

Using linked administrative data bases to study a population of NICU and non-NICU Survivors

(Alphabetical order)

Hillel Goelman, Ph.D., The University of British Columbia

Anne Klassen, D.Phil., McMaster University

Zhen Li, Ph.D, Ministry of Education, Newfoundland and Labrador

Sarka Lisonkova, MD., Ph.D, BC Children and Women's Health Centre

Mari Pighini, Ph.D., The University of British Columbia

Anne Synnes, MDCM, MHSC, FRCPC, BC Women's Hospital and the University of British Columbia

Bozena Zdaniuk, Ph.D, The University of British Columbia

Abstract

Background: The paper describes the use of linked administrative data bases to study the medical trajectories of 71,698 children from birth to age 9 years. Two groups were studied: those who were treated in a Neonatal Intensive Care Units (NICU) and those who were not. Data were drawn from anonymized, universal administrative data bases **Methods:** Retrospective population-based study of health and socio-demographic data. **Results:** The primary diagnoses for the NICU children included low birth weight, prematurity and/or small for gestational age. A higher percentage of the NICU children were from lower socioeconomic backgrounds and were more likely to live in urban settings than did the non-NICU children.

Conclusions: The use of linked, universal data bases provides the opportunity to generate detailed, longitudinal health profiles of at-risk and comparison children which can then be linked to other universal data bases. (139 words).

Key words: longitudinal, at-risk, NICU

Corresponding author: Dr. Hillel Goelman, Ph.D, Chair, Interdisciplinary Studies Graduate Program Room 153 Green College Commons, 6201 Cecil Green Park Road, The University of British Columbia, Vancouver, BC, V6T 1Z1, Tel: 604-822-9903, Fax: 604-822-0470. hillel.goelman@ubc.ca.

This paper describes a full two-year cohort of a population of children born in British Columbia, Canada sub-categorized into the Neonatal Intensive Care Unit (NICU) group and a comparison non-NICU group. The paper presents the children's characteristics at birth including their primary diagnoses and the socio-demographic characteristics of the children's mothers, accessed through existing linked databases.

Among the challenges to longitudinal studies of NICU infants are small sample sizes, the difficulty in recruiting and maintaining samples over time and the lack of socio-demographic and medical heterogeneity of the samples. One solution to these challenges is to conduct population-based studies using administrative data bases. This paper is drawn from an ongoing project comparing the health, developmental, and educational trajectories of children who were admitted to a Level II or III Neonatal Intensive Care Unit (NICU) for at least 24 hours, and a comparison group of non-NICU children within the same birth cohort. The paper describes the process, advantages, challenges and strategies in the utilization of linked administrative data bases to study this birth cohort. The paper also describes the birth-related characteristics of the children, their initial diagnoses at birth and the socio-demographic characteristics of the children's mothers.

Using administrative data bases to study NICU and non-NICU populations

Although the chances for survival for at-risk babies in developed countries have significantly increased over the past 30 years ^{1, 2, 3, 4}, they continue to be at risk for developmental delays due to biological, and environmental risk factors. In addition to immediate neonatal medical attention, follow-up care is recommended since long-term disabilities are significant adverse outcomes for many NICU survivors^{1, 5, 6}.

There are many challenges to longitudinal research with NICU infants ^{7, 8, 9}. These include small sample sizes, the difficulty in recruiting NICU infants representative of the NICU population, limited socio-demographic and medical heterogeneity of the samples, and sample attrition over time. In response to these challenges researchers have identified the benefit of population-based studies¹⁰ that can include pre,

peri- and post-natal medical data, up to and including the children's school years^{11, 12}. Such an approach could be enhanced by linking universal, administrative bases on children's medical histories and services to other universal educational and early intervention administrative bases.

Redden, Mulvihill, Wallander, and Hovings¹¹ summarized the benefits of using linked administrative data bases especially in research on low frequency/high intensity developmental challenges:

The unique ability of this approach to instantly provide population-based, long-term data is valuable, and these data can be applied to developmental issues without the noted limitations of other study designs. Another salient advantage of using record linkage methodology is its applicability for low-prevalence or low-occurrence conditions, as in the case of most disabilities...Record linkage methodology overcomes this problem, as researchers can select an adequate sample size available from an extant database while incurring minimal time and financial burdens (P. 329).

A number of studies have drawn on the Kaiser Permanente Neonatal Minimum Data Set (NMDS) in California¹³. This private insurance company covers 30% of the *insured* population in northern California. By their own estimation, the "very poor and very wealthy were under-represented" (p.21) in their data base. The new NMDS was found to be a major improvement in accuracy and data retrieval over the previous database, but new problems emerged: "Our protocol and software do not handle some admission types very well, such as transported babies, babies who go in and out of the NICU, babies readmitted to the NICU after having gone home, and babies transferred from the NICU to the pediatric ward" (p. 24).

Despite these limitations, the database was used successfully when implemented by Smith, Zupancic, and McCormick¹⁴. This study focused on children who had and who were diagnosed with bronchopulmonary dysplasia (BPD) in the first year of life and included all children born under 33 weeks regardless of birth weight. The researchers captured over 95% of Level III NICU admissions found in the NMDS. These data were linked to State of California birth certificate records in order to find socio-demographic and family data on children. The study succeeded in linking individualized family data in two administrative data bases and confirmed the relationships between the number of NICU days at birth and the number of BPD-related re-admissions to hospitals in the first year of life. However, They also pointed

out that they were limited to the data that were originally entered in the NMDS by other persons, at another time, for other purposes: “A limitation of using a pre-existing database is that we were unable to further explore some of these factors (p. 801)” in BPD rehospitalizations. Thus, the use of pre-existing medical data basis can be of great value for answering some types of research questions as long as the limitations are recognized^{15, 16}.

Method

In the universal health care system in British Columbia, Canada, health data on all citizens are entered into the BC Linked Health Database (BCLHD) and identified through a random number individually matched to the personal health number (PHN) in order to maintain individual anonymity and confidentiality¹⁸. The four components of the BCLHD are the Hospital Separations file (HOSP); the Medical Services Plan Registration and Billing (or MSP R & B) file; MSP Payments file; and the province's Vital Statistics files that include the Birth, Deaths (Mortality) and Clinical Data Births and Deaths files^{15, 16}.

A Data Access Request was submitted to the BC Ministry of Health for data on the 1996/97 two-year birth cohort, matched across the BCLHD and the Vital Statistics data bases by year/month birth date, gender and three-digit postal code. Data access was granted through the Centre for Health Services and Policy Research (CHSPR), a university-based agency with expertise in health care data development and where the BCLHD is housed. Data stewards at CHSPR (currently Population Data BC) proceeded with the BCLHD data organization and linkage of the Vital Statistics, Hospitalization and MSP files. Files were then made accessible to the research team to conduct data analyses and interpretation of the results. The files were checked for missing, redundant, or inconsistent data in these different files and informed decisions were made in order to resolve these issues. For example, files in the Vital Statistics database are organized on a calendar year basis (January 1-December 31), while files in the BCLHD are organized on a fiscal year basis (April 1-March 31).

The crucial step of preparing data from multiple files for analysis was accurate case matching. In order to analyze data extracted from the different Health and Vital Statistics files, the BCLHD utilized the identifier data linkage process¹⁷, with randomly generated substitutes for individual Personal Health Numbers (PHNs). Identifier data linkage refers to "information that identifies an individual or for which it is reasonably foreseeable in the circumstances that it could be utilized, either alone or with other information,

to identify an individual²⁰. PHNs were replaced with randomly assigned individual identification (Study ID) numbers^{iv}. The preparation of the final data set for analyses was guided by the multiple verification of the Study ID and date of birth (year and month), gender, and three-digit postal codes of cases across different files.

A total of 71,713 children were identified in the Vital Statistics Clinical Data-Births files whose records matched the data in both the Birth and Registration files. Subsequent steps of matching and deleting erroneous or redundant records were then conducted. Cases were deleted when they were not included in all five files, did not have identical dates of birth or did not have consistent study identification numbers. In addition to birth place and birth dates, the final inclusion criterion was for children to be continuously registered with the Medical Services Plan (MSP) between 1996/7 and 2004. All subjects who died prior to 2004 were therefore excluded. This yielded a final longitudinal cohort of 71,698.

The cohort was separated into a NICU group (5,750 children admitted to a Level II/III NICU after birth) and a non-NICU group (the 64,551 other children). NICU status could not be established for 1,397 cases for two main reasons. First, 1,368 cases had no birth record in the hospitalization file, possibly because they were born at home. Second, 29 cases originally assigned to the non-NICU group had their status changed to “missing” because their low birth weight or low gestational age made survival outside an NICU setting unlikely.

Results

Table 1 presents the demographic characteristics of mothers and information on children at the time of birth, categorized by children’s NICU status. Analyses of mother’s postal codes at the time of the child’s birth showed that 5,251 (92%) of mothers of NICU children lived in urban areas whereas only 477 (8%) lived in rural locations. The proportion of urban children was lower for the non-NICU group (86%). The data also revealed that more than 1,000 of the NICU children (19%) were ‘outborn’- that is they were transferred to a hospital with a NICU level II or Level III usually because they needed more intensive care.

The NICU families were overrepresented in the two lowest socio-economic (SES) quintiles and underrepresented in the two highest SES quintiles compared to the non-NICU families. Mothers of children in the NICU group were significantly more likely to be unmarried (23% vs. 21%) and to have given birth for the first time (primipara) than mothers in the non-NICU group of children (51% vs. 44%). Aboriginal ethnic origin was not associated with need for NICU care. Children in the NICU group were significantly more likely to be a twin or higher order multiple birth than children in the comparison group (12% vs. 1.5%) and the NICU group had a significantly higher proportion of boys than did the non-NICU group (56% vs. 51%).

As expected, NICU children are highly overrepresented in the low birth weight (BW) categories. Over one percent (N=67) of NICU children weighed 800 grams or less at birth (Extremely Low Birth weight, ELBW) compared to none in the non-NICU group. Similarly, 6% of NICU subjects had a birth weight of 801-1499 grams, the very low birth weight (VLBW) category. Non-NICU subjects with a birth weight less than 1500 grams were excluded. Twenty four per cent of NICU children compared to 2.3% of their non-NICU counterparts fell into the low birth weight (LBW) category (1500-2499 grams). As expected, gestational age (GA) also differentiates the two groups with more premature births in the NICU group. Twelve percent of the NICU children were reported to be born at less than 32 weeks gestational age -- compared to .02% of non-NICU children. Similarly, almost 35% of the NICU children are reported in the 33-37 week brackets (pre-term birth) as compared to only 8% of their non-NICU counterparts.

The primary diagnoses based on ICD-9 codes for the children in the NICU and non-NICU group are presented in Table 2. The most frequently used diagnoses for NICU children were those related to short gestation and low birth weight, followed by those related to respiratory conditions. Smaller percentages were referred to the NICU due to infections, fetal distress, hypoglycemia and other conditions.

Discussion

These data demonstrate the advantages and challenges of using linked administrative bases by generating detailed medical and contextual information on an entire province population of NICU and a non-NICU infants within a given birth cohort.

The results comparing NICU with non-NICU children corroborate and extend previous reports on neonatal risk associated with low SES, first-time mothers, low birth weight, early gestational age and multiple births (e.g., Wilson-Costello¹). Birth weight (BW) and gestational age (GA) are two of the admission criteria to Level II/III NICUs- that are most frequently studied. However, nearly 70% of NICU babies in our study were within the normal birth weight category (2500g or above). In terms of gestational age, 47% of the NICU children were born at less than 37 weeks of gestational age. However, almost 53% were born at 38 weeks or above indicating that lower gestational age alone does not explain the majority of NICU admissions. These data point to the importance of expanding the follow-up efforts to include a wider range of NICU Level II and Level III graduates^{9, 12}.

Using the administrative data bases at our disposal we were able to identify virtually 100% of the population of NICU graduates in a two-year period and to verify that records for these children were intact up to and including age nine. This is extremely promising in terms of further longitudinal work on this population. Using multiple data bases allowed us to verify the identity of the children, and to collect information on contextual maternal, family and demographic variables. Limitations to this approach included the fact that we could examine only the variables that were available in the files, thus limiting hypothesis generation and testing. Another common issue is missing data. Rates of missing data were linked to different reasons and varied across variables. Examples include children with missing or inconsistent birth dates; and children whose birth weights and gestational ages would have resulted in referrals to NICU, but for whom no such referral information was entered., The proportion of missing data out of the final cohort of 71, 698 children ranged between .02% for number of live births variable and 4.6% for mother's marital status at the time of child's birth. Other limitations regard the accuracy and consistency of the primary diagnoses across sites and data

duplication for children in the NICU category who required a transfer from a Level II to a Level III NICU. Other limitations regard the accuracy of medical diagnoses which may be subject to a transcription error.

This paper integrates descriptive information and findings that may be utilized as reference points for future population-based research using linked health data bases. Access to databases containing information on the same subjects over time provides the opportunity to conduct retrospective trajectory research following yearly time intervals. Further research utilizing this full birth cohort will examine NICU and non-NICU children's access and utilization of health services in identifying, for example, predictors for hospital re-admission in both groups of children; and, examining the extent to which geography^{20,21} plays a role in hospitalization patterns and the use other medical and health services. As a part of a larger project we will examine whether NICU referral will impact on their educational outcomes in elementary school.

References

1. Wilson-Costello, D., Friedman, H., Minich, N., et al. (2005). Improved Survival Rates With Increased Neurodevelopmental Disability For Extremely Low Birth Weight Infants in the 1990s. *Pediatrics*, 115, 997-1003.
2. Fischer, N., Steurer, M. A., Adams, M., et al. (2009). Survival rates of extremely preterm infants (gestational age <26 weeks) in Switzerland: impact of the Swiss guidelines for the care of infants born at the limit of viability. *Archives of disease in childhood. Fetal and neonatal edition* . Vol. 94, n6, F407-F413.
3. Riley, K., Roth, S., Sellwood, M., et al. (2008). Survival and neurodevelopmental morbidity at 1 year of age following extremely preterm delivery over a 20-year period: a single centre cohort study. *Acta Pædiatrica*, 97(2), 159-165.
4. Synnes, A. R., Buchanan, L., Ruth, C., et al. (2008). Management of the newborn delivered at the threshold of viability. *BC Medical Journal* 50(9), 498-508
5. Hack, M., Taylor, H. G., Drotar, D., L., et al. (2005). Chronic Conditions, Functional Limitations, and Special Health Care Needs of School-aged Children Born with Extremely Low-Birth-Weight in the 1990s. *The Journal of the American Medical Association*, 294(3), 318-325.
6. Ment, L. R., Vohr, B., Allan, W., et al. (2003). Change in Cognitive Function Over Time in Very Low-Birth-Weight Infants. *The Journal of the American Medical Association*, 289(6), 705-711.
7. Escobar, G., Greene, J., Hulac, P., et al. (2005). Rehospitalisation after birth hospitalisation: patterns among infants of all gestations. *Archives of Disease in Childhood*, 90(2).
8. Engle, W. A., Tomashek, K. M., Wallman, C., et al., and Committee on Fetus and Newborn (2007). "Late-Preterm" Infants: A Population at Risk. *Pediatrics*, 120(6), 1390-1401.
9. Khashu, M., Narayanan, M., Bhargava, et al. (2009). Perinatal Outcomes Associated With Preterm

- Birth at 33 to 36 Weeks Gestation: A Population-Based Cohort Study. *Pediatrics*, 123, 109-113.
10. Stoltenberg, C & Magnus, P. (1995). Children with low birth weight and low gestational age in Oslo, Norway: Immigration is not the cause of increasing proportions. *Journal of Epidemiology and Community Health*, 49:588-593.
 11. Redden, S.C., Mulvihill, B.A., Wallander, .J., et al. (2000). Applications of epidemiological data linkage methodology to examine early risk for childhood disability. *Developmental Review*. 28, 319-349.
 12. Schiariti, V., Farrell, K., Houbé, J.S., et al.(2009). Period prevalence of epilepsy in Children in BC: A population-based study. *Can. J. Neurol. Sci*; 36: 36-41.
 13. Escobar, G., Joffe, S., Gardner, M.N., et al. in M.A ., Folck, B.F. and Carpenter, D.M. (1999). Rehospitalization in the first two weeks after discharge from the Neonatal Intensive Care Unit. *Pediatrics*, 104:1, e2
 14. Smith, V.C., Zupancic, J., McCormick, et al.. (2004). Rehospitalization in the first year of life among infants with bronchopulmonary dysplasia. *The Journal of Pediatrics*, 144(6), pp. 799-803.
 15. Chamberlayne, R., Green, B., Barer, M.L., et al.. (1998). Creating a population-based linked health database: A new resource for health services research, *Can. J. Public Health* 89, (pp. 270–273).
 16. McGrail, K. M. (2008). Income-related inequities: Cross-sectional analyses of the use of medicare services in British Columbia in 1992 and 2002. *Open Medicine*, 2(4), E3-10.
 17. Population Data BC (2009). *History of data access*. Retrieved Nov 25, 2009 from <http://www.popdata.bc.ca/dataaccess/history>.
 18. Liu, S. & Wen, S.W. (2000). Development of record linkage of hospital discharge data for the study of neonatal readmission. *Chronic Diseases in Canada*, 20:3.
 19. Cavoukian, A. (2005). *Frequently Asked Questions Personal Health Information Protection Act*. Toronto, ON: Information and Privacy Commissioner/Ontario.

20. Niven, C, Puderer, H. & Janes, D. (2000). Census Metropolitan Area and Census Agglomeration Influenced Zones (MIZ): A Description of the Methodology. Ottawa: Geography Division, Statistics Canada.
21. Canadian Institute for Health Information (2006). *How Healthy Are Rural Canadians? An Assessment of Their Health Status and Health Determinants*. Ottawa: Canadian Institute for Health Information.

Using Linked Administrative Data Bases September 7 2011

Table 1

Demographic Information and Characteristics of Mothers and Children at the Time of Birth (N=71,698)

		NICU Children N=5,750 (8%)	Non-NICU Children N=64,551 (90%) ¹	Total within category	Total N=71,698	Chi Square Value (p value)
Mother and Children's Residence Location ² (354/1694 missing values ³)	Urban	5,251 (91.7%)	54963 (85.5 %)	60,214 (86%)	70,004	165.99 (p=.000)
	Rural	477 (8.3%)	9,313 (14.5%)	9719 (14 %)		
SES ⁴ (Quintiles) (1,863/3170 missing values)	1 st	1,384 (24.7%)	13,999 (21.6 %)	14,983 (21.9%)	68,528	62.8 (p=000)
	2 nd	1,266 (22.6%)	13,032 (20.7%)	14,298 (20.9%)		
	3 rd	1,111 (19.8%)	12,885 (20.5 %)	13,996 (20.4%)		
	4 th	974 (17.3%)	13,081 (20.8%)	14,055 (20.5%)		
	5 th	879 (15.7%)	10,317 (16.4 %)	11,196 (16.3%)		
Mother's marital status (3267/4543 missing values)	Unmarried	1247 (23.1%)	13,046 (21.1 %)	14,293 (21.3%)	67,155	11.78 (p=.001)
	Married	4147 (76.9 %)	48,715 (78.9%)	52,862 (78.7%)		
Mother's self-reported Aboriginal Status at the time of child's birth (1629/1661 missing values)	Aboriginal	248 (4.3%)	3,068 (4.8 %)	3,316 (4.7%)	70,037	2.47 (p=.116)
	Non-Aboriginal	5,502 (95.7%)	61,219 (95.2 %)	66,721 (95.3%)		
Parity (95/1489 missing values)	First-born	2,924 (51.0%)	28,354 (44 %)	31,278 (44.5%)	70,209	108.73 (p=.000)
	Second-born	1790 (31.2%)	23,684 (36.7%)	25,474 (36.3%)		
	Third- or later-born	1019 (17.8%)	12,438 (19.3%)	13,457 (19.2%)		
Number of Live Births (15/1411 missing values)	Singleton	5059 (88.1%)	63,561 (98.5%)	68,620 (97.6%)	70,287	2654.07 (p=.000)
	Twin	634 (11%)	974 (1.5 %)	1,608 (2.3%)		
	Triplets	52 (.9%)	7 (.01%)	59 (.1%)		
Child Gender (343/1684 missing values)	Female	2,542 (44.4%)	31,485 (49 %)	34,027 (48.6%)	70,014	44.68 (p=000)
	Male	3,187 (55.6%)	32,800 (51%)	35,987 (51.4%)		

Using Linked Administrative Data Bases September 7 2011

Table 1 (continued)

		NICU Children N=5,750 (8%)	Non-NICU Children N=64,551 (90%) ¹	Total within category	Total N=71,698	Chi Square Value (p value)
Outborn ⁵		1,074 (18.7%)		1,074	1,074	
Birth Weight (in grams) (66/1417 missing values ³)	≤800	67 (1.2%)	0 (0%)	67 (.1%)	70,281	11,507.37 (p=.000)
	801-1499	356 (6.2%)	0 (0%)	356 (.5%)		
	1500 -2499	1,406 (24.5%)	1493 (2.3%)	2,899 (4.1%)		
	≥2500	3,919 (68.2%)	63,040 (97.7)	66,959 (95.3%)		
Gestational Age (in weeks) (186/1544 missing values)	<28	138 (2.4%)	0 (0%)	138 (.2%)	70,154	12,221.64 (p=000)
	28-32	581 (10.1%)	16 (.02%)	597 (.9%)		
	33-37	1,985 (34.6%)	5374 (8.3 %)	7359 (10.5%)		
	>37	3034 (52.9%)	59,026 (91.6%)	62,060 (88.5%)		

1. NICU placement data are missing for 1397 (2%) cases – no hospital birth records
2. Residence established on three digit postal codes. Postal codes with intermediate “0” were identified as rural (Canada Post).
3. The first number reflects missing values for the whole cohort, the second number reflects missing values in combination with missing NICU assignment
4. SES is an aggregated variable - Socioeconomic EA decile and quintile
5. Includes cases with recorded transfer from another hospital and/or placement in NICU level II/III at age greater than zero days.

Using Linked Administrative Data Bases September 7 2011

Table 2
Primary diagnoses¹ for NICU and Non-NICU children at time of birth

	Diagnosis number	NICU		Non-NICU	
		N	%	N	%
Single Birth Born at Hospital	V300	911	14.5	51709	80
Congenital Anomalies	740–759.9	264	4.2	1637	2.5
Fetus Or Newborn Affected By Maternal Conditions Which May Be Unrelated To Present Pregnancy	760				
Other	760.8	139	2.2		
Slow Fetal Growth And Fetal Malnutrition	764				
Light-For-Dates" Without Mention Of Fetal Malnutrition	764.0	94	1.5		
Fetal Growth Retardation – Unspecified	764.9	108	1.7		
Disorders Relating To Short Gestation And Unspecified Low Birth weight	765				
Extreme Immaturity	765.0	129	2.0		
Other pre-term births (e.g. LBW)	765.1	1763	28.0	1367	2.1
Intrauterine Hypoxia And Birth Asphyxia	768				
Fetal Distress In Live Born Infant	768.3	98	1.6	700	1.1
Severe Birth Asphyxia	768.5	74	1.2		
Mild to Moderate Asphyxia	768.6			625	1.0
Respiratory distress syndrome	769	375	6.0		
Other Respiratory Conditions Of Fetus And Newborn	770				
Massive Aspiration Syndrome	770.1	110	1.7		
	770.2	60	1.0		
Transitory Tachypnea Of Newborn	770.6	604	9.6	782	1.2
Other Respiratory Problems After Birth	770.8	467	7.4	731	1.1
Infections Specific To The Perinatal Period	771				
Other Infection Specific To The Perinatal Period	771.8	91	1.4		

Using Linked Administrative Data Bases September 7 2011

Table 2 (continued)

Unspecified Fetal And Neonatal Jaundice	774.6	116	1.8	1520	2.4
Endocrine And Metabolic Disturbances Specific To The Fetus And Newborn	775				
Neonatal Hypoglycaemia	775.6	85	1.3		

¹ Primary diagnoses according to ICD-9 Codes (International Classification of Diseases)

Notes: a) only diagnoses affecting more than 1% of children in a given subgroup were listed
b) 665 children were given two primary diagnoses at time of birth