Linking Survey Data with Administrative Health Information

Characteristics Associated with Consent from a Neonatal Intensive Care Unit Follow-up Study

Anne F. Klassen, DPhil[1]
Shoo K. Lee, PhD[2]
Morris Barer, PhD[3]
Parminder Raina, PhD[4]

ABSTRACT

Background: Health services and population health research often depends on the ready availability of administrative health data. However, the linkage of survey-based data to administrative data for health research purposes has raised concerns about privacy. Our aim was to compare consent rates to data linkage in two samples of caregivers and describe characteristics associated with consenters.

Methods: Subjects included caregivers of children admitted at birth to neonatal intensive care units (NICU) in British Columbia and caregivers of a sample of healthy children. Caregivers were asked to sign a consent form enabling researchers to link the survey information with theirs and their child's provincially collected health records. Bivariate analysis identified sample characteristics associated with consent. These were entered into logistic regression models.

Results: The sample included 1,140 of 2,221 NICU children and 393 of 718 healthy children. The overall response rate was 55% and the response rate for located families was 67.1%. Consent to data linkage with the child data was given by 71.6% of respondents and with caregiver data by 67% of respondents. Families of healthy children were as likely to provide consent as families of NICU children. Higher rates of consent were associated with being a biological parent, not requiring survey reminders, involvement in a parent support group, not working full-time, having less healthy children, multiple births and higher income.

Conclusion: The level of consent achieved suggests that when given a choice, most people are willing to permit researcher access to their personal health information for research purposes. There is scope for educating the public about the nature and importance of research that combines survey and administrative data to address important health questions.

MeSH terms: Privacy; databases; questionnaires; neonatology

Significant components of health services and population health research depend on the ready availability of individual-level, longitudinal, linkable, de-identified, administrative health data. Combining administrative data from different sources with information from surveys on health status, demographic and socio-economic characteristics, extensively expands the potential research questions that can be explored and is critical for some forms of research. However, various concerns have been raised about the use of administrative data, and particularly the linkage of survey-based data to administrative data.

Studies that have linked survey data with administrative sources of data for health research purposes, report wide variations in rates of consent.[1-5] In order to properly address concerns about data privacy and to formulate appropriate data linkage and access policies, better information is needed about public willingness to have health data linked, and factors affecting consent to data linkage.

The aim of this paper was to compare consent rates to data linkage and describe characteristics associated with consenters using data provided by parents in a population-based sample of children admitted at birth to neonatal intensive care units (NICU) in BC in 1996-1997 and by parents of healthy children. Each child's mother was sent a questionnaire booklet and a two-page consent letter that described the study and its proposed benefits, and asked for permission to access their and their child's provincially collected health records for research purposes. Since there is little information available on characteristics of Canadian men and women who consent to such data linkage, this paper aims to fill an important gap in the literature.

DESIGN/METHODS

All residents in the province of BC participate in a government-administered, tax-financed health insurance plan. All medically necessary services provided by physicians or received in acute care hospitals are benefits under this plan. Comprehensive patient-level data on hospital and physician services are routinely collected by the province, along with other health information on individuals (e.g., vital statistics

---

La traduction du résumé se trouve à la fin de l'article.

1. Centre for Community Child Health Research, Vancouver, BC
2. Centre for Health Innovation and Improvement, Dept of Pediatrics, University of British Columbia, Vancouver, BC
3. Centre for Health Services and Policy Research, and Department of Health Care and Epidemiology, University of British Columbia
4. Evidence-Based Practice Centre, Dept of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON

Correspondence: Dr. Anne Klassen, Centre for Community Child Health Research, L408, 4480 Oak Street, Vancouver, BC, V6H 3V4, Tel: (604) 875-3712, Fax: (604) 875-3569, E-mail: afk@interchange.ubc.ca

Acknowledgements: The Hospital for Sick Children Foundation (Toronto) provided an operating grant for this study. Anne Klassen was recipient of a Killam Postdoctoral Fellowship. From Canadian Institutes of Health Research, Anne Klassen holds a Senior Research Fellowship, and Parminder Raina holds an Investigator Award. We would like to thank the Canadian Neonatal Network for providing hospital birth record information.

data). These databases have been archived back to 1985, and made linkable and accessible to researchers under carefully developed terms that protect the individual's privacy, as part of the B.C. Linked Health Database (BCLHD).[6] Researchers may link other research data with the BCLHD for research purposes, provided they have institutional ethics approval, observe a strict code governing data access, use and privacy, and have authorization from survey respondents.

Sample

Ethical approval was gained from the university and three hospitals in this study. Mothers of all babies (n=2221) admitted in 1996-97 to the three level III NICUs in BC, and a consecutively recruited sample of healthy babies (n=718) born in 1996-97 at the two hospitals with a hospital-based primary care unit (the 3rd did not have such a unit), were sent a questionnaire booklet and consent letter when their child was 42 months of age. The caregiver who spends the most time with the child was asked to complete the questionnaire, to provide their personal health number (PHN) and that of their child, and to sign the consent form enabling researchers to link the survey information with their provincially collected health records. All non-respondents were followed up with a reminder letter and additional copies of the questionnaire as needed. We excluded from the study any families: who were non-English speaking; in which the child or mother had died; or who completed the questionnaire on the wrong child.

TABLE I

Odds Ratio and 95% Confidence Intervals Comparing Caregivers Who Provided Consent for Linkage Between Questionnaire Data and Children's Health Records with Caregivers Who Did Not Provide Consent on Various Sample Characteristics

	NICU Sample		Healthy Baby Sample	
	OR	95% CI	OR	95% CI
Biological parent	3.3	1.5, 7.2	2.2	0.1, 35.3
Female	1.4	0.8, 2.6	2.2	0.6, 9.0
Married or common-law	1.1	0.8, 1.6	0.8	0.4, 1.6
Age (years)				
19-29	Reference		Reference	
30-39	0.9	0.6, 1.3	1.2	0.7, 2.2
≥40	0.7	0.4, 1.1	1.2	0.6, 2.5
Education level				
University	Reference		Reference	
Trade/technical school/community college	0.8	0.6, 1.1	0.8	0.5, 1.3
High school graduation	0.9	0.6, 1.3	0.9	0.5, 1.8
No high school diploma	0.8	0.5, 1.5	0.6	0.2, 1.7
Not working full-time for pay or profit	1.8	1.3, 2.4	1.2	0.7, 1.9
Household income				
≥$80,000	Reference		Reference	
$50-$79,999	0.8	0.5, 1.2	0.4	0.2, 0.7
$30-$49,999	0.7	0.5, 1.1	0.6	0.3, 1.2
<$30,000	0.7	0.5, 1.1	0.3	0.2, 0.7
Smoker	1.1	0.8, 1.5	1.4	0.8, 2.6
Other smoker in the house	1.3	0.7, 2.2	0.7	0.3, 1.5
Has been involved in parent support group	2.2	1.4, 3.4	2.3	0.9, 5.3
Boy	1.1	0.9, 1.5	1.1	0.7, 1.7
Firstborn	1.2	0.9, 1.6	0.9	0.6, 1.4
Multiple birth	1.7	1.2, 2.5	–	–
No reminders required	1.5	1.1, 2.0	3.0	1.8, 4.9
Moved since child's birth	1.2	0.9, 1.5	1.2	0.7, 1.9
Visited family physician in past year	1.4	1.0, 1.8	1.0	0.6, 1.6
Visited pediatrician in past year	1.0	0.8, 1.4	1.1	0.5, 2.3
Child has health status problem	1.5	1.2, 2.0	1.2	0.8, 1.9
Child behaviour problem	1.1	0.6, 1.8	1.6	0.6, 4.4

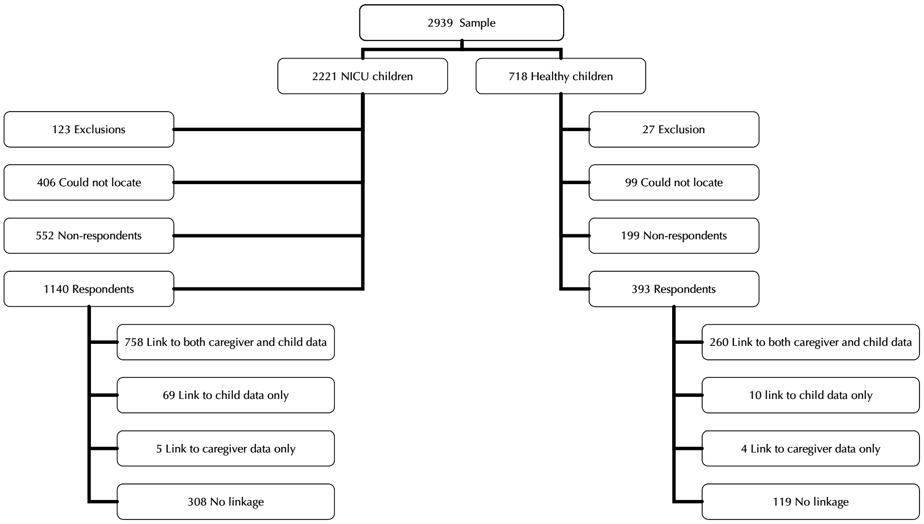

Figure 1. Flow chart for response rate for NICU and healthy child sample

Analysis

The questionnaire booklet included questions to measure socio-economic status, health status, and health care utilization. Caregiver health was measured using the SF-36, which provides a physical and psychosocial health summary score.[7,8] Family function was measured with the Family Assessment.[9] Child health was assessed using the Health Status Classification System Preschool Version (HSCS-PS),[10] which measures 12 problems (including sensory; motor; learning; behaviour). We created a variable to indicate whether the child had one or more health problems. Behaviour problems were measured with the Child Behavior Checklist/ 1.5-5.[11] Scores were recoded into either normal or borderline/clinical. Health care utilization variables were coded as follows: ≥1 visit by child to family doctor in the past year (yes or no); ≥1 visit by child to a pediatrician in the past year (yes or no); ≥1 night in hospital in the past year (yes or no). Bivariate analysis identified variables significantly associated with the outcome variable (consent to link yes or no), which were entered into logistic regression models.

RESULTS

Questionnaires and consent forms were mailed to 2,221 NICU and 718 healthy baby families. Figure 1 shows the number of exclusions, loss to follow-up, non-respondents and respondents for both samples. Overall, 50% of the sample moved at least once since the birth of their child and we located 82.8% (n=2434). The overall response rates (after exclusions) were 54.3% (NICU) and 56.9% (healthy chil-

TABLE II

Odds Ratios and 95% Confidence Intervals from Multivariate Logistic Regression Analysis for Variables Related to Consent vs. No Consent to Linkage Between Questionnaire Data and Children's Health Records

	NICU Sample	Healthy Baby Sample
Biological parent	3.7 (1.6, 8.5)	
No reminders required	1.5 (1.1, 2.1)	3.0 (1.8, 5.0)
Parent support group participant	1.9 (1.2, 3.0)	
Not working full-time	1.5 (1.1, 2.1)	
≥1 health attribute problem	1.4 (1.0, 1.8)	
Annual household income		
≥$80,000		Reference
$50-$79,999		0.4 (0.2, 0.8)
$30-$49,999		0.7 (0.3, 1.4)
<$30,000		0.4 (0.2, 0.8)

TABLE III

Odds Ratios and 95% Confidence Intervals from Multivariate Logistic Regression Analysis for Variables Related to Consent vs. No Consent to Linkage Between Questionnaire Data and Caregivers' Health Records

	NICU Sample	Healthy Baby Sample
Biological parent	2.4 (1.1, 5.3)	
No reminders required	1.6 (1.2, 2.1)	2.9 (1.7, 4.9)
Multiple birth	1.8 (1.2, 2.6)	
Parent support group participant		2.6 (1.1, 6.2)
Annual household income		
≥$80,000		Reference
$50-$79,999		0.5 (0.3, 0.9)
$30-$49,999		0.7 (0.3, 1.4)
<$30,000		0.4 (0.2, 0.7)

dren), and for located families were 67.4% (NICU) and 66.4% (healthy children).

We received 74.2% (n=1138) of consent forms back. Of these, 2.8% (n=32) of families were excluded from our linked data request for the following reasons: 26 signed but gave no PHN; 3 indicated that they did not want data linkage; 2 gave a PHN but did not sign; and 1 never had a PHN (United States citizen).

Of the 72.2% (1106) of respondents (both samples) who provided consent to data linkage and were not otherwise excluded, 9 (0.8%) caregivers gave permission for us to examine their records but not their child's, while 79 (7.1%) granted permission to link only to their child's records. Most consent forms were completed by a biological parent, most commonly (96% child and caregiver forms) the child's mother.

The proportion of parents who provided consent for themselves (66.9%) and their child (72.5%) from the NICU sample did not differ significantly from the proportion of caregivers who provided consent for themselves (67.2%) and their child (68.7%) from the healthy baby sample.

Consent to link to child data

Sample characteristics that discriminated between those who did vs. did not provide consent to access the child data for the NICU sample and healthy child sample appear in Table I. For the NICU sample, consent was higher for the following subjects: biological parent; not working full-time; involved in a parent support group; child part of a multiple birth; did not need survey reminder letters; visited the family physician in the past year; and child had at least one health status problem. For the healthy baby group, consent was higher for the following subjects: higher household income; did not need survey reminders.

For the NICU and healthy baby group, we also explored the relationship between consent to access the child data, and parental physical and psychological health and family function, but no differences were found on these variables.

Consent to link to caregiver data

We performed the same analyses as above, using as the outcome variable those who did vs. did not provide consent to access the caregiver data for the NICU sample and healthy baby sample. For the NICU sample, we found the same pattern of results (i.e., the same variables were statistically significant) except for visited a family physician in the past year, which was no longer significant for the caregiver report. For the healthy baby sample, variables associated with consent included the following: involvement in a parent support group; no reminders required; and higher household income.

Predictors of consent to link to child data

The predictors of consent to link with the child data appear in Table II. In the logistic regression model for the NICU sample that included consent to link survey data to child health records as the outcome variable, the following five variables were associated with consent: biological parent; did not need survey reminders; involvement in a parent support group; not working full-time; and ≥1 health status problems for the child. In the model for the healthy baby sample, two variables were associated with consent: did not need survey reminders; and higher household income.

Predictors of consent to link to caregiver data

The predictors of consent to link to caregiver data appear in Table III. In the logistic regression model for the NICU sample that included consent to link survey data to caregiver health records as the outcome variable, the following three variables were associated with consent: biological parent; did not need survey reminders; and multiple birth. In the model for the healthy baby sample, three variables were associated with consent: did not need survey reminders; participating in a parent support group; and higher household income.

DISCUSSION

The majority of families in our study signed a consent form permitting researchers to link their and their child's personal health records with their questionnaire survey responses. Families of healthy children were as likely to provide consent as families of NICU children. Although the consent rate was lower than the 93% achieved from Ontario respondents to the 1994 National Population Health Survey,[3,4] this is likely due to a difference in methodology – the NPHS used home-based interviews with a random sample of the population, whereas we used a postal survey with parents. Less personal mail-in surveys generally report lower rates of consent.[2,5] However, the level of consent achieved in both studies suggests that when given information and a choice, most people were not concerned about privacy as it related to the use of personal health information for research purposes.

Families who provided consent differed in a number of characteristics from those who did not. For linkage to child and caregiver data, being a biological parent, having less healthy children, not needing survey reminders, involvement in a parent support group, not working full-time, multiple births, and higher household income were all associated with higher rates of consent to data linkage.

There are a number of possible reasons that might explain these findings. That families with children with health problems were more willing to participate may be attributable to a 'motivation effect' – they may believe that the results of the research could benefit them and their children; a 'gratitude effect' – they are simply grateful for the medical attention given to their child, associate the research and the researchers with that medical attention, and want to do whatever they can to support those involved in the child's care; or a 'future benefit' effect – they do not want to see other families go through what they are enduring, and believe that the research is likely to improve the odds of better outcomes for children in similar situations, in the future. These reasons could also account for the higher rate of consent in families of multiple births, and families involved in a parent support group (many of which related to child health issues), since both factors would be associated with a higher use of health care services. Our finding of higher consent in higher income families is in agreement with Young et al.[5] who found that women who provided consent tended to be better educated. A lack of understanding, or perhaps suspicion, about health research, with the result that privacy concerns trump any views about the benefits of the research, might be more common in people in lower socio-economic groups, and might be related to an unwillingness to provide consent.

Our study has certain limitations. Although our response rate is within the range often obtained in a postal survey,[12] non-response can introduce bias. Our results may not be generalizable to all NICU families, nor to non-English-speaking families. Our healthy baby sample was not randomly selected from all low-risk births in the province.

Refusal to consent to having one's survey responses linked to administrative data for research purposes may be related to incorrect perceptions, which may be modifiable by public education about the nature and importance of health research. Important advances in our understandings of etiology, risk factors, and prognosis for disease have been made through the use of population surveillance, disease registries, and other forms of observational research.[13] This research has in common the use of databases containing information on large numbers of people (often without their consent because of the impracticality of securing consent after the fact). Those who are invited to participate in health research should be encouraged to view such invitations as opportunities to assist with research that will benefit society, rather than an invasion of personal privacy.

REFERENCES

1. Jacomb PA, Jorm AF, Korten AE, Rodgers B, Henderson S, Christensen H. GP attendance by elderly Australians: Evidence for unmet need in elderly men. *Med J Aust* 1997;166:123-26.
2. Lillard LA, Farmer MM. Linking Medicare and national survey data. *Ann Intern Med* 1997;127:691-95.
3. Rhodes AE, Lin E, Mustard CA. Self-reported use of mental health services versus administrative records: Should we care? *Int J Methods Psychiatr Res* 2002;11:125-33.
4. Finkelstein MM. Do factors other than need determine utilization of physician's services in Ontario? *CMAJ* 2001;165:565-70.
5. Young AF, Dobson AJ, Byles JE. Health services research using linked records: Who consents and what is the gain? *Aust N Z J Public Health* 2001;25:417-20.
6. Chamberlayne R, Green B, Barer ML, Hertzman C, Lawrence WJ, Sheps SB. Creating a population-based linked health database: A new resource for health services research. *Can J Public Health* 1998;89:270-73.
7. Ware JE, Kosinski M. *SF-36 Physical & Mental Health Summary Scales: A Manual for Users of Version 1*, Second Edition. Lincoln, RI: QualityMetric, 2001.
8. Ware JE, Snow KK, Kosinski M. *SF-36 Health Survey: Manual and Interpretation Guide.* Lincoln, RI: QualityMetric Incorporated, 2000.
9. Byles J, Byrne C, Boyle MH, Offord DR. Ontario Child Health Study: Reliability and validity of the general functioning subscale of the McMaster Family Assessment Device. *Fam Process* 1988;27:97-104.
10. Saigal S, Stoskopf BL, Rosenbaum PL, Hoult LA, Furlong W, Feeny DH. Development of a multi-attribute preschool health status classification system. *Pediatric Res* 1998;43:228A.
11. Achenbach TM, Rescorla LA. *Manual for the ASEBA Preschool Forms and Profiles.* Burlington, VT: University of Vermont Department of Psychiatry, 2000.
12. Asch DA, Jedrziewski MK, Christakis NA. Response rates to mail surveys published in medical journals. *J Clin Epidemiol* 1997;50:1129-36.
13. Al-Shahi R, Warlow C. Using patient-identified data for observational research and audit. *BMJ* 2000;321:1031-32.

Received: December 23, 2003
Revisions requested: April 16, 2004 & August 24, 2004
Revised manuscript received: June 15, 2004 & September 17, 2004
Accepted: September 27, 2004

RÉSUMÉ

Contexte : Les recherches sur les services de santé ou sur la santé de la population dépendent souvent de l'existence de données administratives sur la santé facilement disponibles. Cependant, les maillages de données d'enquête et de données administratives aux fins d'études sur la santé soulèvent certaines craintes quant à la protection des renseignements personnels. Nous avons voulu comparer les taux de consentement au maillage de données dans deux échantillons de personnes soignantes et décrire les caractéristiques des personnes ayant consenti au maillage.

Méthode : Nos sujets étaient des personnes qui prenaient soin d'enfants hospitalisés à la naissance dans des unités de soins néonatals intensifs (USNI) en Colombie-Britannique ou d'enfants sains. Ces personnes ont été priées de signer un formulaire de consentement autorisant l'équipe de recherche à recouper ses données d'enquête avec les dossiers médicaux provinciaux des soignants et des enfants. Au moyen d'une analyse bivariée, nous avons déterminé les caractéristiques de l'échantillon associées au consentement, que nous avons ensuite entrées dans des modèles de régression logistique.

Résultats : Notre échantillon comprenait 1 140 enfants hospitalisés dans des USNI (sur un total de 2 221) et 393 enfants sains (sur un total de 718). Le taux de réponse global était de 55 %, et le taux de réponse dans les familles retracées était de 67,1 %. Le maillage des données avec les dossiers des enfants a été autorisé par 71,6 % des répondants, et le maillage avec les données des adultes soignants, par 67 % des répondants. Les familles d'enfants sains étaient tout aussi susceptibles de donner leur consentement que les familles d'enfants hospitalisés dans des USNI. Des taux de consentement supérieurs étaient associés aux faits d'être un parent biologique, de ne pas avoir besoin d'enquêtes de rappel, de participer à un groupe d'entraide pour parents, de ne pas travailler à plein temps, d'avoir des enfants en moins bonne santé, d'avoir eu des accouchements multiples et de gagner un revenu supérieur.

Conclusion : Le niveau de consentement obtenu donne à penser que lorsqu'ils ont le choix, la plupart des gens sont disposés à autoriser des chercheurs à consulter leur dossier médical personnel à des fins de recherche. Nous avons donc la marge de manouvre nécessaire pour sensibiliser la population à la nature et à l'importance des études combinant des données d'enquête et des données administratives, qui permettent de répondre à d'importantes questions sur la santé.