Child Abuse & Neglect 37 (2013) 120–124

Contents lists available at SciVerse ScienceDirect

Child Abuse & Neglect

# Administrative data linkage as a tool for child maltreatment research[☆]

Marni D. Brownell[a,b,*], Douglas P. Jutte[a,c,d]

[a] Manitoba Center for Health Policy, Winnipeg, Manitoba, Canada
[b] Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada
[c] University of California, Berkeley-University of California, San Francisco, Joint Medical Program, San Francisco, CA USA
[d] School of Public Health, University of California, Berkeley, CA USA

ARTICLE INFO

*Article history:*
Received 24 August 2012
Accepted 13 September 2012
Available online 20 December 2012

*Keywords:*
Child maltreatment
Population-based
Record linkage
Administrative data
Child protection

ABSTRACT

Linking administrative data records for the same individuals across services and over time offers a powerful, population-wide resource for child maltreatment research that can be used to identify risk and protective factors and to examine outcomes. Multistage de-identification processes have been developed to protect privacy and maintain confidentiality of the datasets. Lack of information on those not coming to the attention of child protection agencies, and limited information on certain variables, such as individual-level SES and parenting practices, is outweighed by strengths that include large and unbiased samples, objective measures, comprehensive long-term follow-up, continuous data collection, and relatively low expense. Ever emerging methodologies and expanded holdings ensure that research using linked population-wide databases will make important contributions to the study of child maltreatment.

Child maltreatment is associated with serious and sometimes which underscores need for effective prevention and treatment programs based on sound evidence. Large-scale, longitudinal studies provide the best evidence for informing programs and policies and are powerful tools for unraveling the complicated issues around the causal risk factors for, and the outcomes of, child abuse and neglect (MacMillan et al., 2007). Unfortunately, longitudinal collection of maltreatment data is difficult to coordinate, time-consuming, and expensive. These studies are complicated by issues of problematic sampling that can result in systematic differences at baseline, nonrandom loss-to-follow-up, and low study power if small numbers of subjects report exposure to child maltreatment (MacMillan et al., 2007). In this paper we explore a different approach to data collection and research in the area of child maltreatment that could circumvent some of these difficulties and may shed new light on risk and protective factors, as well as outcomes, associated with child maltreatment.

Linked administrative data comprise information already widely and diligently collected on large populations for other purposes that are merged at the individual level using unique, anonymized identifiers and then made available for academic research. In several jurisdictions, including Canada, Australia, Scandinavia, and the United Kingdom, research centers have used health record linkage to analyze health interventions longitudinally, create cohorts and comparison groups, and compare health measures across geographic and service regions (Roos, Menec, & Currie, 2004), while at the same time maintaining privacy and confidentiality of the data (Jutte, Roos, & Brownell, 2011; Roos et al., 2008). Details about some of these research centers and the academic and government-funded policy research they have produced can be found elsewhere (Black, Devereux, & Salvanes, 2005; Goldacre, Griffith, Gill, & Mackintosh, 2002; Holman et al., 2008; Holman, Bass, Rouse,

---

[☆] The first author acknowledges the financial support of the Government of Manitoba through the Manitoba Center for Health Policy Population-Based Child Health Research Award.

[*] Corresponding author address: Manitoba Center for Health Policy, University of Manitoba, 408-727 McDermot Avenue, Winnipeg, Manitoba, Canada R3E 3P5.

0145-2134/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.chiabu.2012.09.013

& Hobbs, 1999; Ji, Hemminki, Sundquist, & Sundquist, 2010; Kendrick, Douglas, Gardner, & Hucker, 1998; Mekel & Shortt, 2005; Roos et al., 2008; Sundquist, Johansson, Yang, & Sundquist, 2006).

Traditionally, such linkages were confined to health information, such as vital statistics, hospitalizations, physician visits, and pharmaceutical records (Roos et al., 2008). Health information (i.e., hospitalizations and Emergency Department visits for maltreatment-related injuries and vital statistics data on violent deaths) has been used to examine population-level patterns and trends in child maltreatment (Gilbert et al., 2012; O'Donnell et al., 2010; O'Donnell, Nassar, Jacoby, & Stanley, 2012). However, in more recent years, research using administrative data has extended from health information to include linkages with population-wide data from the social service and education realms, including population-based data on child protection, which can include information on allegations, substantiations, type of maltreatment, services received, and out-of-home care. Such linkages have been used to examine the relationship between Emergency Department visits for injury and child maltreatment (O'Donnell et al., 2012; Spivey, Schnitzer, Kruse, Slusher, & Jaffe, 2009), identify the frequency and types of hospitalizations for children with substantiated maltreatment (O'Donnell et al., 2010b), explore factors associated with maltreatment (O'Donnell et al., 2010a), and determine the extent of overlap across different services for children (Gilbert et al., 2012). They have also been used to study outcomes for children in out-of-home care, including involvement with justice, teen pregnancy, and unemployment (Doyle, 2007); suicide and attempted suicide (Katz et al., 2011); and high school performance and completion, teen motherhood, and income assistance receipt (Brownell et al., 2010).

Using linked population-based administrative data can supply the very large samples and long-term follow-up that primary data collection often cannot and at a fraction of the cost. By combining information derived from disparate sources, events occurring to individuals over time and across generations; the ability to do this for large populations allows for a wide range of important and often unique investigations. This type of data can also provide a relatively inexpensive supplement to prospective longitudinal studies on child maltreatment.

## Data Linkage Methodology

Data linkage refers to combining information, believed to be on the same individual, from two different records (Black & Roos, 2005). In the most straightforward linkages, a unique identifying variable is available on both data sources to be linked, such as a personal health number. In those cases the process of linking is relatively simple and involves matching records from the two data sets based on the unique identifier (Roos, Wajda, & Nicol, 1986; Roos & Wajda, 1991; Roos, Walld, Wajda, Bond, & Hartford, 1996). In practice, such unique identifiers rarely exist and linkage is achieved by using a limited set of basic sociodemographic factors, such as name, sex, and birth date, to identify uniquely and reliably an individual across two or more datasets. Because each of these linkage variables may not be unique to a particular individual (e.g., birth date), or they may change over time (e.g., name changes or nicknames), thus the set of identifying characteristics from the two sources must be compared to estimate the likelihood that the potential link is for the same individual. There is an extensive literature on data linkage methodology, particularly the linking of health records (Blakely & Salmond, 2002; Gomatam, Carter, Ariet, & Mitchell, 2002; Jaro, 1995; Méray, Reitsma, Ravelli, & Bonsel, 2007; Qayad & Zhang, 2009; Redden, Mulvihill, Wallander, & Hovinga, 2000; Roos et al., 1986; Roos & Wajda, 1991; Roos et al., 1996, 2008).

Maintaining privacy, confidentiality, and security of the data are always of utmost concern. In order to achieve high-quality linkages without compromising confidentiality and to address legal and political concerns without reducing the value of the information produced, multistage de-identification processes have been developed by various research centers (Kelman, Bass, & Holman, 2002; Roos et al., 2008; Weisbaum, Slaughter, & Collins, 2005). The details vary across sites; however in most cases, in order to enhance privacy and security of data, an outside agency serves as a linkage center.

The merging of information is completed through a series of steps. First, the linkage center receives files from data trustees (e.g., child protection or health agencies) containing only identifying information (i.e., linkage variables such as name, sex, birth date), but no associated content or program data. Second, the linkage variables are used to identify individuals with information present across datasets (e.g., across child protection and hospital data). Third, a unique record number is assigned to the identified individuals. Finally, this unique record number is sent back to the data trustees, who use it to provide the research center with the requested administrative content or program data but without the associated identifying information. Thus confidentiality is maintained because neither the linkage center nor the research center is ever in possession of both the individually identifiable linkage data and programmatic content. Only the trustee maintains those complete records (Brook, Rosman, & Holman, 2008; Holman et al., 1999; Kendrick et al., 1998; Roos et al., 2008). Investigators in the health realm have suggested that this approach can actually enhance individual confidentiality over the traditional investigator-initiated use of private health data (Roos et al., 2004). And in fact, after the introduction of a record linkage system in Western Australia, research projects using name-identified data dropped dramatically (Trutwein, Holman, & Rosman, 2006).

## Privacy and Ethical Issues

Issues of privacy, confidentiality, and security of data are a prominent concern facing the use of administrative data and linkages, whether they include health, education and/or child protection data. Most jurisdictions have legislation that protects the privacy of personal information, however the overinterpretation of protective legislation could represent a risk if it prevents high-quality research of public interest from taking place (Davies & Collins, 2006). Indeed, it has been argued