fails to replicate when later tested in experimental studies [1].

Survey research can be strengthened to help establish a causal connection between exposures and outcomes. One approach is to elicit memories of past exposures in relation to current outcomes: the main limitation is recall bias leading to fallible reporting. Another approach to strengthen survey research involves repeatedly surveying the same person at different times (self-matching): the main limitation is sample attrition due to gaps in follow-up. A third approach relies on panel data to assess the same group at different points (not necessarily the same individuals): the

Conflict of interest: None.

Funding: This project was supported by a Canada Research Chair in Medical Decision Sciences, the Canadian Institutes of Health Research, and the Determinants of Community Health Research at the University of Toronto.

* Corresponding author. Tel.: +1-416-480-6999; fax: +1-416-480-6048.

E-mail address: dar@ices.on.ca (D.A. Redelmeier).

http://dx.doi.org/10.1016/j.jclinepi.2014.09.008
0895-4356/

What is new?

- A new survey approach named the cross-linked survey analysis can identify associations between exposures and outcomes while reducing some ambiguities related to cause and effect.
- Cross-linked survey analysis differs from cohort survey analysis, panel survey analysis, and longitudinal survey analysis by individualizing the timeline to the unique history of each individual respondent.
- Cross-linked survey analysis, unlike self-matching designs, does not track the same person at multiple times for repeated responses.
- Cross-linked survey analysis provides graphical and analytical methods for identifying and reducing temporal confounding.
- A cross-linked survey analysis requires a large sample size involving a population-based community questionnaire with salient specific questions.

main limitation is confounding by time trends [2]. Many other strategies are available to strengthen survey research; no strategy is ideal for all settings; and methodological work developing new designs is a priority for future progress.

The purpose of this article is to highlight a new approach called cross-linked survey analysis. The intent of this analytical approach is to provide an additional strategy for estimating causal connections between exposures and outcomes in survey research. The strategy is illustrated using real data from the Canadian Community Health Survey, a large ongoing household survey that collects data on health determinants throughout the nation. The analysis does not address the original creation of the survey or the mode of administration, questionnaire layout, question formulation, response option structure, or other nuances of the survey content. Cross-linked survey analysis investigates data under the assumption that the available survey responses are valid and important.

2. Methods

2.1. Background

Cross-linked survey analysis was first inspired by the exposure-crossover design, a method for examining sustained changes in the risk of recurrent events [3]. The main commonality of the two approaches is in linking large databases through individual identifiers to provide a before-and-after comparison after a defined exposure. An additional commonality involves structuring observation intervals into

Box 1 Causal inference in survey research. Textbox showing hypothetical survey that reports a positive correlation between coughing and lung cancer. Seven different potential interpretations are presented, of which the first interpretation of direct causality has been confirmed from multiple other lines of evidence. The remaining six explanations are likely to be specious but cannot be excluded based on the survey data.

Survey data

A brief survey is conducted at an oncology center eliciting patients' self-report of coughing (yes/no) and diagnosis of lung cancer (yes/no). Of the 100 respondents, 70 self-reported coughing, 50 self-reported lung cancer, 40 self-reported both, and 20 self-reported neither. These data indicate a significant correlation between coughing and lung cancer (odds ratio, 2.59; 95% confidence interval: 1.08–6.25; $P = 0.029$).

Survey interpretation

The positive correlation might be interpreted in at least seven different ways:

Direct causality	Lung cancer leads to coughing
Reverse causality	Coughing leads to lung cancer
Hidden confounding	Unmeasured factors cause both coughing and lung cancer
Random chance	Spurious finding due to chance association
Self-report bias	Patients with lung cancer become mindful of their coughing
Survivor bias	Coughing reduces case fatality rates and thereby causes an increased apparent prevalence of lung cancer
Selection bias	Only patients concerned about coughing or lung cancer volunteer to participate (forced positive correlation)