

The Importance of Unknowns in Epidemiologic Studies¹

UUCS-85-116

Daniel S. Janik, MD, MPH
Fellow in Environmental Epidemiology
Rocky Mountain Center
for Occupational and Environmental Health

Thomas C. Henderson, PhD
Associate Professor of Computer Science

John L. Lyon, MD
Associate Professor of Family and Community Medicine
University of Utah Medical Center

Ed Sharp, BS
Director, Computer Center

August L. Jung, MD
Associate Professor of Pediatrics
Intermountain Newborn Intensive Care Unit
University of Utah Medical Center

University of Utah
Salt lake City, Utah

¹Completed, in part, with a grant from CETUS Research, A Division of CETUS Systems Corporation, Salt Lake City, Utah

1. Introduction

Epidemiologic study data often include omitted/unobtainable responses (unknowns). In most cases, unknowns are eliminated during data-reduction to facilitate analysis. We examined the effect that elimination of unknowns would have on mortality calculations using data on newborns admitted to a newborn intensive care unit (NICU).

2. Method

Eight domestic and foreign NICU's with advanced NICU computer systems (CETUS Systems Corporation) capable of distinguishing between omitted/unobtainable (unknown) and known responses were identified (Figure 1). In each case, 250 to 450 medical and historical factors were captured on admitted infants [1, 2, 3, 4]. We chose Intermountain Newborn Intensive Care Center data for retrospective analysis because their system had been in operation the longest (5 years), and all consecutive 1984 NICU admissions (N=214) were available for examination.

Location	NICU Name
Albuquerque, NM	University of New Mexico Hospital NICU
Denver, CO	Denver Children's Hospital NICU
Florence, SC	McLeod Regional Medical Center NICU
Louisville, KY	Kosair Children's Hospital NICU
Vancouver, BC	British Columbia Children's Hospital
Salt Lake City, UT	Intermountain NICU
	- University of Utah Medical Center
	- Primary Children's Hospital NICU
Washington DC	Children's Hospital National Medical Center NICU

Figure 1. NICU's with Specialized NICU Computer Systems which Distinguish between Omitted/Unobtainable (Unknown) and Known Responses

Data were divided into unknown (omitted/unobtainable), abnormal and normal response groups. Forty-three factors with at least four unknown responses recorded during 1984 were selected for analysis. NICU mortality (number of NICU deaths per 100 NICU admissions) was selected as the outcome indicator. In order to quantify the effects of eliminating unknowns, NICU mortality rate ratios (NMRR) were calculated, by factor, for abnormal, normal and known (abnormal and normal) both without and with knowns.

$$\text{NMRR} = \frac{\text{NICU mortality rates without unknowns}}{\text{NICU mortality rates with unknowns}}$$

NMRR's were calculated using the formula: Thus, an NMRR = 1 indicates that eliminating unknowns has virtually no effect on reported outcomes, while an NMRR ≥ 1.5 or ≤ 0.5 indicates a substantial over or under reporting of outcomes [1].

To evaluate significance, 90% test-based confidence limits (TBCL) were calculated for each NMRR using the formula:

$$90\% \text{ TBCL} = \text{NMRR} * \exp(1 \pm [1.645/\text{chi}])$$

A 90% TBCL > 1 for NMRR ≥ 1.5 , or a 90% TBCL < 1 for NMRR ≤ 0.5 implies a high degree of confidence that the difference in NICU mortalities measured is statistically valid [5]. To assess the importance of unknowns, as a group, NMRR's and corresponding 90% TBCL were calculated as above for unknowns versus abnormals, unknowns versus normals, and unknowns versus knowns.

It is possible that differences in NMRR's could have been due to the influence of other factors, rather than the particular unknown examined. The factor with the highest NMRR's (RBL = Risk Est Before Labor) was selected for further analysis. Four traditionally strong determinants of NICU mortality (BWT = Birthweight, PNI = Prenatal Infection, PCR = Prenatal Care Received, and MNS = Maternal Nutritional Status) were identified. Each patient record with an unknown RBL (N=24) was individually matched with 5 controls over the selected factors. Controls were matched to plus or minus 100 grams BWT, exactly matched for PNI, PCR and MNS; any known RBL was accepted. Whenever possible, controls were obtained from 1984 INICC admissions. When all five controls were not obtainable from 1984 admissions, 1985 admissions (starting with January 1, 1985 and working up through May 31, 1985), and if necessary 1980-83 admissions (starting with December 1983 and working backwards through January 1980) were used. Summary NICU mortality rates were calculated for study and control groups and compared using NMRR's. In addition, the Mantel-Haenszel chi square (MHCS), maximum likelihood estimate of the odds ratio, and 90% test-based confidence limits for the unknown vs. matched controls by the method of Rothman-Boice [6]. We determined that an MHCS greater than or equal to 13.3 ($p < 0.01$, one tail) would support the conclusion that

differences between unknowns and controls were due to the presence of unknown rather than to other confounding factors.

3. Results

Of the 43 factors examined, 7 (Table 1) had abnormal + unknown versus unknown NMRR's greater than or equal to 1.5, and 1 less than or equal to 0.5. Of these, 6 had NMRR's ≥ 1.5 and 90% TBCL > 1 .

NMRR's for unknowns versus abnormal, unknowns versus normal, and unknowns versus known are summarized in Tables 2-4, respectively. In each case, several NMRR's ≥ 1.5 with 90% TBCL's > 1 were identified. Summary NICU mortality rates for selected unknowns and matched controls were 41.7% (10/24) and 13.3% (16/20), respectively.

The rate ratio for unknowns-to-matched-controls was 3.1 (well above 1.5). The MHCS for selected unknowns versus matched controls was 14.1 ($p < 0.01$). The maximum likelihood estimate of the odds ratio was 9.6, well above 1.5 and well within the 90% test-based confidence limits of 2.9-31.3.

<u>Factor</u>	<u>NICU Mortality Rate (%)</u>		<u>Rate Ratio</u>	
	<u>Abnormals+Unknowns (AU)</u>	<u>Abnormals (A)</u>	<u>AU:A</u>	<u>90% TBCL</u>
Risk Est Bef Labor	15.9 (22/138)	10.5 (12/114)	1.51	1.29-1.77
# Prenat Visits	10.4 (15/144)	3.8 (2/53)	2.74	0.86-8.73
# Prev Postmatures	14.3 (4/28)	33.3 (1/3)	0.43	0.08-2.19
# Prev Hi BWT	9.1 (3/33)	0 (0/9)	inf*	inf
Mat Coombs	11.8 (17/144)	0 (0/6)	inf	inf
Preg Hb/Hcts	9.0 (11/22)	0 (0/10)	inf	inf
Rub Titer	8.0 (9/122)	0 (0/3)	inf	inf
VDRL	10.3 (10/97)	0 (0/1)	inf	inf

*inf = infinity

Table 1. NICU Mortality Rates, Rate Ratios and 90% Test-Based Confidence Limits (TBCL) for Abnormals and Unknowns versus Abnormals for Factors with Rate Ratios ≥ 1.5 or ≤ 0.5 for 214 Consecutive NICU Admissions (INICC, 1984).

Factor	NICU Mortality Rate (%)		Rate Ratio	
	Unknowns (U)	Abnormals (A)	U:A	90% TBCL
Risk Est Bef Labor	41.7 (10/24)	10.5 (12/114)	3.97	2.91-7.18
# Prenat Visits	14.3 (13/91)	3.8 (2/53)	3.76	1.25-11.29
Preg Infections	45.5 (5/11)	16.3 (8/49)	2.79	1.25-6.21
Meconium Stain	25.0 (3/12)	9.1 (1/11)	2.75	0.52-14.4
Mat Hlth Prob	11.8 (2/17)	5.9 (5/85)	2.00	0.55-7.31
Pat Hlth Prob	15.2 (7/46)	8.3 (3/36)	1.83	0.64-5.24
Mec in Trach	25.0 (3/12)	14.3 (3/21)	1.75	0.53-5.80
# Prev Postmatures	10.7 (3/28)	33.3 (1/3)	0.32	0.06-1.73
# Neonatal deaths	7.7 (1/13)	33.3 (5/15)	0.23	0.05-1.00
# Prev Hi BWT	12.5 (3/24)	0 (0/9)	inf*	inf
Mat Coombs	12.3 (17/138)	0 (0/6)	inf	inf
Preg Hb/Hcts	9.8 (11/112)	0 (0/6)	inf	inf
Rub Titer	8.3 (9/109)	0 (0/3)	inf	inf
VDRL	10.4 (10/96)	0 (0/1)	inf	inf

*inf = infinity

Table 2. NICU Mortality Rates, Rate Ratios and 90% Test-Based Confidence Limits (TBCL) for Unknowns versus Abnormals for Factors with Rate Ratios ≥ 1.5 or ≤ 0.5 for 214 Consecutive NICU Admissions (INICC, 1984).

4. Discussion

NICU mortality rates for abnormals from which unknowns were eliminated were substantially different from abnormals from which unknowns were not eliminated for 19% (8/43) factors. Of these 8 factors, 6 were determined to be statistically significant by the methodology employed. NICU mortality was under reported by up to 63% and in one case, over reported by up to 133%. NICU mortality rates were substantially under reported for 88% (7/8) of the above factors, and for 16% (7/43) of factors in general. Clearly elimination of unknowns can affect reported outcome.

NICU mortality rates for unknowns were substantially different from abnormals, normals and knowns for 33% (14/43), 23% (10/43) and 26% (11/43) factors, respectively. In each case, at least three factors were determined to be statistically significant. For one factor, the unknown NICU mortality rate was up to 4 times the abnormal, 16 times the normal and 6 times the known NICU mortality rate. In at least some instances, unknowns may

NICU Mortality Rate (%)		Rate Ratio		
Factor	Unknowns (U)	Normals (A)	U:N	90% TBCL
Risk Est Bef Labor	41.7 (10/24)	2.6 (2/76)	16.04	7.80-33.0
Preg Infections	45.5 (5/11)	7.1 (11/154)	6.41	3.55-11.62
Mat Preg Wt	16.7 (13/78)	6.5 (7/107)	2.57	1.27-5.22
Mec Staining	25.0 (3/12)	9.9 (18/181)	2.53	0.99-6.46
Risk Est Dur Labor	10.0 (1/10)	4.5 (1/22)	2.22	0.86-5.73
Sib Hlth Prob	18.2 (6/33)	8.4 (12/143)	2.17	1.02-4.64
Mat Wt Gain	15.9 (11/69)	8.8 (7/80)	1.81	0.87-3.76
Mec in Trach	16.7 (1/6)	10.8 (22/203)	1.55	0.67-3.56
Fet Monitor	4.8 (1/21)	13.3 (17/128)	0.36	0.08-1.63
Mon Abnor	4.3 (1/23)	15.0 (20/133)	0.29	0.06-1.27

*inf = infinity

Table 3. NICU Mortality Rates, Rate Ratios and 90% Test-Based Confidence Limits (TBCL) for Unknowns versus Normals for Factors with Rate Ratios ≥ 1.5 or ≤ 0.5 for 214 Consecutive NICU Admissions (INICC, 1984).

exceed abnormal, normal, or known in importance, and deserve to be analyzed and reported separately.

Unknown versus matched-controls' NICU mortality rates and Mantel-Haenszel chi square support the premise that observed differences due to the unknown and not other confounding factors. While the methodologies employed were limited both in scope and application, our data suggest that not only should unknowns be eliminated with care, but in most instances they should either be reported as abnormal, or analyzed separately for their particular contribution to outcome. A practical suggestion is that unknowns be included with abnormal, normal and known where sensitivity is of primary concern, and be analyzed and reported separately where specificity is of primary concern.

The methodologies we used reflect time, knowledge and monetary constraints common to most epidemiologic research efforts. Some constraints, however, deserve special comment:

1. NICU admissions served as our reference population. NICU admissions

Factor	NICU Mortality Rate (%)		Rate Ratio	
	Unknowns (U)	Knowns (K)	U:K	90% TBCL
Risk Est Bef Labor	41.7 (10/24)	7.4 (14/190)	5.64	3.19-9.98
Preg Infections	45.5 (5/11)	9.4 (19/203)	4.84	1.94-12.08
Mec Staining	25.0 (3/12)	10.4 (21/202)	2.40	0.95-6.07
Mat Preg Wt	16.7 (13/78)	8.1 (11/136)	2.06	1.11-3.84
Sib Hlth Prob	18.2 (6/33)	9.9 (18/181)	1.84	0.89-3.80
Mat Wt Gain	15.9 (11/69)	9.0 (13/145)	1.77	0.95-3.30
# Prenat Visits	14.3 (13/91)	8.9 (11/123)	1.61	0.85-3.05
Pat Hlth Prob	15.2 (7/46)	10.1 (17/168)	1.50	0.75-2.99
Mec in Trach	16.7 (1/6)	11.1 (23/208)	1.5	0.31-7.23
Fet Monitor	4.8 (1/21)	11.9 (23/193)	0.40	0.09-1.85
Mon Abnor	4.3 (1/23)	12.0 (23/191)	0.36	0.08-1.65

*inf = infinity

Table 4. NICU Mortality Rates, Rate Ratios and 90% Test-Based Confidence Limits (TBCL) for Unknowns versus Knowns for Factors with Rate Ratios ≥ 1.5 or ≤ 0.5 for 214 Consecutive NICU Admissions (INICC, 1984).

constitute a select and biased subset of pregnancies, newborns or ill newborns.

2. NICU mortality served as our outcome indicator. NICU mortality is an uncommon and select outcome indicator incorporating a variable observation period. It is a biased subset of neonatal or infant mortality. However, had we selected neonatal or infant mortality as our outcome indicator, bias would have been introduced due to patient attrition during the longer observation period required. Bias is thus introduced through the use of any of the above outcome indicators. We found NICU mortality to be a convenient, timely and inexpensive outcome indicator, well-suited to this particular study.
3. Few patient records or medical data systems specifically capture unknown responses. Where unknown responses are not well accommodated, unknowns often appear to occur infrequently and randomly. The NICU's considered for this study all had advanced NICU computing systems capable of distinguishing and accommodating unknowns including distinguishing between unknowns due to unobtainable data (unobtainable data) and unknowns due to failure to even solicit the information indicated (omitted question). We did not extend our analysis to these subgroups. However, our unknowns appeared to consist mostly of non-random unobtainables. The assumption that unknowns as a group can be eliminated without affecting outcomes may originate from not recognizing this distinction.

Our observations are probably applicable to a wide variety of analytic methods and disciplines. For example, image processing, where noise and data ambiguity may contribute unknowns, and expert medical systems, where unknowns may actually be important delimiters.

5. Summary

We examined the effect that eliminating unknowns had on NICU mortality rates using several different methods. For many factors elimination of unknowns substantially biased outcome reports. Differences observed were probably due to the unknown and not other confounding factors. Unknowns should be eliminated with caution from medical, epidemiologic and other studies. In most instances, they should probably be reported as abnormals or analyzed separately for their particular contribution to the outcome.

References

- [1] Janik, D.S. et al.
A Computerized Single Entry System for Recording and Reporting Data on High-Risk Newborn Infants.
J. Pediatr 93(519), September, 1978.
- [2] Janik, D.S. et al.
Computerized Newborn Intensive Care Data Recording and Reporting II: An Online System.
J. Pediatr 94(328), February, 1979.
- [3] Janik, D.S. et al.
Computerized Newborn Intensive Care Data Recording and Reporting III: A Practical Microcomputer System.
J. Pediatr 97(497), September, 1980.
- [4] Janik, D.S. and Ed Sharp.
A Computerized for Recording, Reporting and Researching Data for Newborn Intensive Care Units (NICU) Transferable to Most Sites.
In Lindberg, D.A.B. and P.L. Reichertz (editor), *Lecture Notes in Medical Informatics* - 24. 1984.
- [5] Kleinbaum, D.G, L.L. Kupper and H. Morgenstern.
Epidemiologic Research.
Wadsworth, Inc., Belmont, CA, 1982.
- [6] Rothman, K.J. and J.D. Boice.
Epidemiologic Analysis with a Programmable Calculator.
Epidemiology Resources, Inc., Boston, Massachusetts, date unknown.

