Cross-sectional reporting of previous Cesarean birth was validated using longitudinal linked data

Jian Sheng Chen[a,*], Christine L. Roberts[a], Jane B. Ford[a], Lee K. Taylor[b], Judy M. Simpson[c]

[a]Clinical and Population Perinatal Health Research, The Kolling Institute of Medical Research, University of Sydney, Sydney, New South Wales, Australia
[b]Surveillance Methods, Centre for Epidemiology and Research, NSW Department of Health, Sydney, New South Wales, Australia
[c]School of Public Health, University of Sydney, Sydney, New South Wales, Australia

Abstract

Objective: The aim of this study was to demonstrate the feasibility of using linked health records to assess data quality in population health data.

Study Design and Setting: Reproductive histories of 155,897 women were constructed by longitudinal linkage of the New South Wales (Australia) birth records in 1998–2005, and 127,952 birth and hospital discharge records in 2000–2005 were cross-sectionally linked. History of Cesarean section (CS) derived from the longitudinal linkage (''gold standard'') was used to validate the CS history fields (i.e., ''Was the last birth by Cesarean section?'' and ''Total number of previous Cesarean sections?'') in birth records and to validate ''vaginal birth after previous Cesarean (VBAC)'' and ''maternal care for uterine scar'' in hospital records.

Results: The reporting of CS at last birth was reliable with sensitivity, specificity, positive predictive value (PPV), and negative predictive value all >95% as was the number of previous CS (weighted kappa = 0.97). For the hospital data, sensitivity and PPV were 46% and 99% for VBAC, 92% and 99% for maternal care of uterine scar, and 85% and 99%, respectively, for any prior CS.

Conclusion: Assessing data quality by record linkage is feasible and can be done more quickly and cheaply than by any traditional validation study. © 2010 Elsevier Inc. All rights reserved.

Keywords: Birth records; Cesarean section; Hospital records; International classification of diseases; Medical record linkage; Reliability and validity; Vaginal birth after Cesarean

1. Introduction

Population health data (PHD), such as birth records, hospital discharge, pharmaceutical, and registry data, are collected for health surveillance, needs assessment, administrative, policy, and planning purposes. These routinely collected PHD are increasingly being used in epidemiological studies that range from identifying potential risk factors, assessing effects of health determinants, monitoring disease trends, assessing health service utilization, and evaluating intervention programs [1–5]. Readily available PHD provide a cost-efficient means of conducting health and medical research, which can be undertaken more quickly than collecting data specifically from patients. Also, extremely large data collections allow investigation of rare outcomes that would be difficult or impossible to study using traditional methods such as case–control, cross-sectional, and cohort study designs [6]. However, the usefulness of a PHD set is largely dependent on the completeness and validity of the data [7,8].

Assessment of data quality for PHD in validation studies usually requires access to and data abstraction from original data sources such as medical records. Such studies are difficult, expensive, and time consuming, and consequently, they are infrequently undertaken [9]. If the quality of the data from at least one PHD source is known, a more cost-efficient method of assessing quality in another source could be to link and compare one data source with another of known reliability.

The aim of this study was to demonstrate the feasibility of using linked health records to assess data quality in routinely collected data. Specifically, we constructed reproductive histories of women by longitudinal linkage of birth records and used the history of Cesarean section (CS) derived from this longitudinal linkage to validate (1) the

* Corresponding author. Institute of Bone and Joint Research, Royal North Shore Hospital, St Leonards, New South Wales 2065, Australia. Tel.: +61-2-99267328; fax: +61-2-99061859.

E-mail address: jschen@med.usyd.edu.au (J.S. Chen).

What is new?

- This study showed that the quality of routinely collected population health data could be assessed using record linkage.

- Using record linkage to assess data quality was reliable.

- The study was done in a short time frame with few financial resources.

- The method is likely to have more widespread application in other fields.

reporting of history of in the birth record at any single time-point and (2) the reporting of vaginal birth after previous Cesarean (VBAC) and maternal care for uterine scar in International Classification of Diseases (ICD) coded hospital discharge data through linkage of birth and hospital data.

2. Materials and methods

Two data collections were used: birth data and hospital discharge data. Birth data were used to develop a longitudinal method of identifying history of previous CS, which was then used to validate the history of previous CS recorded at each birth and in the hospital discharge record.

2.1. Data sources and data linkage

The New South Wales (NSW) Midwives Data Collection ("birth data") includes all births in the state of NSW regardless of mother's residence. Information on maternal health, the pregnancy, labor, delivery, and perinatal outcomes are collected when a live birth or a stillbirth of at least 20 weeks of gestation or at least 400 g birth weight occurs. Demographics and information about previous births are usually collected at the initial antenatal care visit. Since 1998, the birth data also include information on whether women have had a CS at the last (previous) birth and number of previous CS.

The NSW Admitted Patient Data Collection ("hospital data") covers every inpatient admission in NSW and includes demographic and episode-related data. Data from the medical records are coded according to the tenth revision of the International Classification of Diseases Australian Modification (ICD10-AM) [10]. ICD codes are available for VBAC (ICD10: O75.7) and maternal care for uterine scar (ICD10: O34.2). The latter code should be assigned to all women with a Cesarean or other operative uterine scar who have an elective Cesarean, a trial of labor that proceeds to Cesarean delivery or where the uterine scar requires care but does not proceed to delivery, for example, antepartum care for uterine pain because of

previous scar. Together, these two codes should identify all women with a prior Cesarean.

In NSW, record linkage is conducted at the Centre for Health Record Linkage (CHeReL) using probabilistic matching techniques and a system of weights to discriminate and compare patient records [11]. Personal identifiers, such as name, address, sex, date of birth, hospital code, and record number, are used for matching of records belonging to individual people. Once linked, personal identifiers (such as mother's name, address, and date of birth) are removed and a person project number (unique to each mother and project) is assigned to each record, which can be used to link individual records in a PHD set or in different PHD sets. Evaluation of the performance of matching and data linkage is determined using sensitivity analysis and manual clerical review of indeterminate matches (~2%).

2.2. Study design

For this study, birth records were longitudinally linked within the birth data from January 1, 1998 to December 31, 2005 and cross-sectionally linked to the hospital data for the period of July 1, 2000 to December 31, 2005. In birth records, CS for current birth is reported with a sensitivity and specificity of 100% [12]. So the CS history identified by the longitudinal method was considered the "gold standard." It was used to validate the quality of the data for the CS history fields (i.e., "Was the last birth by Cesarean section?" and "Total number of previous Cesarean sections?") in birth records and to validate "VBAC" and "maternal care for uterine scar" in hospital records.

In the case of multifetal pregnancy, only the last record of each "birth" was used because the delivery method for a twin birth could change from vaginal birth for the first twin to CS for the second twin. Because the reproductive histories need to be consecutive to validate CS at last birth and number of previous CS in birth records, women with nonconsecutive records were excluded (e.g., those with information on a first and third birth but not on the second) as were those without information on mode of delivery. This step ensures completeness of the reproductive histories and is also likely to improve the accuracy of the selected sample [13]. Selection procedures for the study sample are presented in Fig. 1.

2.3. Assessments

First, accuracy of reporting of a CS at last (previous) birth was assessed separately for second, third, and fourth births. Because few fifth or subsequent births were available, these were excluded. Accuracy of reporting of a CS at last birth was also assessed by year. We sought to identify factors that may contribute to the misclassification of a CS at last birth in the birth records (i.e., records for second, third, and fourth births). Factors assessed included pregnancy factors for last birth (adverse infant outcome