ELSEVIER

Social Science & Medicine 66 (2008) 117–129

SOCIAL SCIENCE & MEDICINE

www.elsevier.com/locate/socscimed

From health research to social research: Privacy, methods, approaches

Leslie L. Roos*, Marni Brownell, Lisa Lix, Noralou P. Roos, Randy Walld, Leonard MacWilliam

Manitoba Centre for Health Policy, Department of Community Health Sciences, Faculty of Medicine, University of Manitoba, 408-727 McDermot Avenue, Winnipeg, MB, Canada R3E 3P5

Available online 4 October 2007

Abstract

Information-rich environments in Canada, Australia, and the United Kingdom have been built using record linkage techniques with population-based health insurance systems and longitudinal administrative data. This paper discusses the issues in extending population-based administrative data from health to additional topics more generally connected with well being. The scope of work associated with a multi-faceted American survey, the Panel Study in Income Dynamics (PSID), is compared with that of the administrative data in Manitoba, Canada. Both the PSID and the Manitoba database go back over 30 years, include families, and have good information on residential location. The PSID has emphasized research design to maximize the opportunities associated with expensive primary data collection. Information-rich environments such as that in Manitoba depend on registries and record linkage to increase the range of variables available for analysis.

Using new databases on education and income assistance to provide information on the whole Manitoba population has involved linking files while preserving privacy, scaling educational achievement, assessing exposure to a given neighborhood, and measuring family circumstances. Questions being studied concern the role of the socioeconomic gradient and infant health in child development, the comparative influence of family and neighborhood in later well being, and the long-term effects of poverty reduction. Issues of organization of research, gaps in the data, and productivity are discussed.

Keywords: Administrative data; Population health; Registry; Policy analysis; Social research; Canada

Introduction

Weiner (1999) has stressed how science is shaped by 'entry points' that facilitate important research. One such entry point has been the development of record linkage, "the bringing together of information from two records that are believed to relate to the same individual or family. Linkage is achieved by comparing a limited subset of the total available

*Corresponding author.
E-mail addresses: Leslie_Roos@cpe.umanitoba.ca, jo-Anne_baribeau@cpe.umanitoba.ca (L.L. Roos), Marni_Brownell@cpe.umanitoba.ca (M. Brownell), Lisa_Lix@cpe.umanitoba.ca (L. Lix), Noralou_Roos@cpe.umanitoba.ca (N.P. Roos), Randy_Walld@cpe.umanitoba.ca (R. Walld), Leonard_MacWilliam@cpe.umanitoba.ca (L. MacWilliam).

0277-9536/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.socscimed.2007.08.017

information, using specified 'linkage variables,' selected for their ability to uniquely and reliably identify an individual'' (Black & Roos, 2005). Record linkage is now being used routinely by a number of Canadian and Australian research centers to maintain registries, to manage multi-file databases, to create longitudinal histories, and to work across data sets. Linkage is critical for expanding population-based research beyond its historical 'home' with health care information to additional topics more generally connected with well being. Such linkage has proven to be efficient, relatively inexpensive, and protective of privacy, compared with approaches that would collect new information for each study question (Black & Roos, 2005; Trutwein, Holman, & Rosman, 2006).

Previous work has noted the creation of ''information-rich environments'' built using population-based health insurance systems, longitudinal administrative data, and record linkage techniques. At least six groups working in such environments (Oxford, Scotland, Western Australia, and three Canadian centres: Manitoba, Ontario, and British Columbia) have demonstrated high research productivity while maintaining privacy and confidentiality. Characteristics of these centres—including their use of data for both government-funded policy work and investigator-initiated research—have been described elsewhere (Goldacre, Griffith, Gill, & Mackintosh, 2002; Holman, Bass, Rouse, & Hobbs, 1999; Kendrick, Douglas, Gardner, & Hucker, 1998; Mekel & Shortt, 2005). These environments provide the capacity to analyze interventions longitudinally; to draw cohorts and construct control groups; to compare regions, areas and hospitals in defined populations; to combine information on patients and physicians; and to compile expenditures for different services (Roos, Menec, & Currie, 2004).

Understanding the potential of administrative data is facilitated by comparison with longitudinal clinical and survey studies carried out in several countries (Lawlor et al., 2005; Power & Hertzman, 1999; Silva & Stanton, 1996). Perhaps the most influential, widely used survey has been the Panel Study in Income Dynamics (PSID) based at the University of Michigan's Institute for Social Research. The PSID was the only social science study noted on ''the National Science Foundation's list of its 50 most significant projects in its 50 year history'' (Duncan, Hofferth, & Stafford, 2004, p. 158).

New Manitoba databases have provided opportunities for using administrative data to study health, education, labor force participation, and general well being. Despite the Manitoba Centre for Health Policy's (MCHP) many years of experience, incorporating information from the Ministry of Education, Citizenship, and Youth and from the Ministry of Family Services and Housing involved a substantial learning curve. Comparisons with the PSID highlight the issues in extending population-based administrative data to facilitate social research.

Comparing approaches

Panel study in income dynamics (PSID)

Costly primary data collection and considerable attention to design are characteristic of the PSID and other longitudinal surveys (Table 1). Beginning in 1968, the PSID has included annual or biennial waves of data collection on about 5000 nationally representative households, the inclusion of families and a second generation of respondents, corrections for immigration, the reconstruction of residential location to aid research on neighborhood effects, and the wide variety of questions asked over the years. The capacity to study the connections between generations is a particular strength of the PSID (Duncan et al., 2004).

Manitoba data

Fig. 1 outlines the data housed at MCHP with an anonymized population-based research registry playing a central role. Data sets are cleaned, checked, and transformed to become linkable using a unique encrypted identifier; these data sets are typically put together as needed for each study (Table 1). The registry includes an encrypted personal health identification number (PHIN), date-specific demographic characteristics (e.g. age and sex), residential postal code, and family composition for each registrant. The Manitoba registry, supplemented by Vital Statistics files, provides information on all registrants (approximately 2.5 million individuals), as well as their dates of arrival and departure (births, deaths, and moves) for any date since 1970. High levels of follow-up have benefited both the PSID and the Manitoba data (Duncan et al., 2004; Fitzgerald, Gottschalk, & Moffitt, 1998; Roos & Nicol, 1999). Time-sensitive