A Validation Study of the Physician Notices of Birth Database in the Regional Municipality of Ottawa-Carleton

by:

Beth Henderson-Tadeson

Supervisor:

Dr. Paula J. Stewart

ISBN 0-612-15723-7

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

TABLE OF CONTENTS

LIST OF EQUATIONS

LIST OF FIGURES

ABSTRACT

The Regional Municipality of Ottawa-Carleton maintains a database of information obtained from routinely assembled Physician Notices of Birth Forms. The validity of this database was called into question when it was discovered that substantial over-reporting of stillbirths existed when compared to the hospital caseroom logbooks. The literature shows that key variables in perinatal research, such as birthweight, gestational age, mother's age and sex of baby are generally reasonably accurate. To date, however, there is no piece of research investigating chance-corrected agreement of whether a baby was born alive or still.

The study problem examined in this paper is the quality of the Ottawa-Carleton Health Department's Physician Notices of Birth (PNOB) database. All stillbirths and babies with unknown outcomes born in an Ottawa hospital to women residing in Ottawa-Carleton were included in the study in addition to a random sample of livebirths. Study variables were gestational age, mother's age, birthweight, number of previous livebirths, sex, postal code and whether the baby was born alive. PNOB forms corresponding to those selected from the existing database were retrieved and the information reported on the forms was re-entered into the computer. This new database served as the record of the actual information on the PNOB form. Data from the existing PNOB database were compared to the newly-entered PNOB database and to the hospital caseroom logbooks. The maternal medical records were examined in all cases of discrepancy

for the above three-way comparison. A random sample of all cases without discrepancy in the above three-way comparison was also drawn and the mother's medical records chart reviewed.

The reliability of data entry from the PNOB form as measured by comparing the existing and newly-entered database was quite high for all variables with most intraclass correlation coefficients and kappas in excess of 0.95. Agreement for the reliability of entering vital status was 0.90 which is high in an absolute sense, but not high relative to the importance of the measure itself.

Excellent agreement between the PNOB database and the logbook was found for all variables, including gestational age, mother's age, birthweight, number of previous livebirths and sex of baby. In most cases, agreement was higher with the newly-entered PNOB database. When limiting postal code to the first three digits, it too was quite accurate.

Among records with discrepancies between the existing and new databases and logbook, agreement with the medical records varied greatly. For both PNOB databases, excellent agreement was obtained for gestational age and poor agreement was achieved for mother's age. Nearly perfect agreement was obtained by the newly-entered database for birthweight while poor agreement was achieved by the existing database. Better agreement for number of previous livebirths was obtained with the newly-entered database.

Among records for which both PNOB databases and the logbook were in perfect agreement, virtually perfect agreement was obtained when compared to the medical records, indicating that when the databases and logbook are consistent, they can also be counted on to be accurate.

For the existing database, sensitivity of detecting stillbirths was moderately high (86%) while positive predictive value for detecting stillbirth was lower at 71%. The newly-entered database was much more sensitive in identifying stillbirths (99.1%), but only minimally better for positive predictive value (76%). These properties of the newly-entered database are likely near the upper limits of the ability of the present system to distinguish between livebirths and stillbirths.

The conclusions of this study indicate that most variables of importance to perinatal research are valid and reliable, with the notable exception of whether a baby was born alive. As a result of the inability of the PNOB database to accurately identify stillbirths, research using the PNOB database involving stillbirths should be avoided at the regional health department level.

ACKNOWLEDGMENTS

This thesis would not have been possible without the guidance and support of several very generous people. First, I would like to thank Mrs. Janice Potter for acquainting me with the Ottawa-Carleton Health Department's PNOB database. Thank you also to Dr. Brian Ivey, Ms. Patricia Niday, Dr. Carl Nimrod and Dr. Robin Walker for their direction in a study leading up to this thesis in its present form. Thanks to Dr. Rama Nair for his statistical advice and patience in reading an early draft of this manuscript. To Dr. Robert Spasoff for his guidance in preparing the final draft of the thesis proposal. To my thesis examiners, Drs. Nicholas Birkett and Ian McDowell for their very meticulous and thought-provoking appraisal of this manuscript.

Many thanks to Dr. Paula Stewart, my mentor, who has always made herself available for consultation, whether at her office or at her home. Without her patience and constant encouragement, this manuscript may never have seen the light of day.

Thank you to my family for their concern, understanding and their generosity with numerous invitations for lunch or dinner or even simply an afternoon tea!

Finally to my husband, Don Tadeson, for his commitment to and sacrifice for my education. Your love and understanding has meant more to me than you will ever know. It's time to get on with life!

1. INTRODUCTION AND STATEMENT OF THE PROBLEM

The utilization of large databases in public health research has been expanding over the past few years. The use of administrative databases as a secondary source of data can be very attractive to those doing epidemiological research. Such data are usually available at a fraction of the cost of collecting one's own data and often cover the entire population of interest. In maternal/child health, one such source of data is the Physician Notice of Live or Still Birth forms (PNOB). In addition to their more traditional use in vital statistics, PNOB forms are becoming more widely used as a source of data for numerous research endeavours. At the Perinatal Health Information Workshop held in Toronto in 1993 by the Association of Public Health Epidemiologists in Ontario (APHEO), it was found that 12 Ontario Health Units were using the PNOB forms to establish their own local database. The extent to which this data is used and the accuracy of it is not known.

Wadhera, Millar and Nimrod, from the Canadian Centre for Health Information and from the Ottawa General Hospital, used over 6.6 million of these birth registration records to track trends in Canadian birthweights of singletons from 1971 to 1989 [1]. They were able to show that the median birth weight has been steadily increasing from 3316g to 3420g (3.1%) between 1971 and 1989. The tremendous number of records available in this database allowed them to analyze the birthweight distribution among strata which are not usually examined because

of problems of small cell frequencies and unstable estimates, including the extremes of mother's age and high birth orders. For instance, median birthweight was lowest among mothers 15-19 years of age (3365g) and rose steadily to a peak of 3458g among mothers 30-34 years of age and slowly dropped off to 3410g in women over 44 years of age. Median birthweight among very young mothers under 15 years of age was slightly higher (3385g) than among the 15-19 year olds. Median birthweight steadily increased with rising birth order. Birthweight did not drop off for babies born fifth or later.

Millar, Strachan and Wadhera extended the analysis to examine trends in low birthweight (LBW) over the same time period [2]. Using the same 6.6 million birth registration records, this group found that the prevalence of LBW (<2500g) decreased by 30% from 6.6% to 4.6% over the 18-year study period. A decrease in the prevalence of babies born with a medium low birthweight (MLBW) between 1500g and 2499g was responsible for most of the decline, while the prevalence of those born with a very low birthweight (VLBW) of <1500g barely changed at all. LBW was also examined in relation to the mother's age. As expected, LBW was the least prevalent in women between the ages of 20 and 39, while those under 20 years and over 40 years of age had similar, yet higher rates. Despite the extremely large number of births examined, estimates of prevalence of LBW among these women at the extremes of their childbearing years still exhibit remarkable statistical variation.

In the Regional Municipality of Ottawa-Carleton (RMOC) the validity of the PNOB database became questionable when it was discovered that substantial errors existed in the reporting of stillbirths. When the statistics were completed for the year 1991, it was shown that the rate of stillbirth was extremely high at 9.1 per 1000 deliveries. Embarking on a research project to determine the cause of this epidemic of stillbirth, it became clear that the singular cause of excess stillbirth was that the PNOB database contained many cases of healthy, living babies who had been incorrectly coded as being stillborn.

As a result of this cautionary lesson, the present study was born with the purpose of examining the quality of the PNOB database in the Regional Municipality of Ottawa-Carleton.

2. LITERATURE REVIEW

2.1 Flow of Official Vital Statistics Information

The term "stillbirth" is defined in the *Vital Statistics Act, Chapter V.4, Section 1* as "the complete expulsion or extraction from its mother of a product of conception either after the twentieth week of pregnancy or after the product of conception has attained the weight of 500 grams or more, and where after such expulsion or extraction there is no breathing, beating of the heart, pulsation of the umbilical cord or movement of voluntary muscle."

An overview of the flow of official information is presented in Figures 1 and 2 (courtesy of Dr. George Pasut). In all cases of live or still birth, a physician's *Notice of Live Birth or Stillbirth* (PNOB) is completed in duplicate and signed by either the attending medical practitioner or nurse (Figure 3). The original copy must be sent within 48 hours to the Division Registrar of Births and Deaths while the duplicate copy is sent (within 48 hours) to the Ministry of Health. The *Notice of Live Birth or Stillbirth* is passed from the Division Registrar to the Office of the Registrar General and then on to the Ministry of Health. The Ministry of Health sends a copy of the PNOB to the local Health Departments. In some locales, health units have set up a process with the hospitals to inform the local boards of the birth for program purposes. This *Notice of Live Birth or Stillbirth* is the source of data for the Ottawa-Carleton Health Department's database.

In the case of livebirths a *Statement of Live Birth* (Figure 4) is completed and signed by either an attending medical practitioner or nurse and forwarded to the Division Registrar. Unlike the *Notice of Live Birth or Stillbirth*, which contains information about the outcome of a pregnancy, the *Statement of Live Birth* contains demographic information about the parents and the baby, including its name. Information contained on this *Statement of Live Birth* is forwarded to the local boards.

Figure #1

Live Births

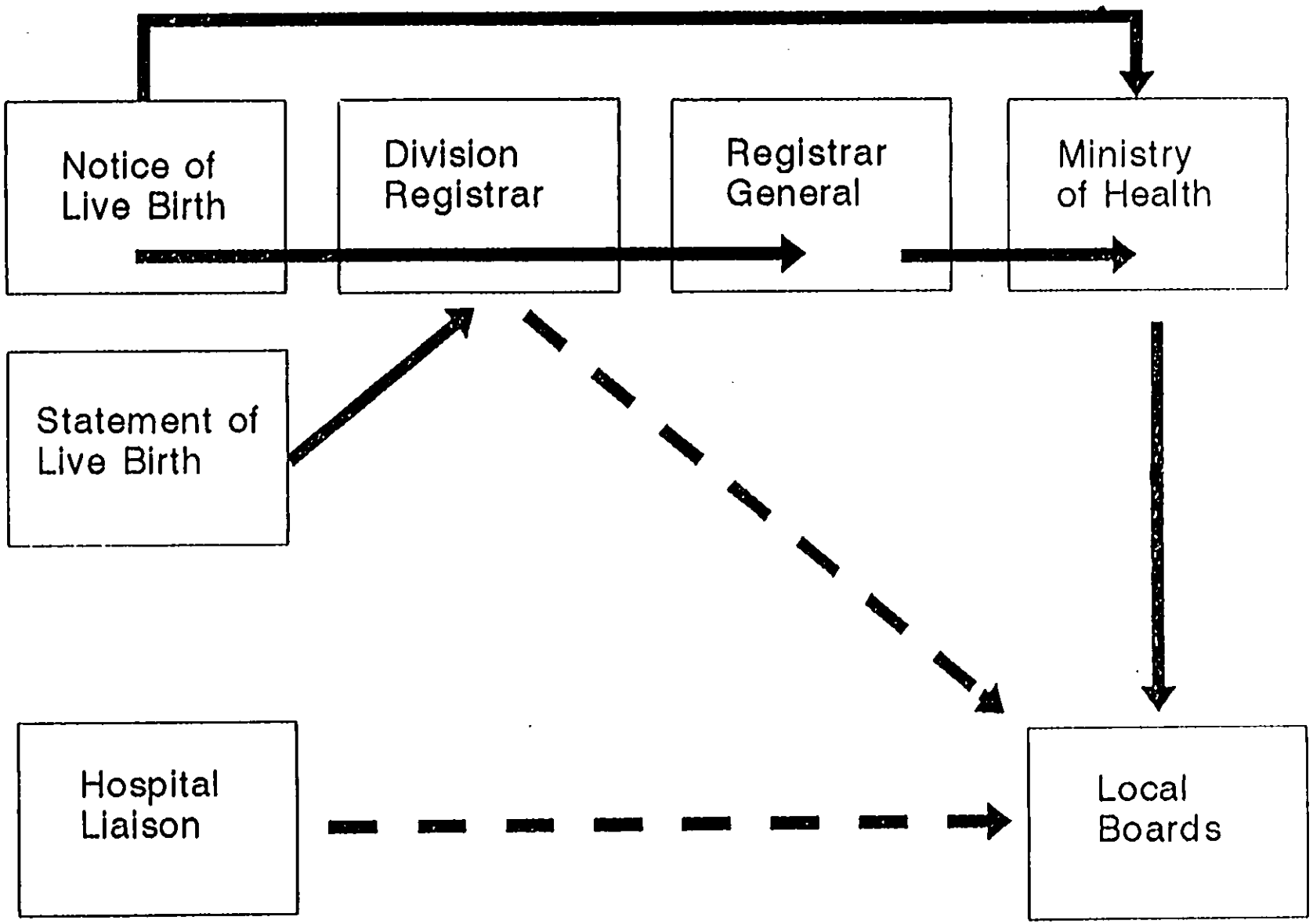


Figure #2

Still-Births

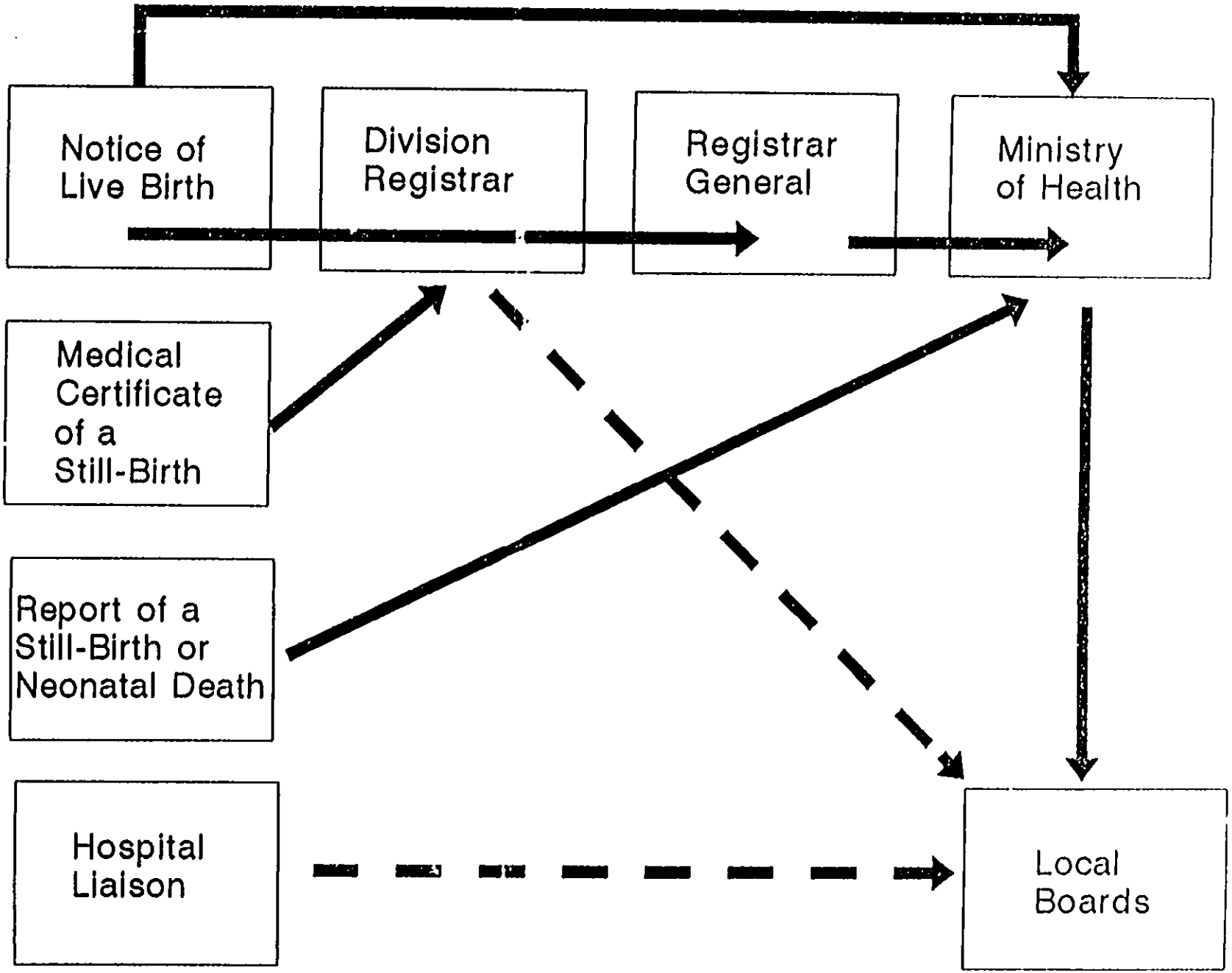


Figure #3

PROVINCE OF ONTARIO

VITAL STATISTICS ACT, SECTION 5

FORM 1

NOTICE OF LIVE BIRTH OR STILLBIRTH

ORIGINAL FOR DIVISION REGISTRAR

OHIP NUMBER

1 MOTHER	Surname / Forename(s)		2 AGE (years)
3 MARITAL STATUS	☐ Single ☐ Married ☐ Widowed ☐ Divorced	4 OCCUPATION	5 Telephone Number
6 PERMANENT ADDRESS OF MOTHER	(Street and number)		
	(City, town or village)	(County/District) / Province	Postal Code
7 PLACE OF BIRTH OF CHILD	(Home) ☐ (Hospital) ☐ (Name and location of hospital)		8 DATE OF BIRTH OF CHILD (Month, by name) (Day) Year)
9 SEX OF CHILD	(Male) ☐ (Female) ☐	10 BIRTHWEIGHT OF CHILD (Lbs and oz) or (Grams)	11 GESTATION PERIOD (In completed weeks)
12 WAS CHILD BORN ALIVE	(Yes) ☐ (No) ☐	13 KIND OF BIRTH (Single) ☐ (Twin) ☐ (Triplet) ☐ (Other) ☐	14 NUMBER OF PREVIOUS BIRTHS (Live births) / (Stillbirths)
15 WERE ANY CONGENITAL ANOMALIES NOTED AT BIRTH	(Yes) ☐ (No) ☐ If "Yes", describe		
CERTIFICATION	I CERTIFY THAT I WAS THE ATTENDING MEDICAL PRACTITIONER ☐ NURSE ☐ AT THIS BIRTH		Signature
NAME OF ATTENDING PHYSICIAN OR NURSE	Surname (please print) / Forename(s)		Date signed (Month, by name) (Day) (Year)
POST OFFICE ADDRESS	(Street and number, city and province)		Postal Code

INSTRUCTIONS PLEASE REMOVE CARBON AND THIS STUB BEFORE MAILING

The ORIGINAL (black copy) of this form to be sent WITHIN 48 HOURS to the DIVISION REGISTRAR OF BIRTHS, DEATHS (The Vital Statistics Act, Sec. 5)

The DUPLICATE (red copy) of this form to be mailed WITHIN 48 HOURS to the ONTARIO MINISTRY OF HEALTH

57655

8-2370-691 1992-11-17

Figure #4

Form 2
VSA

Province of Ontario (Canada)
Office of the Registrar-General

STATEMENT OF LIVE BIRTH

Registration No. (Office use only)

This is a permanent legal record.
Type or print plainly in blue or black ink and complete all items.

IMPORTANT: See reverse side for instructions

Cette formule est disponible en français.

CHILD'S SURNAME	1	2 Sex of child
FORENAME(S)		
DATE OF BIRTH	3 Month (By name) Day Year	4 Name of hospital (If not in hospital give exact location where birth occurred)
PLACE OF BIRTH	5 City, town, village, township (By name) Regional Municipality, county or district	For office use only

PARENTS	FATHER	MOTHER
NAME	6 Present surname	9 Present surname
	Forename(s)	Surname at birth
	Surname at birth	Forename(s)
	Other surname(s)	Other surname(s)
BIRTHPLACE	7 City/town/village	10 City/town/village
	Province/Country / For office use only	Province/Country / For office use only
DATE OF BIRTH	8 Month (By name) Day Year / Age	11 Month (By name) Day Year / Age
RESIDENCE OF MOTHER FOR STATISTICAL PURPOSES ONLY	12 Complete street address (If rural give exact location) City, town, village, township	Postal Code / For office use only
MAILING ADDRESS	13 Complete street address (If different from above) If rural give Post Office or Rural Route address	Postal Code
OTHER	14 Duration of pregnancy (in weeks)	15 Number of children ever born to this mother (including this birth): Number Liveborn; Number Stillborn after 20 weeks pregnancy
	16 Weight of child at birth: Grams ____ or ____ lb ____ oz	17 Kind of birth: single ☐ twin ☐ triplet ☐ other ☐
	18 If twin, triplet, state whether this child was born 1st, 2nd or 3rd	
ATTENDANT	19 Name and address of attendant at birth	Physician ☐ Nurse ☐ Other ☐

BEFORE SIGNING SEE ITEM F ON REVERSE SIDE

CERTIFICATION OF INFORMANT	20 I (We) certify the foregoing to be true and correct to the best of my (our) knowledge and belief		Day / Month / Year	
		Signature of Mother		
	20a We have agreed that the child's last name will be as shown in item 1. above ☐ Yes ☐ No	Signature of Father		
		Signature of Informant (other than Mother or Father)		
	21 Before completing this section see item E on reverse side ☐ Check here if the name selected is determined to be consistent with the child's Cultural, Ethnic or Religious Heritage			

DO NOT WRITE BELOW THIS LINE – OFFICE USE ONLY

CERTIFICATION OF DIVISION REGISTRAR	I am satisfied as to the correctness and sufficiency of the statement and register the birth by signing this statement. Signature of Division Registrar		
	Registration Number	Code Number	Date Month Day Year
For office use only			

For stillbirths, there are two documents that need to be completed in addition to the Notice of Live Birth or Stillbirth. A *Report of Stillbirth or Neonatal Death* (Figure 5) is completed by the hospital administration, containing demographic information about the father, the mother and a brief obstetrical history, and the sex, weight, gestation and probable time and cause of death of the fetus. This form is then forwarded to the Ministry of Health. The other document is the *Medical Certificate of a Stillbirth* (Figure 6). This form is completed by either the attending physician or coroner and contains more information regarding the circumstances of the birth itself, including birthweight, kind of birth (single/twin, etc.), the actual medical cause of stillbirth, type of delivery, whether the fetus died antepartum or intrapartum, and autopsy findings. This form must be filed with the Division Registrar before a burial permit can be issued.

Figure #5

Ministry of Health
Ontario

Form 3
Public Hospitals Act

Report of
Stillbirth or Neonatal Death

As required by Subsection 16(4) of Regulation 518 under the Public Hospitals Act.

I, _____ administrator

of the _____ hospital in the _____
(city, town, township or village)

of _____ in the _____
(county or as the case may be)

of _____ report that a _____
(stillbirth or neonatal death)

occurred in the above-named hospital on the _____ day of _____, 19____

at _____ o'clock in the _____ noon.

I have the following information with respect to this _____
(stillbirth or neonatal death)

1. Father,
   - *(a)* name: _____
   - *(b)* address: _____
2. Mother,
   - *(a)* name: _____
   - *(b)* hospital register number: _____
   - *(c)* address: _____
   - *(d)* marital status: _____
   - *(e)* racial origin: _____
   - *(f)* number of previous pregnancies,
     - (i) abortions: _____
     - (ii) stillbirths: _____
     - (iii) live births: _____
   - *(g)* any abnormalities associated with previous deliveries: _____
3. If mother is dead give,
   - *(a)* date of her death: _____
   - *(b)* cause of her death: _____
4. If a stillbirth give,
   - *(a)* sex: _____
   - *(b)* weight: _____
   - *(c)* period of gestation at time of extraction or expulsion: _____
   - *(d)* probable time of death: _____
5. If a neonatal death give,
   - *(a)* sex: _____
   - *(b)* weight at birth: _____
   - *(c)* date, hour and minute of death: _____
   - *(d)* cause of death:

I		
IMMEDIATE CAUSE - State the disease, injury or complication that caused death, not the mode of dying, such as heart failure, asphyxia, asthenia, et cetera.	i	_____
		due to
MORBID CONDITIONS, if any, giving rise to immediate cause (state in order backwards from immediate cause)	ii	_____
		due to
	iii	_____
II		
OTHER MORBID CONDITIONS (if important) contributing to death but not causally related to immediate cause.		_____

Instructions

1. The morbid conditions relating to death are divided into two parts. In Part I are those conditions causally related to the "IMMEDIATE CAUSE" and in Part II those not causally related thereto. In most cases the completion of Part I will be sufficient. Detailed certification is not required, the entry of a single cause is preferable in cases where a single cause is sufficient, as in Example I. Where the person completing the certificate finds it necessary to record more than one

(continued on reverse)

Figure #6

Form 8
VSA

Province of Ontario (Canada)
Office of the Registrar-General

To be completed by attending physician or coroner.

Cette formule est disponible en français

MEDICAL CERTIFICATE OF STILLBIRTH

Registration No. (office use only)

THIS IS A PERMANENT LEGAL RECORD. TYPE OR PRINT PLAINLY AND COMPLETE ALL ITEMS
(See reverse for instructions, and legal requirements under the VITAL STATISTICS ACT.)

PERSONAL PARTICULARS OF DECEASED CHILD

NAME OF CHILD

1. Surname — Forename(s)
2. SEX

DATE OF STILLBIRTH

3. Month (by name), day, year
4. Name of hospital (if not in hospital, give exact location where stillbirth occurred)

PLACE OF STILLBIRTH

5. City, town, village or township (by name) — Regional municipality, county or district

OTHER PARTICULARS

6. Mother – Surname / Forename(s)
7. Weight of child at birth — Grams _____ or lb _____ oz
8. Kind of birth — single ☐ twin ☐ triplet ☐
9. If twin, triplet, state whether this child was born 1st, 2nd, or 3rd

CAUSE OF STILLBIRTH

10. MEDICAL CERTIFICATE OF STILLBIRTH

I

Disease or condition directly leading to stillbirth (This does not mean the mode of dying. It means the disease, injury, or complication which caused stillbirth.) (a) _____ due to (or as a consequence of)

Antecedent causes
Morbid conditions, if any, giving rise to the above cause, stating the underlying condition last. (b) _____ due to (or as a consequence of)

(c) _____

II

Other significant conditions contributing to the stillbirth, but not related to the disease or condition causing it.

OTHER INFORMATION

11. a) Was there manipulative, instrumental or other operative procedure for delivery? Yes ☐ No ☐ b) If so, was foetus dead before the procedure? Yes ☐ No ☐

c) State nature of procedure _____
(Such as low, middle or high forceps; version and extraction, Caesarian section, craniotomy)

12. Did death occur before labour? Yes ☐ No ☐ During labour? Yes ☐ No ☐ Was labour induced? Yes ☐ No ☐

13. a) Was there an autopsy? Yes ☐ No ☐ b) If so, state findings _____

CERTIFICATION (Attending physician, coroner, etc.)

14. I Certify that I was ☐ was not ☐ in attendance at this stillbirth and that the statements herein are true and correct to the best of my knowledge and belief. Signature (attending physician, coroner, etc.)

15. Designation — Attending physician ☐ Coroner ☐

16. Name of physician or coroner (print or type) — Date signed – Month (by name), day, year

Address

CERTIFICATION OF DIVISION REGISTRAR

I am satisfied as to the correctness and sufficiency of this medical certificate of stillbirth and the statement of stillbirth and I register the stillbirth by signing this certificate and the statement of stillbirth. Signature of Division Registrar

Registration Number — Code Number — Date: Month (by name), day, year

This form and Form 7 (Statement of Stillbirth) must be filed with a Division Registrar before a burial permit can be issued.

For Office Use Only

8-2370-58.1: 1993-02-02

2.2 Framework for Evaluating Public Health Surveillance Systems

Public Health surveillance systems and attributes which make them useful and effective have been discussed by Klaucke [3]. He lists the following important attributes of a surveillance system:

1) Simplicity
2) Flexibility
3) Acceptability
4) Sensitivity
5) Predictive Value Positive
6) Representativeness
7) Timeliness

The PNOB form, itself, is a fairly simple form to complete. None of the information contained in it requires extensive questioning of the mother on sensitive issues. The form contains only one piece of information (eg. were any congenital anomalies noted at birth) which may not be known within moments of birth. The processing of the form is a bit more complicated. The form is sent from the hospital to the Division Registrar, the Registrar General, the Ministry of Health, and finally to the Regional Health Department.

A useful surveillance system should have flexiblity to adapt to changing information. The present definition of a stillbirth is one born either after 20 weeks or greater than 500 grams. The definition of stillbirth is not constant across geography and time. Reporting of stillbirths and neonatal mortality was discussed at a recent Association of Public Health Epidemiologists of Ontario workshop [4].

Prior to 1988, the Public Hospitals Act (Provincial) only specified gestational age (20 weeks) in its definition of stillbirth. Since 1988 it has become consistent with the Federal Vital Statistics Act. Other provinces have various definitions, many specifying weight but not gestational age. Australia, for example uses the definition of 22 weeks or 500 grams. The WHO defintion (1977) specifies 500 grams but requires 22 weeks if birthweight is unavailable. The decision of whether a baby meets the criteria of a stillbirth (20 weeks or 500 grams) is left to the person completing the form. A baby born today in Ontario at only 19 weeks and weighing 520 grams would be classified as a stillbirth. Prior to 1988 this baby would have been called a late miscarriage. For the sake of historical comparisons within a province, and comparisons to provinces and countries with different definitions of stillbirth, the system would be more flexible to change if all deliveries and miscarriages (within reason) were required to be reported.

Acceptability reflects the willingness of individuals to participate in the system. Given that the completion of the form is required by the Vital Statistics Act, individuals do not truly have a choice in whether or not to participate in the system. Klaucke lists several factors which may influence acceptability. Some of these factors are the public health importance of the event, time burden, and legislative requirements for reporting. Overall, the surveillance system does appear to be acceptable, largely as a result of the simplicity of completing the form.

The vast majority of PNOB forms are complete upon arrival at the Health Department.

The next two attributes of the system (sensitivity and predictive value positive) are the major focus of this study and will be discussed in a following section.

PNOB database operated by the Ottawa-Carleton Health Department is repr ll births in the region. Since PNOB completion is required by the Vital Sta t, it is reasonable to expect a virtually 100% completion rate. One potential threat to its representativeness lies in the reporting of the smallest, most immature stillbirths, the reporting of which may have been affected by changing defintions of stillbirth.

Timeliness of reporting information is the final attribute of a surveillance system listed by Klaucke. The PNOB form must be sent to the Division Registrar within 48 hours of the birth. From the Division Registrar, the PNOB form must make a series of stops before arriving at a regional health department, perhaps months later. For detecting and controlling epidemics of stillbirths, this time delay may be unacceptable, but for long-term planning by the health department, this delay may be quite acceptable.

2.3 Issues in Data Quality

Along with an increased use of large databases comes concern regarding the quality of each of these databases, and databases in general. In a study of 500 births in the State of Georgia, Floyd, Lavoie and Terry examined the rate of 'perfect agreement' between birth certificates and hospital records [5]. They found a range of agreement from 99.5% (95% CI 98.8-100%) for date of birth to 12.9% (CI 9.6-16.2%) for number of prenatal visits. Of particular note was the agreement regarding sex (93.8%; CI 91.5-96.2), birthweight (80.9%; CI 77.1-84.7%), number of living children (75.4%; CI 71.2-79.6%), and date of last normal menstrual period (35.0%; CI 30.3-39.7%).

In a similar study from New York City, researchers compared the birthweights reported to the Health Department to those contained in the medical records charts of nearly 4000 low birthweight neonates [6]. A discrepancy was defined as a disagreement between the two sources of more than 30g. The authors found an overall rate of 87% (CI 85.9-88.1%) concordance within 30g (or ± 1 ounce) and a statistically significant positive relationship between birthweight and the proportion of concordance, ranging from 79% (CI 67.9-90.1%) for babies weighing less than 500g to 94% (CI 93.0-95.0%) for babies weighing between 2000 and 2500g. The authors also found that birthweights tended to be lower in the Health Department's database as compared to the medical records, indicating that an

analysis of the Health Department's database alone could have led to a false conclusion of an increased rate of low birthweight babies.

In a study from Finland, Teperi studied all 1987 births (n=59370) registered in the Finnish Medical Birth Registry to examine the accuracy of thirty variables [7]. A stratified random sample of 775 maternal hospital medical records was abstracted. Only 760 of these records corresponded to a record in the database. Excellent or satisfactory agreement was obtained for most (26) variables, including sex (99.7%; CI 99.3-100%), number of previous livebirths (99.1%; CI 98.4-99.8%), birthweight (exact 99.2%; CI 98.6-99.8%; ± 100g 99.9%; CI 99.7-100%) and place of residence (97.4%; CI 96.3-98.5%). The author found the most notable example of unsatisfactory agreement for gestational age (exact 82.6%; CI 79.9-85.3% ± 6 days 93.9%; CI 92.2-95.6%). Teperi did examine the accuracy of whether a baby was born alive and found a simple agreement of 99.9% (CI 99.7-100%). Based on the extremely low rate of stillbirth in the Finnish registry of 4.3 per 1000 total births, the rate of simple agreement is expected to be extremely high. Without further analysis of sensitivity and positive predictive value, this figure of 99.9% accuracy is not easily interpreted.

The distributions of birthweight and gestational age are critical due to their relative important in perinatal research. On these two key variables, Teperi found that birthweight and gestational age were missing in 0.04% and 0.2% of cases, respectively [7]. Of the 48 babies with extreme birthweights (<500g or >5500g) 45

had consistent data in the medical records chart to support such extreme values. In contrast to birthweight data, 69 of 79 cases with extreme gestations (<20 weeks or >44 weeks) were obvious errors. Missing gestational age data were slightly more likely to come from the population of babies weighing less than 1500g, although this was not statistically significant.

Although comparing available data to that contained in a hospital medical records chart is an excellent method for assessing data quality, it is not the only method. David described and examined the quality and completeness of all live births to residents of North Carolina over a three year period by examining descriptive statistics (range and pattern of distribution of values) and assessing patterns of missing values [8]. He found that the reporting of maternal characteristics and birthweight was essentially complete, but gestational age was missing in 18.3% (CI 18.2-18.5%) statewide and varied considerably by county (0.4% to 60.4%) and by hospital (0% to 65.6%). The quality of gestational age data as measured by the range of values was found to be highly suspect ranging from 1 to 99 completed weeks of gestation, with at least one birth in each of the 99 categories. He also found that 2.3% of all reported gestational ages of livebirths were reported between 46 and 56 weeks, while a further 0.5% were reported under 20 weeks or beyond 56 weeks. Unfortunately, these errors do not seem to be randomly distributed among the entire population of births; among the records with missing or suspicious gestational age, the babies were, on average, 100 grams

lighter and were more likely to be born to young unmarried non-white women with less education than the rest of the population of mothers. David also found, not surprisingly, evidence of digit preference in the reporting of birthweights, with overrepresentation of full, half, and quarter pounds. Otherwise, the frequency distribution of birthweights approximated a normal distribution, though slightly skewed to the left.

More recently, Buescher et al compared 395 birth certificates, also from North Carolina, to the corresponding maternal hospital medical records [9]. They found excellent agreement for birthweight and one-minute Apgar score (100%), and moderate agreement for month prenatal care began, number of prenatal visits, maternal weight gain, and method of delivery ranging from 78.9% (CI 73.4-84.4%) to 91.9% (CI 87.9-95.9%). For tobacco and alcohol use agreement was assessed for those cases where both the birth certificate and medical record indicated some use. Agreement was estimated at 84.4% (CI 74.1-94.7%) and 56.2% (CI 35.0-77.4%), respectively. If all cases, including those with no tobacco or alcohol use, were assessed, the agreement was 96.1% and 98.1%. The group concluded that the data contained on the new U.S. birth certificate will be able to support valid conclusions of aggregate analyses.

All of these studies attempted to quantify the amount of agreement found among a number of different variables. The sources of data and the countries in which the data was collected varied greatly and, not surprisingly, so did the level of

agreement. Most found the greatest degree of agreement for such easily measured variables as date of birth, sex and birthweight. Lesser agreement was obtained for less-easily observed variables like number of prenatal visits, number of previous livebirths, gestational age, and alcohol and tobacco use. Accuracy of whether a baby was born alive after correction for chance, has not been established for any perinatal database. Given the large range of agreement reported in the literature, it is important to examine the PNOB database used in Ottawa-Carleton to assess its own quality and to compare it to systems used in other countries.

2.4 Methods Used to Assess Quality of Data

As will be shown, the methods used in this thesis to evaluate the accuracy of the reporting of stillbirth/livebirth are similar to those routinely used in evaluating the performance of diagnostic tests; namely sensitivity, specificity, and positive and negative predictive values. If the PNOB database is thought of as a screening test and the logbook and/or medical records as the "gold standard", then the rationale behind this type of analysis becomes more clear. This is a similar approach recently taken in Israel by Pipel et al. [10].

Pipel et al compared information from a centrally located multi-hospital database to information in the operating room logbooks. The purpose of this database was to provide admission, transfer and discharge data, and included such information as name, date of birth, sex, marital status, diagnoses, surgical procedures, and other valuable information. The statistics measured to evaluate

the database were sensitivity in detecting whether a patient underwent a certain surgical procedure, and positive predictive value of the same. All cases were identified from both the computerized database and the operating room logbooks. When there was a discrepancy between the database and the logbook, the patient's medical chart was reviewed.

Results of this study showed sensitivites for three surgical procedures ranging from 90 to 99%. Positive predictive values were even more impressive, ranging from 96-99%. The study was limited to identifying surgical procedures and did not include an evaluation of any other types of data available to them.

When comparing two methods of measuring the same information, a clear understanding of the issue of association versus agreement is imperative. A common method used to compare two associated variables is the Pearson product-moment correlation coefficient (r). Pearson's r is an index of trend, ranging from -1 to +1, representing the strength of the tendency for changes in one variable to be accompanied by changes in the other. When r equals +1, there is a perfect linear relationship between an increase in x and an increase in y. Perfect correlation does not in any way mean that x equals y, for x may be systematically greater than y, yet still produce this relationship. Therefore, the effects of systematic bias are left unaddressed if one attempts to assess agreement with the Pearson correlation coefficient.

Altman and Bland have described other methods often used inappropriately in analyzing data from method comparison studies [11]. Comparing group means for the purpose of evaluating agreement seems logical. Some may conclude that if the two methods yield the same mean value, that the methods are equivalent. Comparison of means is remarkably similar to the concept of ecological fallacy, in which one cannot extrapolate the findings of a group to individuals within the group. Similarly, one cannot conclude that if the difference of the means of grouped data is not statistically significant that the individual pairs are in agreement.

A third technique commonly used but often misused for assessing agreement between pairs of measurements is the linear regression model. Significance testing for the least squares model usually involves testing the regression coefficient against zero, which, as in the case of the correlation coefficient, is inappropriate. If two measurement methods were equivalent, one could reasonably expect the slope of the regression line to be equal to the line of unity (slope = 1). According to Altman & Bland [11] the expected slope of the regression line is:

Equation 1: Expected Slope of Regression Line

$$\beta = \sigma_{T^2} / (\sigma_{A^2} + \sigma_{T^2})$$

where σ_{T^2} is equal to the variance of the true values and σ_{A^2} is equal to the variance of measurement error for method A. The value of this equation can only

be equal to or less than 1. How much lower than 1 will depend on the amount of measurement error of the method chosen as independent (i.e. method A).

Fortunately, methods have been developed which specifically address the need to assess agreement rather than association. For nominal data, the index of choice is the kappa statistic, which corrects for agreement expected by chance [12]. Kappa was first proposed in its current form by Jacob Cohen [13]. The formula for kappa is as follows:

Equation 2: Kappa

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where p_o is the overall proportion of observed agreement and p_e is the overall proportion of chance-expected agreement based on the marginal probabilities. Kappa is a ratio of the observed excess agreement beyond chance to the maximum possible excess agreement beyond chance. Kappa takes on its maximum value of +1 when there is perfect agreement. If observed agreement is greater than that expected by chance alone, k>0; if observed agreement is less than that expected by chance, k<0. Kappa takes on its minimum value of -1 only "in the special case when one or both observers choose each of two categories exactly half the time" [12]. The standard error of kappa was also approximated by Cohen [13]:

Equation 3: Standard Error of Kappa

$$\sigma_\kappa = \sqrt{\frac{p_o(1-p_o)}{N(1-p_e)^2}}$$

Kappa can be expanded to assess agreement for nominal data with more than two observers or more than two categories. The overall value of kappa is calculated as above except that it is based on the sums of individual observed proportions of chance-expected agreement and the sums of the individual expected proportions of chance-expected agreement.

This overall kappa index could also be used in the analysis for agreement of ordinal data. Although intuitively attractive, the overall kappa does not utilize the fact that the data are ordered, and treats all pairs not in perfect agreement equally. With ordinal data, it is often important to account for near agreement. Another version of kappa, known as the weighted kappa, gives partial credit for near agreement. Perfect agreement, on the leading diagonal in a k x k table, is given the highest weight of +1, while less than perfect agreement is given a lesser weight between 0 and +1. Several authors have differing opinions regarding which weighting system is used for less than perfect agreement. Cohen proposed that weights be equal to the square of the deviation of each pair from exact agreement [14]. Fleiss and Cohen later demonstrated that when the squared deviation weights are used, the weighted kappa becomes equivalent to the Intraclass Correlation Coefficient (ICC), another well-recognized index of agreement which will be discussed next [15]. Although Cicchetti and Allison proposed that the absolute deviation be used for weighting the errors [16], the consensus appears to

be that the most logical choice for weighting near agreement is the squared deviation version because of its equivalency to the ICC [17].

In assessing reliability of measurements on a continuous scale, the intraclass correlation coefficients (ICC) are often used. The ICC family is based on the ANOVA model, either one-way or two-way with fixed or random effects. The ICC is the proportion of total variability accounted for by the variability among persons. A property of the ICC which is difficult but essential to appreciate is that for the same degree of within-subject variance, the greater the between-subject variance, the greater the ICC. The situation in which there is very little between-subject variation is known as a 'homogenous setting'. In homogenous settings measures of reliability are difficult to interpret because these settings do not allow raters to use the full range of the measurement system, leaving most of the scale untested [18].

There are several variations of the ICC, each of which can give quite different results when used on the same data. Each form is appropriate for specific situations, based on the design and conceptual intent of the study. For the comparison of individual ratings (as opposed to a 'consensus' or the mean of several ratings), Shrout & Fleiss give three versions of the ICC [19].

The first version is based on the one-way ANOVA and is appropriate when each subject is rated by a different set of raters, randomly selected from a larger population of raters. The formula is as follows:

Equation 4: ICC (1,1)

$$ICC(1,1) = \frac{BMS - WMS}{BMS + (k-1)WMS}$$

where k is the number of raters, BMS is the between-subjects mean square and WMS is the within-subjects mean square. ICC(1,1) provides the most conservative estimate of reliability and as a result, will underestimate the ICC if used under inappropriate conditions [20].

The remaining two versions of ICC are both based on the two-way ANOVA model, and are appropriate when each subject is rated by the same k raters. In the second, ICC(2,1), it is assumed that the k raters have been randomly selected from a larger population of raters, while in the third, ICC(3,1), it is only the fixed set of raters in the study that are of interest. The equations for the second and third variants of the ICC follow:

Equation 5: ICC (2,1)

$$ICC(2,1) = \frac{BMS - EMS}{BMS + (k-1)EMS + k(JMS - EMS)/n}$$

Equation 6: ICC (3,1)

$$ICC(3,1) = \frac{BMS - EMS}{BMS + (k-1)EMS}$$

where WMS has been partitioned into EMS (mean square of the error) and JMS (between-judges mean square) and n is the number of persons or targets.

Suppose that we have data from a reliability study where each rater assesses each subject. When it comes to the analysis, the researcher may decide to combine all data, in which the effect of the different raters will contribute to the overall variability in the ratings. In this scenario, it would be appropriate to use the ICC(2,1), with a two-way random-effects ANOVA. Now, suppose the researcher analyzed the ratings from each rater separately and then pooled the results. In this case, the ICC(3,1) would be more appropriate, using a two-way fixed-effects ANOVA. Shrout and Fleiss suggest that the second variant gives more conservative estimates of interobserver agreement than the third variant [19].

Confidence intervals can be calculated for ICC(1,1) and ICC(2,1), but for ICC(3,1) "no simple method is available, either exact or approximate, for constructing a confidence interval for the parameter" [21]. For ICC(2,1) (the version of ICC to be used in this paper) the two-sided 95% confidence interval can be approximated by the following two-step formula:

Equation 7: Degrees of Freedom for the F-Ratio

$$v^* = \frac{(n-1)(k-1)(kR_1F + n[1+(k-1)R_1] - kR_1)^2}{(n-1)k^2R_1^2F^2 + (n[1+(k-1)R_1] - kR_1)^2}$$

where v* is used to determine the degrees of freedom for the F value, R_1 is the ICC(2,1) previously defined above, k is the number of raters (in this paper k = 2), and where F is from the two-way ANOVA and is equal to Mean Square (raters) divided by Mean Square (residual). $F^* = F_{1-\frac{1}{2}\alpha}[(n-1), v^*]$ and $F_* = F_{1-\frac{1}{2}\alpha}[v^*, (n-1)]$

Equation 8: Two-Sided 95% Confidence Intervals

$$\frac{n(BMS - F_{n-1,v^*}EMS)}{F_{n-1,v^*}[kJMS + (kn-k-n)EMS] + nBMS} < R_1 < \frac{n(F_{v^*,n-1}BMS - EMS)}{kJMS + (kn-k-n)EMS + nF_{v^*,n-1}BMS}$$

The fact that several variants of the ICC exist does not seem to be well-recognized in the literature. Authors using the ICC formula, including those writing articles of a more technical nature often fail to acknowledge which they are using and even that there is more than one [12,22].

Other methods have been proposed for analyzing interobserver agreement studies. In the epidemiological literature, the logistic model has been suggested for assessing interobserver agreement [23]. This method is particularly useful in that it allows for stratification or adjustment for covariates. While the authors claim that it can be used to examine the degree of agreement (akin to the weighted kappa), conventional computer software allows for only a dichotomous assessment of agreement, similar to the simple, unweighted kappa.

For the assessment of agreement of continuous variables, Altman & Bland suggest a simple approach based on the ANOVA and simple graphical techniques

[11]. A plot of method A vs. method B with a linear regression line should be a mandatory first step. The important second step is to plot the difference between the methods (A-B) against the average (A+B)/2. With this type of a plot, patterns of disagreement can be readily observed. According to Altman & Bland the mean of these (A-B) differences will be the relative bias and the standard deviation of these differences will be an estimate of the error [11], allowing one to separate systematic and random error which the ICC combines into a single measure [24]. In addition to calculating the mean difference, one can also examine the data to see if there is any association between the size of the differences and the magnitude of the measurements.

In the statistical literature, log-linear models have been used for assessing agreement for categorical data [25,26]. Another index, the concordance correlation coefficient (CCC), has been suggested. Similar to the ICC, the CCC assesses the agreement between the observed data and a 45° line passing through the origin. First postulated in 1989 by Lin, this statistic can apparently be broken down into two indices, one assessing the goodness of fit to the regression line and the other comparing the slope and intercept of the observed line to a 45° line passing through the origin, thereby allowing the reader to interpret whether poor agreement was due to poor correlation or poor placement of the regression line between the axes [27]. In the paper describing this new statistic, there is no description of how to calculate the confidence intervals to accompany the CCC. While it has been said

that the CCC provides more information than the more widely recognized and accepted ICC [22], in this study the ICC and CCC were virtually identical, down to the fourth decimal place. For this reason, the resulting concordance correlation coefficients are not reported. This statistic has yet to receive any attention in the biomedical literature.

One other approach to the measurement of agreement deserves mention. Bartko, the father of the intraclass correlation coefficient [28], in the last few months has published a paper describing a 'bivariate confidence ellipse' which provides boundaries for dispersion [18]. Based on the graphical methods proposed by Altman and Bland [11], Bartko's ellipse gives visual guidelines for the scatterplot and is a convenient, simple way to appreciate the relative sizes of the within- and between-subject variances.

In summary, a well-developed body of literature regarding inter-rater agreement exists, straddling the traditional mathematical/statistical and the epidemiological/medical bodies of literature. The statistical methods for analyzing inter-rater agreement studies and method-comparison studies are still evolving. Many of the sophisticated statistical models and methods have yet to find their way into the mainstream epidemiological literature. Each method has its advantages as well as its limitations. No single method appears to be capable of providing the definitive answer to the question of inter-rater reliability. Tests of equality of means will be performed on data in this thesis: the paired t-test for equality of two

means and the Friedman test for equality of more than two means from paired data [29]. Levene's F will be used to test for homogeneity of variance [30,31].

Misclassification of a baby's vital status (liveborn vs. stillborn) plays a large role in this study. Several complementary epidemiological measures will be used to assess the accuracy of the PNOB database as a test of whether a baby was stillborn. Sensitivity is defined as the probability that an individual (baby) with a given condition (stillborn) will be classified by a certain test as having that condition. Specificity is the probability that an individual without a given condition (not stillborn) will be classified as not having that condition. Two other complementary measures which address the quality of a "test" are positive predictive value and negative predictive value. Positive predictive value is the likelihood that a baby with a positive test is truly stillborn whereas negative predictive value is the likelihood that a baby with a negative test is truly liveborn. These two measures, unlike sensitivity and specificity, vary as the prevalence of the condition under study changes in the population. For any given degree of sensitivity and specificity, as the prevalence of the condition decreases, the positive predictive value of the test decreases and the negative predictive value increases.

3. STUDY PROBLEM AND OBJECTIVES

The primary problem to be examined is the quality of the Ottawa-Carleton Health Department's Physician Notices of Birth (PNOB) database. Errors in the database could arise in two ways: 1) the data was entered incorrectly at the health department and; 2) information on the PNOB form itself is wrong when compared to the logbook and/or medical records chart. This study was designed to assess the effects of both possibilities. The specific objectives are:

1) To identify how PNOB records are completed in each maternity care hospital in the Regional Municipality of Ottawa-Carleton.

2) To assess the reliability of extracting information from the PNOB records to create a computer database, for the following:

   a) Health Department PNOB database records indicating livebirth - Sample A

   b) Health Department PNOB database records indicating stillbirth - Sample B

   c) Health Department PNOB database records indicating missing vital status - Sample C

   d) False negatives found in hospital logbooks but not in Health Department PNOB database

3) To compare the accuracy of the information in the Physician Notice Of Birth records to that in the hospital logbooks and medical charts for the following:

a) Health Department PNOB database records indicating livebirth - Sample A

b) Health Department PNOB database records indicating stillbirth - Sample B

c) Health Department PNOB database records indicating missing vital status - Sample C

c) False negatives found in hospital logbooks but not in Health Department PNOB database - Sample D

d) A subset of PNOB records from Samples A-D with perfect agreement between the existing and new PNOB databases and logbook - Sample E

The study protocol was submitted to the Research Ethics Board of each of the five Regional Municipality of Ottawa-Carleton hospitals with Labour and Delivery facilities. Ethical approval was given in writing from four hospitals and verbally from one hospital. The letters of approval are found in Appendix A.

4. METHODS

4.1 Study Population and Sample Size

This study includes records extracted from the existing PNOB database for the following samples: A) a random sample of livebirths from 1989-92; B) all stillbirths from 1989-92; C) all babies for whom vital status is missing in the database from 1989-92; D) False negatives (stillbirths incorrectly identified as livebirths in the existing PNOB database) and; E) a subset of babies in samples A, B, C, D for whom the existing PNOB database, new PNOB database and logbook are in perfect agreement for certain variables (to be defined in the following section).

All babies were born to women residing in the Regional Municipality of Ottawa-Carleton at one of the five Ottawa hospitals with labour and delivery facilities between January 1, 1989 and December 31, 1992. Both single and multiple births have been included in the study. Approximately 1.1% of women from Ottawa-Carleton deliver their babies at home or in a hospital outside the region. This study is, therefore, a population study since it encompasses roughly 99% of all births to women residing in the Regional Municipality of Ottawa-Carleton over the specified study period. Between January 1, 1989 and December 31, 1992 there were 39,149 records of births to Ottawa-Carleton women in Ottawa-Carleton hospitals that were present in the Health Department's PNOB database.

Sample A is a stratified random sample of 450 babies recorded as being liveborn in the health department's existing PNOB database. There were two ways to select this random sample. First, one could have selected births in proportion to the number born in each hospital and second, by selecting a fixed number from each hospital. The latter method was chosen for the practical purpose of distributing the workload among the hospital medical records departments.

A random sample stratified by hospital was drawn: 100 each from Hospitals A, B, C and D, and 50 from Hospital E. Fewer were chosen from Hospital E on the basis of expected language difficulties for this researcher in reading the caseroom logbooks and medical records in French. The sample size of 450 was selected because it was seen as a balance between a sufficient sample size and the maximum number of records that could be processed by this researcher, given the limitation of time and resources and the uncompensated work that was required by each hospital's medical records staff. Formal sample size calculations were not undertaken because the maximum possible sample size based on the above limitations was selected.

Sample B includes all 270 babies who were recorded as being stillborn in the health department's existing PNOB. Selecting all stillbirths was done to determine the number of false positives, allow the calculation of predictive value positive and to examine the accuracy of the database with respect to this small, but important group of births.

Sample C includes all 224 babies for whom vital status was missing in the existing PNOB database. Sample D included all babies whom the logbook indicated were stillborn but whom the existing PNOB database indicated were liveborn.

For Sample D a sample size of 100 records in perfect agreement between the existing and new databases and the caseroom logbook was selected. The variables examined for perfect agreement were gestational age, mother's age, number of previous livebirths, and sex of baby. For the variable birthweight, perfect agreement was defined as ± 28g (1 ounce). A similar allowance for perfect agreement for birthweight has been documented elsewhere [6]. Postal code was not one of the criteria for perfect agreement and will, therefore, not be included in the analysis of these perfect records.

4.2 Study Design

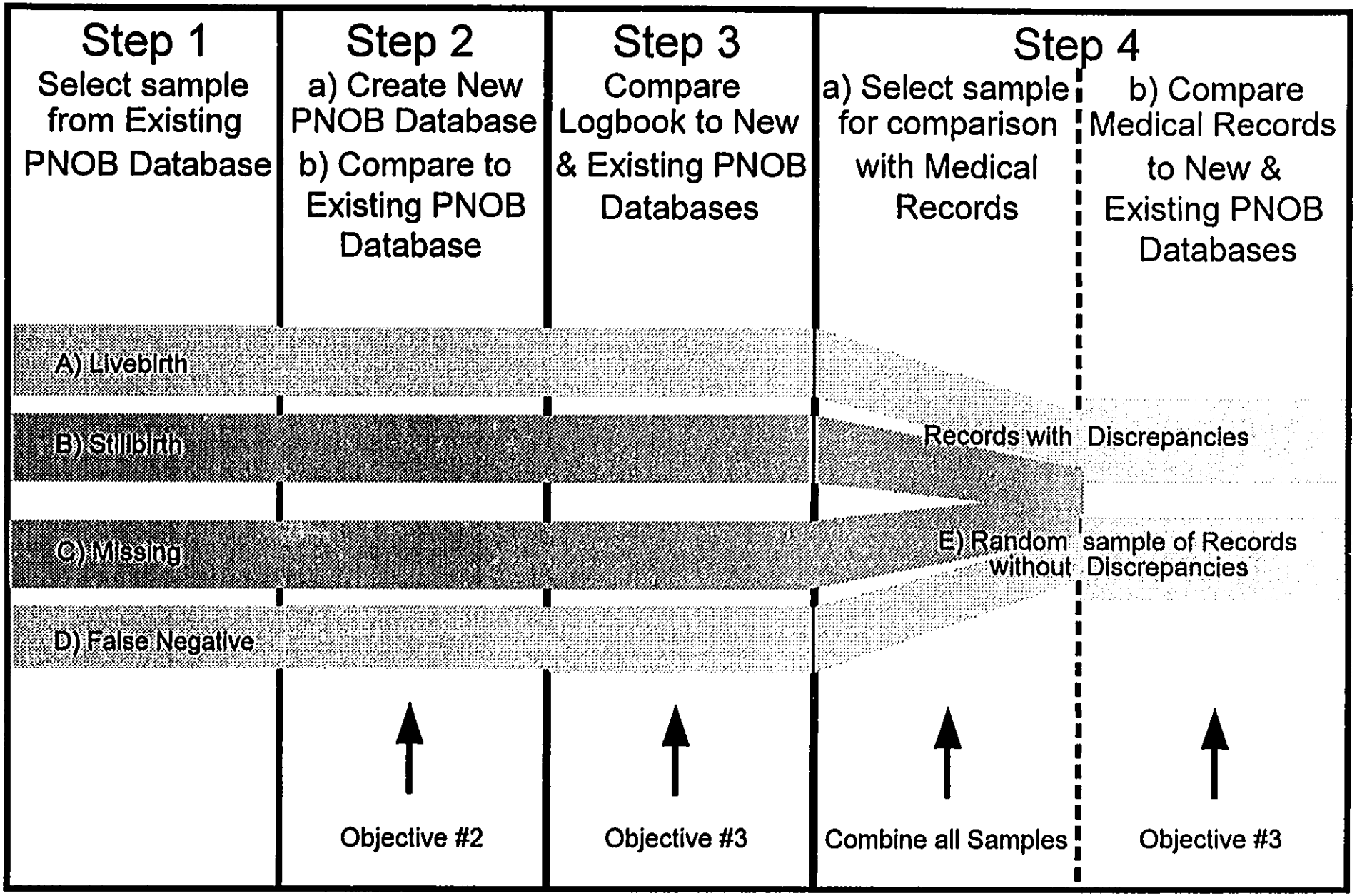
The first objective was addressed by conducting face-to-face interviews with the nurse managers of the obstetrical/labour & delivery department of each of the five hospitals providing maternity care in Ottawa-Carleton.

An overview of the study design for Objectives #2&3 is provided in Figure #7. The second objective (reliability of entering data from the PNOB form into a computerized PNOB database) was addressed by comparing the existing 1989-92 dataset created routinely by the clerk at the Ottawa-Carleton Health Department with a new dataset created by the primary investigator of this study. The data for this new dataset was abstracted very carefully from the PNOB forms to ensure it was a true representation of the actual data on the PNOB form.

The third objective was assessed by comparing the information on the PNOB forms to that in the caseroom logbooks. The logbook was presumed to be a reasonable and accessible proxy for the gold standard which was the hospital chart. If there was any discrepancy between the PNOB databases and the logbook, the medical chart was examined as the ultimate gold standard.

Objectives #2 and #3 of the study were studied for each of the first four sample populations (A - livebirths, B - stillbirths, C - missing vital status, D-false negatives). Both the existing and new PNOB databases and the caseroom logbook may be in agreement yet still be incorrect. Therefore, to complete Objective #3, a

Figure #7
Overview of Study Design



fifth sample was constructed from a random sample of babies for whom information was in perfect agreement among the existing and new PNOB databases and caseroom logbooks. Information contained in the PNOB databases and logbook was compared directly to that contained in the mother's hospital chart.

4.3 Data Collection

Since January of 1989 the Ottawa Carleton Health Department has employed a data entry clerk to enter the data from the PNOB form into a computerized database. Each PNOB that arrives at the Health Department is assigned a unique identification number which is printed on the PNOB form by the data entry clerk and used in the database. No identifying information (name, address, telephone number, doctor's name) is entered into the existing database.

The PNOB computerized files for each year from 1989 to 1992 were obtained from the Ottawa-Carleton Health Department. The files were then combined into one large SPSS-PC system file. This combined file was searched to identify all records of stillbirth, all records with a missing value for live or stillbirth, and a random sample of 450 livebirths, 100 from each large hospital, 50 from the smallest. Variables abstracted from this PNOB database file are shown in Table 1. A list of each I.D. number, day, month and year of birth was then prepared for each subgroup identified above.

Table 1
Variables Examined in the Study

VARIABLE	DESCRIPTION
ID	Unique identifier
DOBY	Year of birth (range 89-92)
DOBM	Month of birth (range 01-12)
DOBD	Day of birth (range 01-31)
MAGE	Age of mother (range 10-55)
SEX	Sex of infant (1 = male; 2 = female)
GEST	Gestational age in completed weeks (range 18-45)
WEIGHT	Birthweight in grams (range 250-6000)
VS	Vital status at birth (1 = live; 2 = stillbirth)
PLIVE	Previous live births (range 0-15)
PC	Postal code
NAME*	Mother's first and last name (used only to establish a positive match for the data linkage)

* The mother's name does not appear in the existing PNOB database but was extracted from the PNOB forms and appears in the newly-entered PNOB database

The Ottawa-Carleton Health Department archives all PNOB forms on-site. For each record of birth in the study (i.e. all stillbirths, random sample of livebirths, all births with unknown outcome and all false negatives) the computer record was matched by the unique I.D. number to the corresponding PNOB form. Each match was verified by I.D. number and date of birth only (i.e. not by mother's name). Data specified in Table 1 were abstracted from the PNOB forms and were entered on an

SPSS-PC Data Entry screen using a 386DX laptop computer to create the new PNOB database used to assess inter-rater reliability (Objective #2). Security of the data file was assured with one level of password protection on the computer file. The computer itself was kept in a locked room when not in use. Appropriate value limits were applied to each variable on the data entry screen (Table 1). In order to assess intra-rater reliability of data entry, ten percent of cases were double-entered at the end of each data entry session.

The next stage of data collection was designed to meet Objective #3 and involved identifying each case, based on PNOB data, in the hospital caseroom logbooks. For each hospital, a list was prepared for all study cases including the day, month and year of birth, as well as the mother's name (taken from the newly-entered PNOB database). Each list was taken to the appropriate hospital and the caseroom logbooks were examined to locate all study cases based on hospital, day, month and year of birth, and the mother's name. Although not ideal, this information was all that was available for data linkage and in virtually all cases was sufficient.

If a case was not located in the logbook for the particular date of birth, several other strategies were used to locate it. First, the days preceding and following the presumed date of birth were searched. Secondly, the numbers representing the day and month were switched when both were less than or equal to 12 and the cases on this revised date were examined. Finally, other hospital

logbooks were searched for the presumed date of birth (in case the wrong hospital was entered or the handwriting was illegible, e.g. OCH versus OGH).

Once a study case from the database was located in the logbook the same information listed in Table 1 was either transcribed to paper (4 hospitals) or entered directly to the database (1 hospital), based on the availability of the Health Department's laptop computer. At the end of each day, 10% of the cases were relocated and either re-transcribed to paper or re-entered to the computer. Once again, this was done to assess intra-rater reliability in abstracting and entering data.

The caseroom logbooks at all five hospitals for every day of the four year study period were manually examined once more to identify any additional cases of stillbirth that were not identified in the database. When an additional case was identified, the study information was recorded. If the address and postal code indicated that the mother was not a resident of the Regional Municipality of Ottawa-Carleton, the case was discarded. If she was a resident of Ottawa-Carleton, the existing PNOB database was then searched and all unique I.D. numbers for all births on that day, as well as the preceding and following days were recorded. Each potential match was then pulled from the archived PNOB forms until the match was found, based on the mother's name, hospital, and proximity to the correct date of birth. All but eight stillbirths discovered using this method were matched to a

record in the existing PNOB database. These eight stillbirths were not included in the data analysis.

At this point, a brief analysis was done to identify any discrepancies between the Health Department's existing PNOB database, the newly entered PNOB database and the caseroom logbooks . A random sample of 100 was drawn from all cases (live and stillborn) in which there was perfect agreement among all three sources. This group of 100 makes up Sample E.

In cases where the newly-entered database differed from both the existing database and the logbook, the PNOB form itself was examined for a second time. If there was an error made by this investigator, the error was corrected in the newly entered database, ensuring that the newly-entered PNOB database was as accurate a representation of the actual PNOB forms as possible. Furthermore, in all cases of discrepancy involving the logbook, a list was prepared with the mother's name and the baby's date of birth for each hospital and then submitted to the Medical Records department. At one hospital I was personally responsible for looking up the hospital identification number of each mother before submitting the list to the Medical Records Department.

Once the charts became available, the study data were extracted and transcribed to paper since no portable computer was available. Ten percent of cases were re-abstracted and re-transcribed for estimation of intra-rater reliability. All data were entered into SPSS-PC Data Entry.

The computerized files for the newly entered PNOB forms, the logbook and the medical charts were then linked, based on a match of unique I.D. number and year of birth. Once the linkage was made, a list was prepared of the mothers' names from each of these three data files in separate columns. A line-by-line visual inspection of the three columns was made to verify that the names were the same across each line, and that the record linkage was successful. After this manual verification, this combined file was linked by unique I.D. number and year of birth to the existing database file at the Health Department. A similar record linkage verification at this point was not possible since the existing PNOB database contains no personal identifying information.

Although the aim of this study did not include estimating the rate of duplicate entries, when found, the two entries were compared and then one was discarded. Duplicate entries could only be identified when both entries appeared among either the stillbirths or the random sample of livebirths.

4.4 Data Analysis

Data analysis was performed using the SPSS-PC, BMDP New System, and Microsoft Excel. Descriptive statistics were calculated for stillbirths and livebirths separately since it has been shown in New York City that among low birthweight babies, there is a positive relationship between increasing birthweight and the likelihood of concordance on birthweight between a health department's birth records and the neonate's medical chart [6]. If this finding was also true in Ottawa-

Carleton, it would be unwise to combine livebirths and stillbirths (often having low birthweight). Any measure of agreement for this non-random sample of live- and stillbirths would be invalid because being a stillbirth was the factor upon which this sample was stratified, yielding an artificial prevalence of stillbirth of nearly 25% in this sample.

Descriptive statistics appropriate to each variable type (nominal, ordinal, continuous) were calculated for the existing and newly entered databases and logbook, separately. They are presented for all three sources to demonstrate the comparability of the descriptive statistics obtained from databases of varying degrees of accuracy. For continuous variables (mother's age, birthweight, gestational age) means, medians, standard deviations, minimum and maximum values, and the proportion of missing values were calculated. (No missing values were imputed from the remaining data.) Values were plotted in a histogram to examine the overall distribution of values. Particular attention was paid to the most extreme values in both tails of the distributions for any outliers that may be present. If a biologically implausible value was present, the data collection papers were consulted to determine if it was a typographical error. In the case that it was such an error, the entry was corrected; if not, it was left alone. Frequency tables were produced for nominal and ordinal data (sex, previous live births, liveborn/stillborn).

For Objective #2 a comparison between the newly entered PNOB form database and the existing Health Department's PNOB database was made as a measure of inter-rater reliability prior to correcting the few errors that existed in the newly-entered PNOB database. Crosstabulations and kappa coefficients for ordinal and nominal data, and intraclass correlation coefficients for continuous data were examined. As outlined by Altman and Bland in their article on the analysis of method comparison studies [11], pairwise scatterplots of the raw data, as well as scatterplots of the difference between two methods against the average measurement are important tools in examining concordance. All of these graphs have been assembled and examined. However, the plots of raw data rarely demonstrate any additional information that can't be readily appreciated from the more useful plots of difference versus mean. For this reason, the raw data scatterplots will not be routinely presented in this paper. Because of the large sample sizes used for some subsamples, 'sunflowers' will be used to indicate coordinates on the graph to which more than one data point belongs; the more 'petals', the more data points a sunflower represents. Each sunflower is located at the mean of all points which it represents, rather than the geometric centre.

For Objective #3, pairwise comparisons were made between the health department's existing database, and the logbook and medical records to assess the database presently being used by the health department. Pairwise comparisons were also made between the newly entered database, and the logbook and medical

records to assess the quality of information on the PNOB form itself. Although overall comparisons between the four sources are technically feasible [19], it is the pairwise comparisons which are of interest in this study; therefore, intraclass correlation coefficients ((ICC(2,1)) were calculated for each pairwise comparison. This version of the ICC was selected because, in the words of Shrout and Fleiss, "in this instance the question being asked is whether the judges are interchangeable" [19]. That this version gives a more conservative estimate of agreement than the ICC (3,1) is also desirable. Weighted mean ICC values are calculated across all four variables for each pairwise comparison to give an overall idea of the confidence intervals around intraclass correlation coefficients are also given. Scatterplots identical to those used to meet objective #2 are also presented.

Sensitivity and positive predictive value were calculated for the variable "stillbirth" since this was the variable which had been previously identified as being inaccurate in the existing database. Since all babies identified as a stillbirth in the existing PNOB database and the logbook were selected, the 'a', 'b', and 'c' cells in a two-by-two table can be ascertained. However, the 'd' cell (true negatives) can not be determined. Thus, specificity cannot be calculated. While high specificity is a valuable attribute of a diagnostic test, Klaucke does not list it as an important public health surveillance system attribute [3]. Stillbirth is a rare "disease" occurring at a rate typically in the range of 5/1000 livebirths. Specificity of a diagnostic test in a population with such a low incidence of disease is expected

to be very high. To illustrate with a simplified version, imagine that there were exactly 40,000 total births in Ottawa-Carleton and that the stillbirth rate was exactly 5/1000, resulting in 200 stillborn babies. Now imagine that the screening test (in this case, the existing PNOB database) randomly assigned birth outcomes to incoming PNOB forms. The resulting hypothetical 2x2 table is shown in Table 2.

Table 2
Statistically Independent Classification of Stillbirth

		Reference Standard		
		Stillbirth	Livebirth	Total
Existing Database	Stillbirth	1	199	200
	Livebirth	199	39,601	39,800
	Total	200	39,800	40,000

In this 2x2 table the rows and columns are statistically independent of each other. The resulting specificity of this table is 99.5% (39601/39800). This figure means that if birth outcome were decided by randomly calling 5 babies stillborn out of every 1000 babies in the PNOB database, 99.5% of all true livebirths would still be correctly classified as livebirths in the existing PNOB database. Clearly, such a low incidence of stillbirth (statistically rather than clinically speaking) renders estimation of specificity needless.

A baby was a 'confirmed stillbirth' when the mother's medical record verified that the baby was indeed a stillbirth. When the mother's medical record was not found, a baby was a 'confirmed stillbirth' when the caseroom logbook so indicated.

If the baby was not entered in the logbook, or if there was no indication whether it was born alive or still (i.e. missing Apgar scores) and the mother's chart was not found, the baby was excluded from the analysis.

5. RESULTS

5.1 Population of Births

Between January 1, 1989 and December 31, 1992 39,149 babies were born to women residing in the Regional Municipality of Ottawa-Carleton. Of these births, 38,660 were born in an Ottawa-Carleton hospital (Table 3). Four of these births (2 stillbirths from Hospital B; 2 stillbirths from Hospital D) were discovered to be duplicate entries and were, therefore, removed from the analysis. Eight additional stillbirths born to women residing in Ottawa-Carleton which did not match any record in the existing PNOB database, live or stillborn, were discovered upon examination of the logbooks for a total of 38664 births for this analysis.

Table 3
Location of Births by Hospital

Location of Birth	Number of Reported Births	Duplicate Entries	Additional Stillbirths	Adjusted Number of Births[a]
Hosp. A	11493	2	1	11492
Hosp. B	8768	0	3	8771
Hosp. C	8455	0	1	8456
Hosp. D	8109	2	0[b]	8108
Hosp. E	1834	0	3	1837
Home	149	N/A	N/A	N/A
Non-RMOC	298	N/A	N/A	N/A
Missing	43[c]	N/A	N/A	N/A
TOTAL	39149	4	8	38664

a This column adjusts only for missed stillbirths and the duplicate entries that happened to be found. It does not adjust for missed livebirths and the majority of possible duplicate entries.

b,c One stillbirth with a missing value for 'place of birth' was located in the logbook of Hospital 'D'. It is not included in the 'additional stillbirth' column since it was identified as a stillbirth in the database but does increase the 'adjusted number of births' for Hospital 'D' by one.

Descriptive statistics for the eight stillborn babies for whom there was no match in the existing PNOB database are presented in Table 4.

Table 4
Descriptive Statistics
Stillbirths Missing from Existing Database

	Year of Birth	Hospital	Gestation (weeks)	Birthweight (grams)
1	1989	A	33	2040
2	1989	C	39-40	3260
3	1990	B	36	2800
4	1990	E	21	480
5	1990	B	22	234
6	1990	B	32	1640
7	1990	E	21	410
8	1991	E	20-21	325

Four of these babies are extremely premature, barely meeting the Vital Statistics Act criteria for distinguishing a stillborn baby from a miscarriage. At one hospital three of these babies were not listed in the logbook, but rather in a separate notebook for late spontaneous abortions.

5.2 Proportion of Records Located in Logbooks & Medical Records

Nine hundred eighty cases from the PNOB database were selected for the study. Of these 980 cases, 972 (99%) were located in the logbook. Among the eight that were not found, 2 were from Hospital A, 2 from Hospital B, 2 from Hospital C

and 2 from Hospital D. An additional eight stillbirths were identified in the caseroom logbooks, for a total of 980 cases included in the study and abstracted from the caseroom logbooks.

A mother's medical chart was requested if there was a discrepancy between the existing database, the newly entered PNOB database or the logbook information for any one variable. Most charts requested represented records with a discrepancy on a single variable (n=181) while a few had two (n=23) or three (n=5) variables with a discrepancy. In total, there were 242 discrepancies belonging to 209 individual records. Some variables were more likely than others to have a discrepancy. Live/stillborn status had the highest number of discrepancies (n=76) followed by gestational age (n=45), while discrepancies in the sex of the baby was relatively uncommon (n=8; Table 5).

Table 5
Frequency of Discrepancies Between the
PNOB Databases and Caseroom Logbooks

Variable	Number of Discrepancies
Sex of Baby	8
Mother's Age	26
Postal Code	27
Number of Previous Livebirths	28
Birthweight	32
Gestational Age	45
Vital Status	76
TOTAL	**242**

Requests for 487 charts were given to the Medical Records departments of the appropriate hospitals. These 487 charts represented the random sample of 100 with perfect agreement, 209 cases with discrepancies and 168 cases with missing data. One hospital did not wish to retrieve the 1989 medical records charts from storage (n=13), which brought the number of requested charts down from 487 to 474. The effect of not retrieving these 13 charts is likely to be small. Although not specifically examined in this study, if year of birth was in some way related to the accuracy of information on the PNOB form, then it is possible that the effect of not retrieving the 1989 charts could bias the results. Although not tested statistically, year of birth seemed only to be related to stillbirths never being reported on a PNOB form; seven of eight stillbirths missing from the existing PNOB database were from the first two years, 1989 and 1990. Four hundred twenty-eight charts (90%) were eventually located and reviewed (Table 6).

Table 6
Proportion of Medical Records Found

Hospital	Number Requested	Number Found	Success Rate
A	138	131	95%
B	111	92	83%
C	93	87	94%
D	98*	87	89%
E	34	31	91%
TOTAL	474	428	90%

* One hundred eleven charts were originally requested, but the hospital staff did not wish to retrieve the 1989 charts (n=13) from storage.

5.3 *Accuracy of the Investigator's Data Entry*

Double-data entry was performed for a sample of the data entered by the investigator from the new PNOB database, the logbook and medical records. The number of errors made in any of the study variables and the reliability by variable are shown in Table 7.

Table 7
Reliability of Data Entry by Variable

Data Source	Number of Errors	Number of Variables Double-Entered	Reliability by Variable
New Database	11	1045	98.9%
Logbook	7	1067	99.2%
Medical Records	4	468	99.1%

Errors made in the process of data entry could be divided into two major categories: those which were incorrectly copied from the original source to paper (abstraction errors), and those which were copied correctly but were incorrectly keyed into the computer (typographic errors). During some of the data entry (especially from the PNOB forms) the data was abstracted from the source and immediately keyed into the computer, without any paper records. Errors occurring when there were no paper records could not be classified as abstraction or typographic errors and, therefore, are called unclassifiable errors. A summary of the types of errors that occurred for each of the PNOB form, logbook and medical records data entry is shown in Table 8.

Table 8
Types of Errors Identified in Double-Entry

Data Source	Typographic	Abstraction	Unclassifiable	Total
New Database	--	--	11	11
Logbook	4	2	1	7
Medical Records	1	3	--	4

5.4 Descriptive Statistics

Descriptive statistics including mean values or frequency probabilities, and 95% confidence intervals were calculated for all variables from the three complete data sources (i.e. existing and new PNOB databases, and logbook). Nine hundred eighty records for this study were examined in total.

5.4.1 Sample A: Random Sample of Livebirths

Point-estimates of birthweight across the three complete data sources (existing and new PNOB databases, and logbook) were similar, ranging from 3356g in the newly-entered database to 3376g in the logbook. The mean birthweight has been calculated twice for the logbook, first including all babies and again excluding a clear outlier (a 7885g preterm female infant). This outlier was one of the random sample of records entered twice, both times as 7885g. The confidence intervals of the mean values were much larger than the differences between the point estimates, ranging from 3297 to 3438g and the differences were not statistically significant (Friedman's test; p=0.932. Differences among gestational ages were

small and not significant (p=0.888) with means ranging from 38.90 to 38.95. Confidence intervals yielded a range from 38.68 to 39.17. Actual values ranged from 22 to 44 weeks of gestation. Differences among mother's ages were also small and not significant (p=0.970). The means ranged from 28.68 to 28.75, confidence intervals ranged from 28.23 to 29.20 and actual ages ranged from 16 to 44 years of age (Table 9).

Table 9
Descriptive Statistics for Continuous Data
Sample A: Livebirths

Variable	Source	Mean	LL 95% CI	UL 95% CI	n
Birthweight (grams)	Existing Database	3363	3306	3421	446
	New Database	3356	3297	3416	443
	Logbook	3376[a]	3314	3438	443
		3366[b]	3307	3425	442
Gestation (completed weeks)	Existing Database	38.90	38.68	39.12	449
	New Database	38.90	38.68	39.12	447
	Logbook	38.95	38.73	39.17	442
Mother's Age (years)	Existing Database	28.75	28.30	29.20	443
	New Database	28.73	28.28	29.18	443
	Logbook	28.68	28.23	29.13	444

[a] Descriptive statistics calculated including 7885g preterm baby
[b] Descriptive statistics calculated excluding 7885g preterm baby

Male babies represented slightly less than half of all babies, ranging from .473 in the new database to .477 in the logbook. Confidence intervals ranged from .428 to .522. Half of all women delivering babies were primiparous (.505 to .508; 95% CI .460 - .552). A further third (.315 to .324) had delivered one previous livebirth (95% CI .274 to .366). Fewer still had delivered two or more previous livebirths (Table 10).

Table 10
Descriptive Statistics for Ordinal/Nominal Data
Sample A: Livebirths

Variable and Value	Source	Proportion	LL 95% CI	UL 95% CI	n
Sex (male)	Existing Database	.476	.432	.520	485
	New Database	.473	.428	.518	482
	Logbook	.477	.432	.522	480
# Previous Livebirths = 0	Existing Database	.508	.464	.552	486
	New Database	.506	.461	.551	482
	Logbook	.505	.460	.550	481
= 1	Existing Database	.315	.274	.356	486
	New Database	.315	.274	.357	482
	Logbook	.324	.282	.366	481
= 2	Existing Database	.126	.097	.155	486
	New Database	.129	.099	.159	482
	Logbook	.121	.092	.150	481
≥ 3	Existing Database	.051	.031	.071	486
	New Database	.049	.030	.068	482
	Logbook	.049	.030	.068	481

5.4.2 Sample B: Presumed Stillbirths

The mean birthweight of presumed stillbirths was on average 1300g lower, ranging from 2036 to 2081 grams (p=.977). Individual birthweights ranged widely from an extremely low value of 50g to a high value of 5237g. Confidence intervals ranged from 1866 to 2206g. The mean gestational age ranged from 32.23 to 32.49 weeks (p=.907). Confidence intervals ranged from 31.35 to 33.36 and individual values ranged from 20 to 42 completed weeks of gestation. The mean age of women delivering stillbirths varied little from 28.34 to 28.43 years (p=.993), and individual ages ranged from 16 to 42 years (Table 11).

Table 11
Descriptive Statistics for Continuous Data
Sample B: Presumed Stillbirths

Variable	Source	Mean	LL 95% CI	UL 95% CI	n
Birthweight (grams)	Existing Database	2081	1912	2249	267
	New Database	2036	1866	2206	266
	Logbook	2074	1904	2244	267
Gestation (completed weeks)	Existing Database	32.23	31.35	33.12	267
	New Database	32.32	31.44	33.21	269
	Logbook	32.49	31.62	33.36	266
Mother's Age (years)	Existing Database	28.42	27.79	29.05	269
	New Database	28.43	27.81	29.06	269
	Logbook	28.34	27.71	28.97	266

Among presumed stillbirths, the proportion of male babies varied from .581 to .597 with confidence intervals from .522 to .656. Half of the women had never delivered a liveborn baby (.502 to .522) with confidence intervals from .442 to .582. About a third had previously delivered one liveborn baby (.307 to .340) with confidence intervals from .252 to .397. About a sixth had previously delivered two or more livebirths (Table 12).

Table 12
Descriptive Statistics for Ordinal/Nominal Data
Sample B: Presumed Stillbirths

Variable and Value	Source	Proportion	LL 95% CI	UL 95% CI	n
Sex (male)	Existing Database	.597	.538	.656	268
	New Database	.586	.527	.645	268
	Logbook	.581	.522	.640	267
# Previous Livebirths = 0	Existing Database	.522	.462	.582	270
	New Database	.511	.451	.571	268
	Logbook	.502	.442	.562	265
= 1	Existing Database	.307	.252	.362	270
	New Database	.317	.261	.373	268
	Logbook	.340	.283	.397	265
= 2	Existing Database	.119	.080	.158	270
	New Database	.123	.084	.162	268
	Logbook	.117	.078	.156	265
≥ 3	Existing Database	.053	.026	.080	270
	New Database	.049	.023	.075	268
	Logbook	.042	.018	.066	265

5.4.3 Sample C: Missing Vital Status

Among babies for whom the existing PNOB database was missing information on vital status, the mean birthweight ranged from 3328 to 3394g (p=.909). Individual birthweights ranged from 840 to 5550g and confidence intervals ranged from 3310 to 3476g. Mean values of gestational age ranged from 38.68 to 39.13 (p=.766) with individuals means ranging from 29 to 42 completed weeks. Confidence intervals ranged from 38.28 to 39.38. The mean mother's age ranged from 28.35 to 28.42 years, with individual values ranging from 16 to 41 years. Confidence intervals ranged from 27.67 to 29.09 (Table 13).

Table 13
Descriptive Statistics for Continuous Data
Sample C: Missing Vital Status

Variable	Source	Mean	LL 95% CI	UL 95% CI	n
Birthweight (grams)	Existing Database	3393	3310	3476	210
	New Database	3394	3311	3477	209
	Logbook	3328	3231	3425	221
Gestation (completed weeks)	Existing Database	39.09	38.83	39.36	209
	New Database	39.13	38.88	39.38	209
	Logbook	38.68	38.28	39.09	220
Mother's Age (years)	Existing Database	28.35	27.67	29.04	215
	New Database	28.35	27.67	29.04	215
	Logbook	28.42	27.76	29.09	217

About half of the babies were male (.467 to .486) with confidence intervals from .395 to .552. Half of the mothers were primiparous (.486 to .491; CI .419 to .557). A further third had had one previous livebirth (.330-.339; CI .267-.402). About 18% had delivered two or more previous livebirths (Table 14).

Table 14
Descriptive Statistics for Ordinal/Nominal Data
Sample C: Missing Vital Status

Variable and Value	Source	Proportion	LL 95% CI	UL 95% CI	n
Sex (male)	Existing Database	.467	.395	.540	182
	New Database	.470	.397	.543	181
	Logbook	.486	.420	.552	220
# Previous Livebirths = 0	Existing Database	.486	.419	.553	212
	New Database	.488	.421	.555	213
	Logbook	.491	.425	.557	218
= 1	Existing Database	.330	.267	.393	212
	New Database	.333	.270	.396	213
	Logbook	.339	.276	.402	218
= 2	Existing Database	.108	.066	.150	212
	New Database	.099	.059	.139	213
	Logbook	.101	.061	.141	218
≥ 3	Existing Database	.076	.040	.112	212
	New Database	.080	.044	.116	213
	Logbook	.070	.036	.104	218

Male babies accounted for between 50.9% and 51.9% of all stillbirths. The number of previous livebirths for women delivering stillbirths ranged from zero to eight among all three data sources. Half of all women had had no previous liveborn baby (51.2 - 52.6%). A further third (29.4 - 31.0%) had had one previous livebirth.

Between 12.8% and 13.1% had delivered two previous livebirths prior to the delivery of the stillbirth under study. Women who had had three or more previous livebirths accounted for between 4.7% and 5.2% of all women having a stillbirth in this sample.

5.4.4 Sample D: False Negatives

Among babies for whom the existing PNOB database was missing information on vital status, the mean birthweight ranged from 2820 to 3033g (p=.909). Individual birthweights ranged from 340 to 4590g and confidence intervals ranged from 2438 to 3346g. Mean values of gestational age ranged from 36.32 to 36.54 (p=.766) with individuals means ranging from 20 to 41 completed weeks. Confidence intervals ranged from 34.35 to 38.47. The mean mother's age ranged from 28.89 to 28.96 years, with individual values ranging from 19 to 40 years. Confidence intervals ranged from 27.04 to 30.79 (Table 15).

Table 15
Descriptive Statistics for Continuous Data
Sample D: False Negatives

Variable	Source	Mean	LL 95% CI	UL 95% CI	n
Birthweight (grams)	Existing Database	3033	2721	3346	35
	New Database	2820	2438	3202	34
	Logbook	2865	2500	3230	36
Gestation (completed weeks)	Existing Database	36.54	34.61	38.47	35
	New Database	36.32	34.35	38.29	34
	Logbook	36.50	34.63	38.37	36
Mother's Age (years)	Existing Database	28.96	27.13	30.79	36
	New Database	28.96	27.11	30.77	36
	Logbook	28.89	27.04	30.74	36

About half of the babies were male (.344 to .400) with confidence intervals from .189 to .560. Half of the mothers were primiparous (.528; CI .365 to .691). A further third had had one previous livebirth (.222; CI .086-.358). About 25% had delivered two or more previous livebirths (Table 16).

Table 16
Descriptive Statistics for Ordinal/Nominal Data
Sample D: False Negatives

Variable and Value	Source	Proportion	LL 95% CI	UL 95% CI	n
Sex (male)	Existing Database	.400	.240	.560	35
	New Database	.344	.189	.499	32
	Logbook	.400	.240	.560	35
# Previous Livebirths = 0	Existing Database	.528	.365	.691	36
	New Database	.528	.365	.691	36
	Logbook	.528	.365	.691	36
= 1	Existing Database	.222	.086	.358	36
	New Database	.222	.086	.358	36
	Logbook	.222	.086	.358	36
= 2	Existing Database	.222	.086	.358	36
	New Database	.222	.086	.358	36
	Logbook	.222	.086	.358	36
≥ 3	Existing Database	.028	0	.082	36
	New Database	.028	0	.082	36
	Logbook	.028	0	.082	36

5.5 *Objective #1:*

To identify how PNOB records are completed in each maternity care hospital in the Regional Municipality of Ottawa-Carleton

To meet objective #1, face-to-face interviews were carried out with the Labour & Delivery Nurse Manager or, in one case, a senior nurse at each of the five hospitals. All nurse managers attested to the timely completion of the PNOB forms within at least one hour of birth. Four managers reported that the PNOB forms were completed right in the caseroom. One of these four hospitals reported that the

babies were weighed in the nursery rather than in the caseroom and that the birthweight was phoned back to the nurses' station and then documented. Only one hospital reported that the PNOBs were completed at the nurses' station following birth.

Two hospitals reported that the PNOB forms were completed from information in the mothers' medical charts while two reported that the PNOB forms were completed in the caseroom independent of the mothers' medical charts. One hospital stated that the PNOB form was completed from the caseroom nurses' notes.

Although all five hospitals reported that the PNOBs were checked for missing or duplicate entries, only two had systems in place for checking the accuracy of information contained on the PNOB form. In both cases, the ward clerk checked the accuracy of PNOB forms against either the logbook or the medical chart. At one hospital, incomplete forms were returned to the caseroom for completion prior to their delivery to the office of the Registrar.

5.6 OBJECTIVE #2:

To assess the reliability of extracting information from the PNOB records to create a computer database

Reliability of the acts of deciphering and entering data onto the health department's computer from the PNOB forms is assessed quantitatively using the intraclass correlation coefficient for the pairwise comparison of the existing and

newly-entered databases and qualitatively using scatterplots of differences in values against mean values (A-B vs. (A+B)/2). The number of pairs upon which the ICCs were calculated varies slightly among the different variables, reflecting the number of babies for whom a value for both the existing and new databases was present.

5.6.1 Sample A: Random Sample of Livebirths

Among livebirths, the weighted mean ICC for all variables was extremely high (.9672). ICCs for the variables gestational age (0.9985), mother's age (0.9981) and birthweight (0.9803) were extremely high while number of previous livebirths was slightly lower (0.8922; Table 17).

Table 17
Comparison of Existing & New Databases
Sample A: Livebirths

Variable	n	ICC	LL 95% CI	UL 95% CI
Gestational Age	447	.9985	.9978	.9989
Mother's Age	443	.9981	.9973	.9987
Birthweight	443	.9803	.9720	.9862
# Previous Livebirths	446	.8922	.8493	.9234
Weighted Mean ICC		.9672		

The difference versus mean scatterplots for gestational age (Figure 8a), mother's age (Figure 8b) and birthweight (Figure 8c) visually demonstrate the excellent agreement between data sources and show little trend (slope of the regression lines close to 0) and little systematic bias (mean value of the X-Y differences =.01, .02, and 7.17 respectively). Random error was also minimal with only a few data points outside the boundaries of plus and minus two standard deviations. The standard deviations were very small and of little clinical importance. The plot for number of previous livebirths (Figure 8d) showed a very slight trend towards either overestimation of parity by the existing database or underestimation by the newly-entered database but virtually no systematic bias (mean X-Y difference = .02).

Kappa for agreement of sex of baby was perfect (k=1.0). Simple agreement for postal code is more difficult to describe (see first row of Table 18). Although postal code is on the PNOB form, it wasn't until the year 1990 that postal codes were entered in the existing database. As a result all births from 1989 are excluded from pairwise analyses involving the existing database but not from the newly-entered database, explaining the smaller sample size for those two comparisons involving the existing PNOB database. In one percent of records, there was an error among the first three digits. Errors in the second half were more common in 2.2% of cases. No records contained discrepancies in both the first and second halves of the postal code.

Figure # 8a

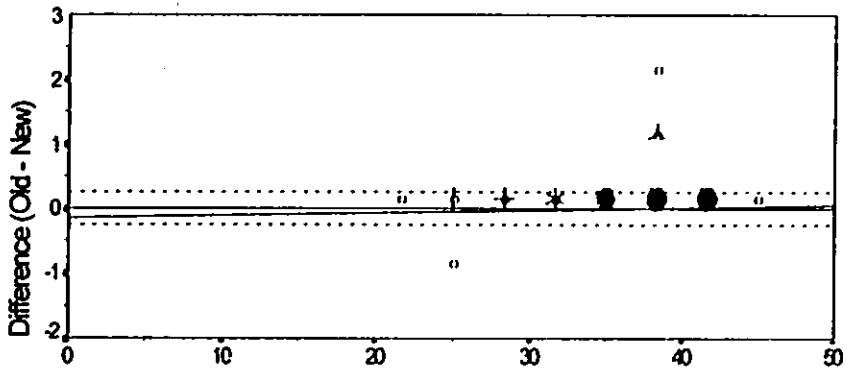


Figure # 8c

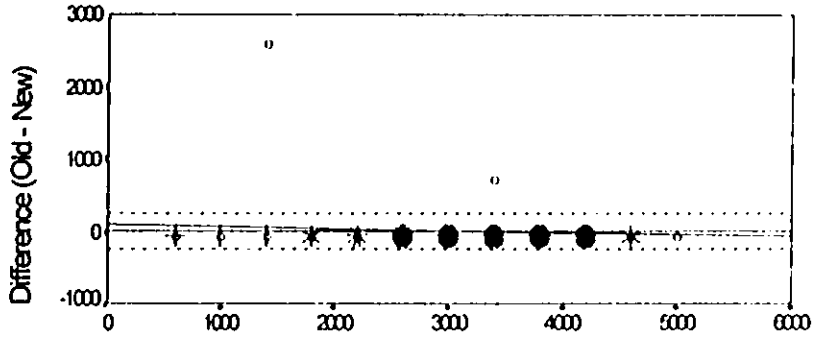


Figure #8b

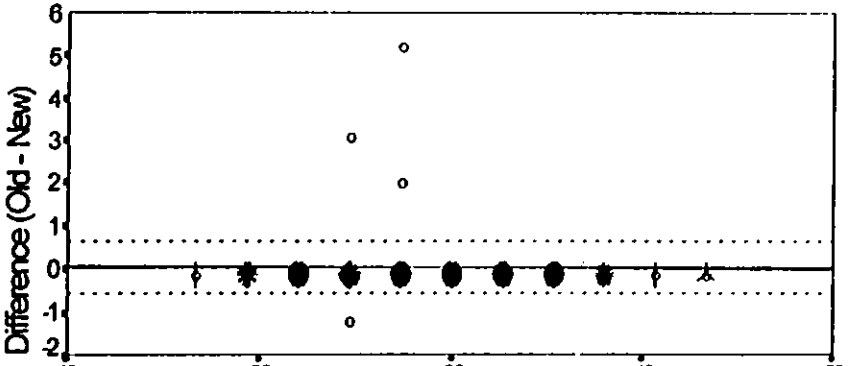


Figure # 8d

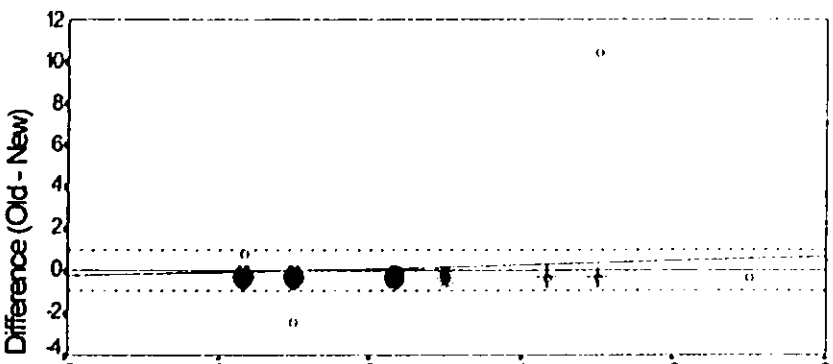


Table 18
Comparison of Postal Codes
Existing Versus New Databases

Sample	N[a]	Wrong First 3 Digits	Wrong Last 3 Digits	Wrong First & Last 3 Digits	Missing
Random Sample of Livebirths	448	5 (1.1%)	10 (2.2%)	0	8 (1.8%)
Presumed Stillbirths	212	5 (2.4%)	11 (5.2%)	0	0
Missing Vital Status	134	2 (1.5%)	3 (2.2%)	2 (1.5%)	0
False Negatives	29	0	2 (6.9%)	0	0

[a] Excluding 1989 births for which postal code was not entered into the existing database

5.6.2 Sample B: Presumed Stillbirths

Among presumed stillbirths, there was excellent agreement for all variables with a weighted mean ICC of .9863 and no point estimates of ICC below 0.96 (Table 19). Virtually perfect agreement was found for gestational age (ICC=.995), mother's age (ICC=.999) and number of previous livebirths (ICC=.99). Agreement for birthweight was slightly lower but still outstanding at .96. The lower limits of the 95% confidence intervals all exceeded .94.

Table 19
Comparison of Existing & New Databases
Sample B: Presumed Stillbirths

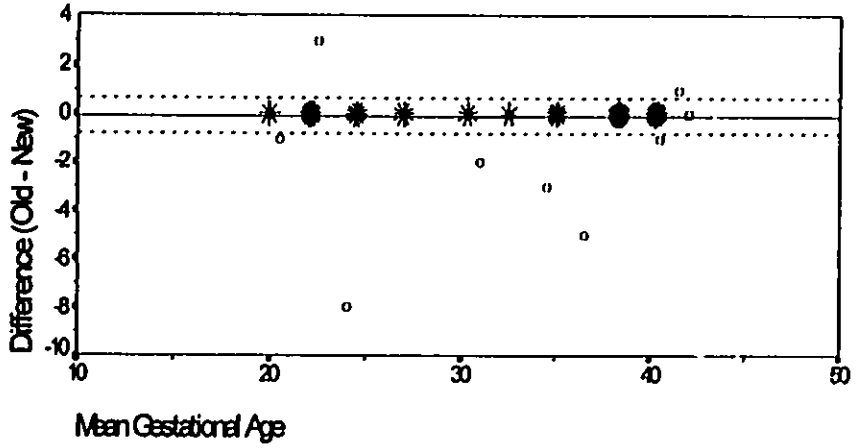
Variable	n	ICC	LL 95% CI	UL 95% CI
Gestational Age	268	.9952	.9932	.9967
Mother's Age	269	.9993	.9990	.9995
Birthweight	265	.9615	.9454	.9729
# Previous Livebirths	268	.9889	.9841	.9922
Weighted Mean ICC		.9863		

The mean versus difference scatterplots (Figures 9 a-d) reflected this excellent agreement with little or no sign of trend of systematic bias. There are several grossly discrepant values for birthweight lying on a straight line and for which an extra zero has been added to the correct value. There were also three pairs where the magnitude of the discrepancy was 1000, indicating that an incorrect first digit was entered for one of the data sources. The raw-data scatterplot of birthweight appears in Figure 10.

Figure # 9a

Errors Associated with Gestational Age

Presumed Stillbirths

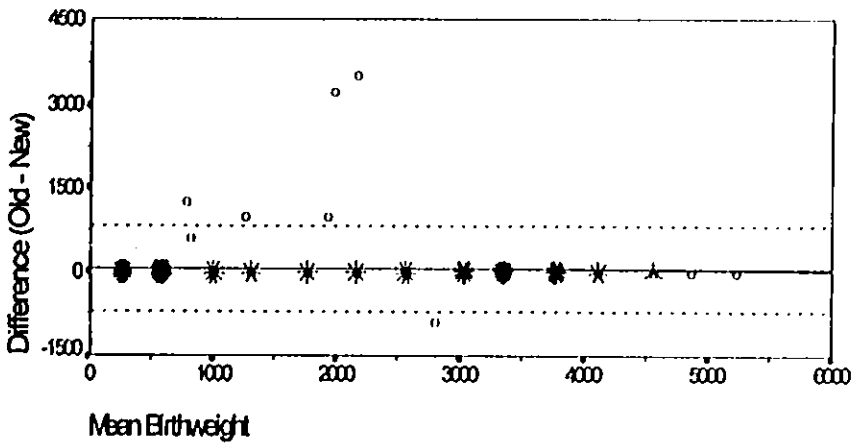


Mean X - Y Difference: -0.08

Figure # 9c

Errors Associated with Birthweight

Presumed Stillbirths

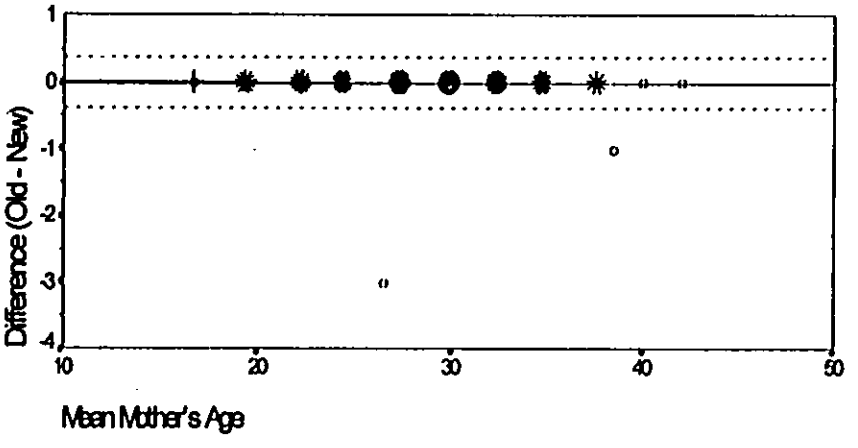


Mean X - Y Difference: 49.74

Figure # 9b

Errors Associated with Mother's Age

Presumed Stillbirths

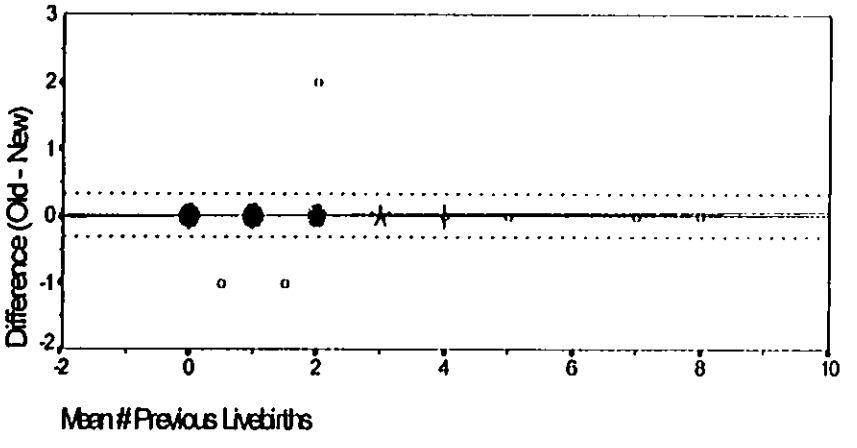


Mean X - Y Difference: -0.01

Figure # 9d

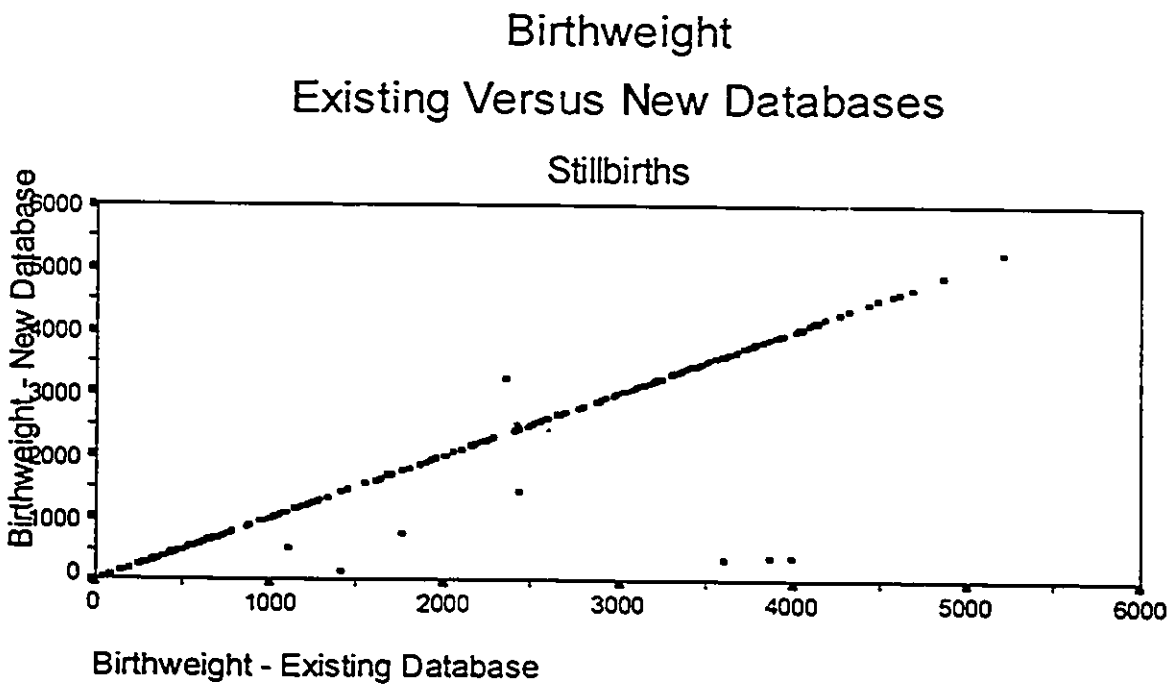
Errors Associated with # Previous Livebirths

Presumed Stillbirths



Mean X - Y Differen 0

Figure 10



Kappa for sex of baby was also high at 0.9690. For postal code (see row #2 of Table 18) 2.4% of records contained errors in the first half of the postal code. Discrepancies among the last three digits of the postal code were present in 5.2% of all records and none had errors in both the first and last half of postal code.

5.6.3 Sample C: Missing Vital Status

Among babies with missing information regarding whether they were born live or still (Table 20) the weighted mean ICC was extremely high at .9854 and point estimates of agreement for all variables exceeded 0.94, including perfect agreement for mother's age (ICC=1.0). The lower limits of the 95% confidence intervals all exceeded 0.92.

Table 20
Comparison of Existing & New Databases
Sample C: Missing Vital Status

Variable	n	ICC	LL 95% CI	UL 95% CI
Gestational Age	209	.9454	.9228	.9615
Mother's Age	215	1.000	1.000	1.000
Birthweight	210	.9998	.9997	.9999
# Previous Livebirths	212	.9956	.9938	.9969
Weighted Mean ICC		.9854		

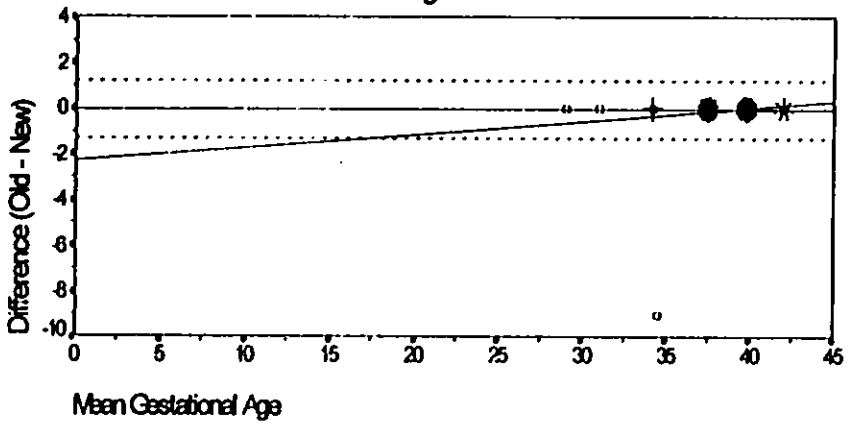
Scatterplots (Figures 11a-d) gave a visual picture of the agreement. The scatterplot for gestational age (Figure 11a) showed some trend toward either underestimation of gestational age by the existing PNOB database or overestimation of gestational age by the newly-entered PNOB database among less mature babies. There was one clearly discrepant value lying outside the boundaries of ± 2 standard deviations. All other plots (Figures 11 b-d) reflected the virtually perfect agreement associated with ICCs in the .995 - 1.00 range.

Agreement for the variable 'sex of baby' was high at k=.966 for 177 data pairs. There were discrepancies in the first half of the postal code in 1.5% of cases while 2.2% had discrepancies in the last half of the postal code. A further 1.5% had errors in both the first and last halves of the postal code (Table 18, row #3).

Figure #11a

Errors Associated with Gestational Age

Missing Vital Status

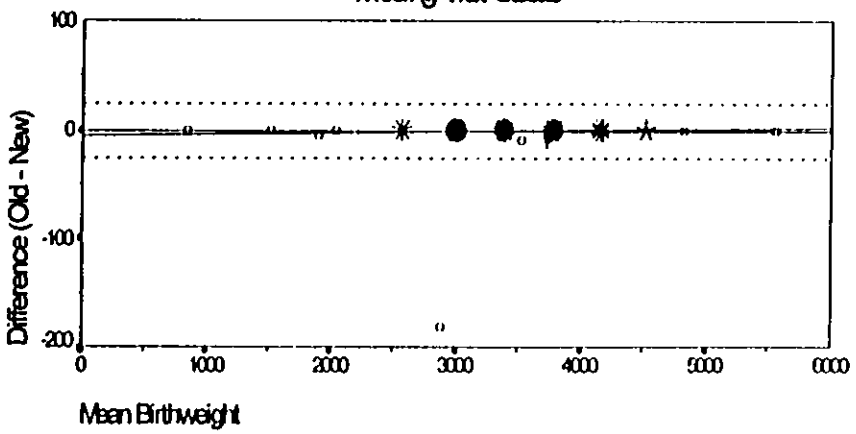


Mean X - Y Difference: -0.04

Figure #11c

Errors Associated with Birthweight

Missing Vital Status

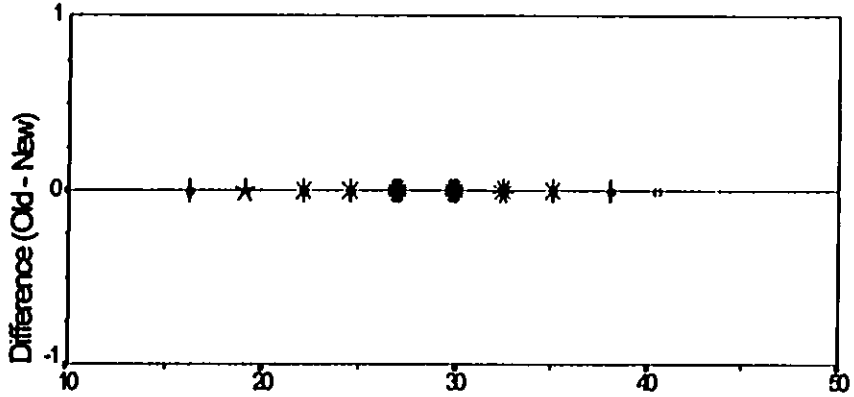


Mean X - Y Difference: -1.02

Figure #11b

Errors Associated with Mother's Age

Missing Vital Status

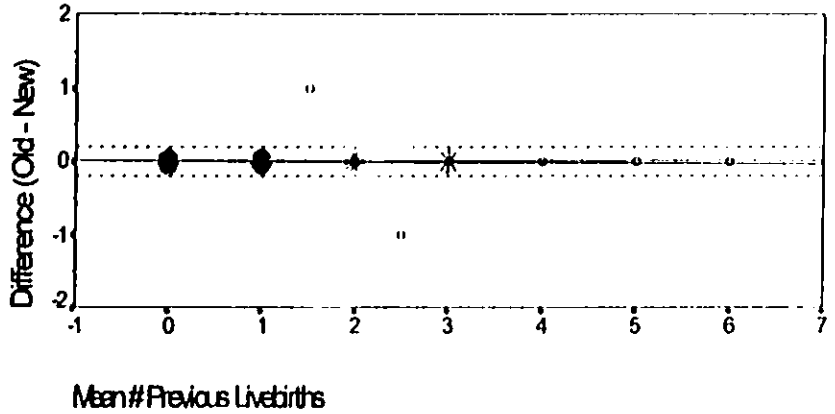


Mean X - Y Difference: 0

Figure # 11d

Errors Associated with # Previous Livebirths

Missing Vital Status



Mean X - Y Difference: 0

5.6.4 Sample D: False Negatives

Among babies indicated in the existing PNOB database as liveborn but later found to be stillborn according to the logbooks, there was virtually perfect agreement for gestational age (ICC=.9977), mother's age (ICC=1.0) and number of previous livebirths (ICC=1.0; Table 21). Agreement for birthweight was much lower with ICC=0.70. Upon examination of the raw-data scatterplot (Figure 12), all data pairs are in perfect agreement except two. There are two babies for whom birthweight was over-estimated by a factor of 10, indicating incorrect decimal point placement.

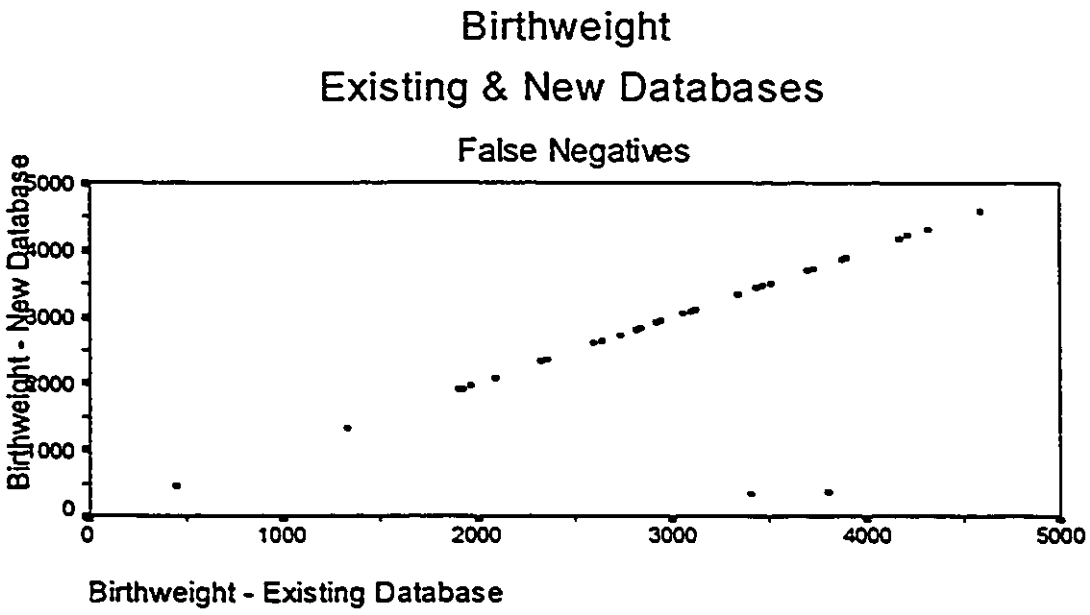
Table 21
Comparison of Existing & New Databases
Sample D: False Negatives

Variable	n	ICC	LL 95% CI	UL 95% CI
Gestational Age	34	.9977	.9955	.9988
Mother's Age	36	1.000	1.000	1.000
Birthweight	34	.6985	.4838	.8343
# Previous Livebirths	36	1.000	1.000	1.000
Weighted Mean ICC		.9262		

The mean versus difference scatterplots (Figures 13 a,b,d) reflect the excellent agreement obtained for gestational age, mother's age and number of previous livebirths. The two discrepant pairs for birthweight appear again in Figure 13c and in the raw-data scatterplot (Figure 12).

Kappa for sex of baby was 1.00, indicating perfect agreement. There were no errors in the first half of postal code and only 2 errors in the last half (Table 18, row 4)

Figure 12



5.6.5 Liveborn/Stillborn Status

For the variable live/stillborn, a 3x3 table is presented (Table 22). Ignoring the missing values, simple agreement is 0.9544 and kappa is 0.9025. Most of the discrepant pairs lie in the 'c' cell of a standard 2 x 2 table, indicating that 27 babies called 'stillborn' in the newly-entered database were entered as having been born live in the health department's existing database. Seven liveborn babies (according to the new database) were entered as stillbirths in the existing PNOB database.

Figure # 13a

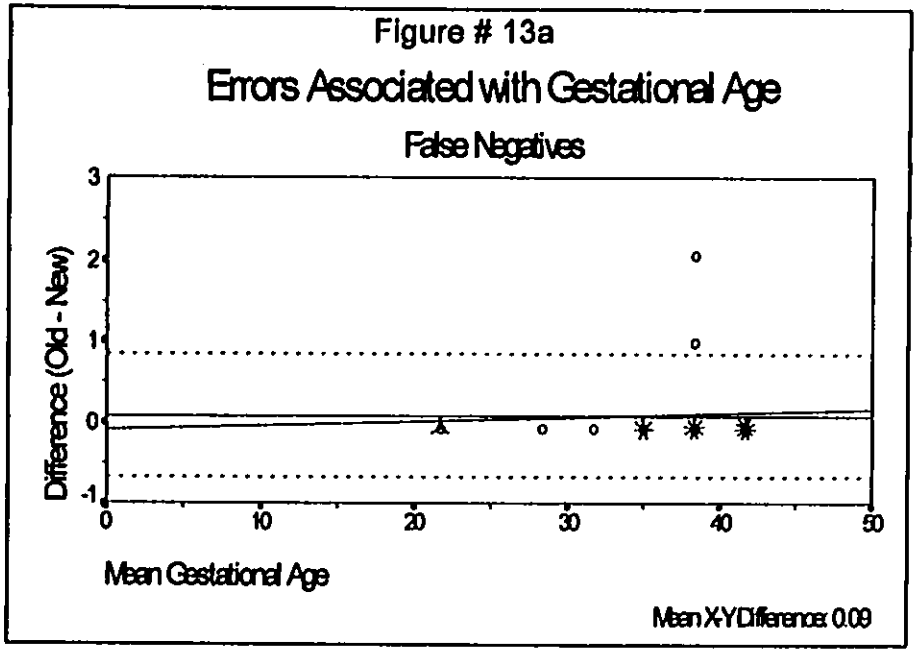


Figure # 13c

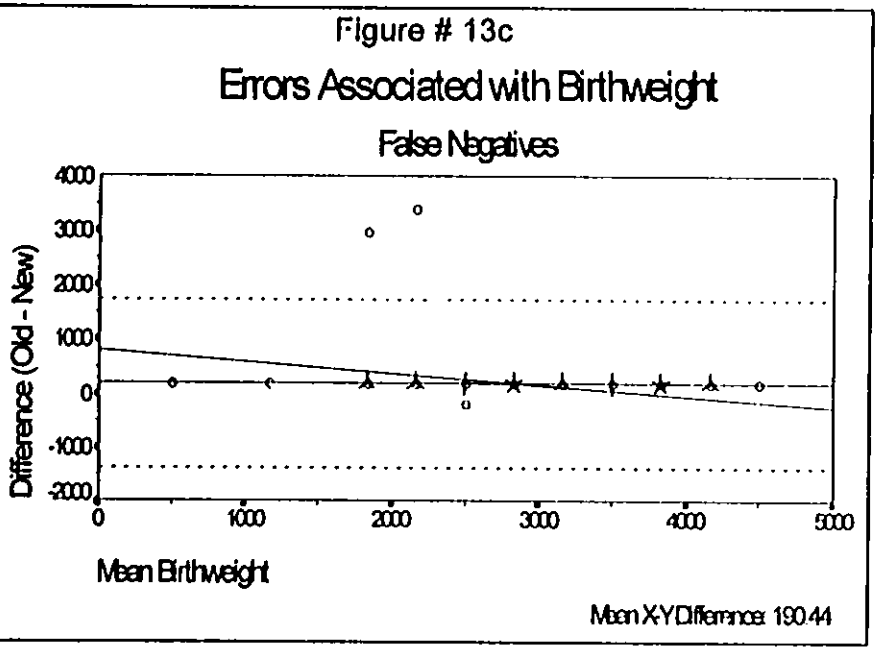


Figure #13b

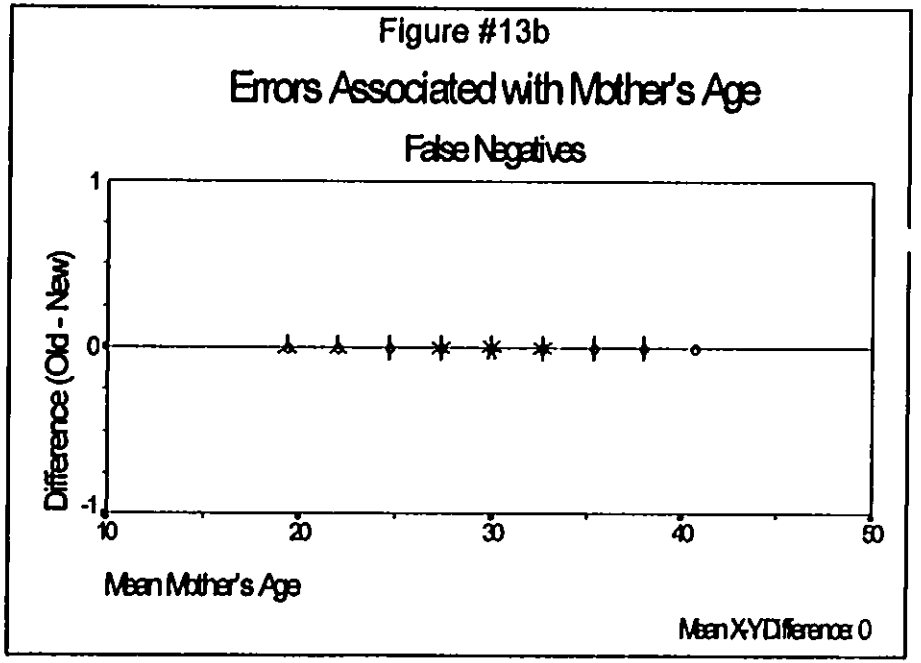
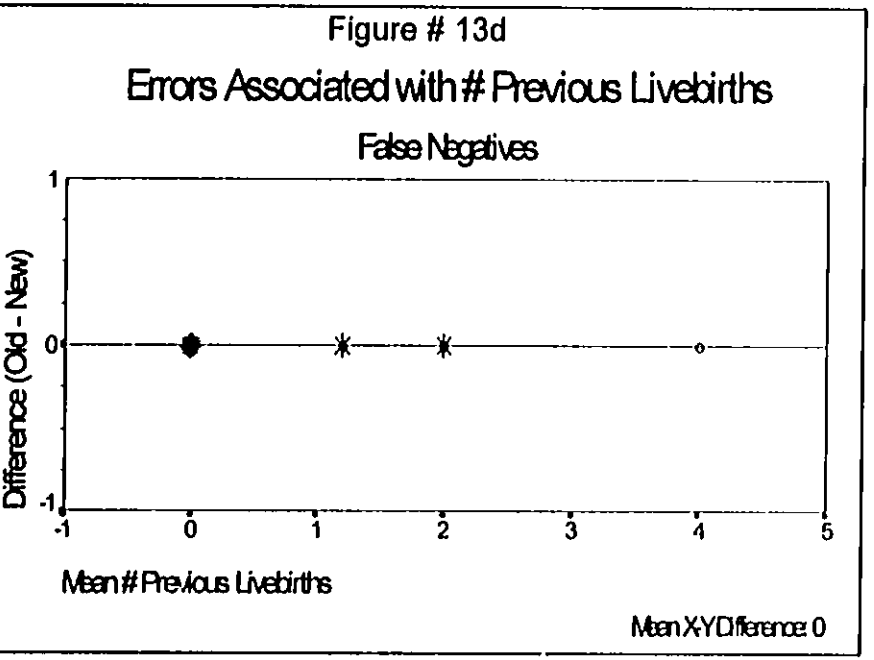


Figure # 13d



Even though kappa is quite high, a discrepancy among 34 babies on such a crucial variable is clearly not acceptable.

Table 22
Comparison of Existing and New Databases
Liveborn/Stillborn Status

Source		Newly-Entered Database			Total
		Stillborn	Liveborn	Missing	
Existing PNOB Database	Stillborn	261	7	2	270
	Liveborn	27	450	9	486
	Missing	0	2	222	224
Total		288	459	233	980

5.7 Objective #3:
Validity of Data in PNOB Databases

5.7.1 Sample A: Random Sample of Livebirths Compared to Logbook

5.7.1.1 Newly entered PNOB Database

Overall weighted agreement between variables in the newly-entered database and the logbook for livebirths (Table 23) were high at 0.945. Agreement for gestational age and mother's age (ICC=.98) was extremely high. Agreement with the logbook was also quite high for number of previous livebirths (ICC=0.93) and a bit lower for birthweight (ICC=0.89).

Table 23
Comparison of Newly Entered Database & Logbook
Sample A: Livebirths

Variable	n	ICC	LL 95 % CI	UL 95% CI
Gestational Age	439	.9819	.9743	.9873
Mother's Age	437	.9770	.9672	.9838
Birthweight	436	.8935	.8511	.9243
# Previous Livebirths	441	.9257	.8955	.9475
Weighted Mean ICC		.9446		

Among the difference versus mean scatterplots, some random variation without systematic bias is seen between the new database and the logbook for gestational age (Figure 14a), with several values lying outside the mean plus and minus two standard deviation boundary. Nonetheless, there does not appear to be any appreciable degree of systematic bias or trend. In Figure 14b, a clearly discordant value is seen, with the new PNOB database value for age being about 20 years higher than the logbook value. In figure 14c for the variable birthweight, an outlier with a discrepancy of 5100g is clearly seen in the lower right corner of the scatterplot. The remaining scatterplot of number of previous livebirths for the new database and logbook (figure 14d) is unremarkable.

Figure #14a

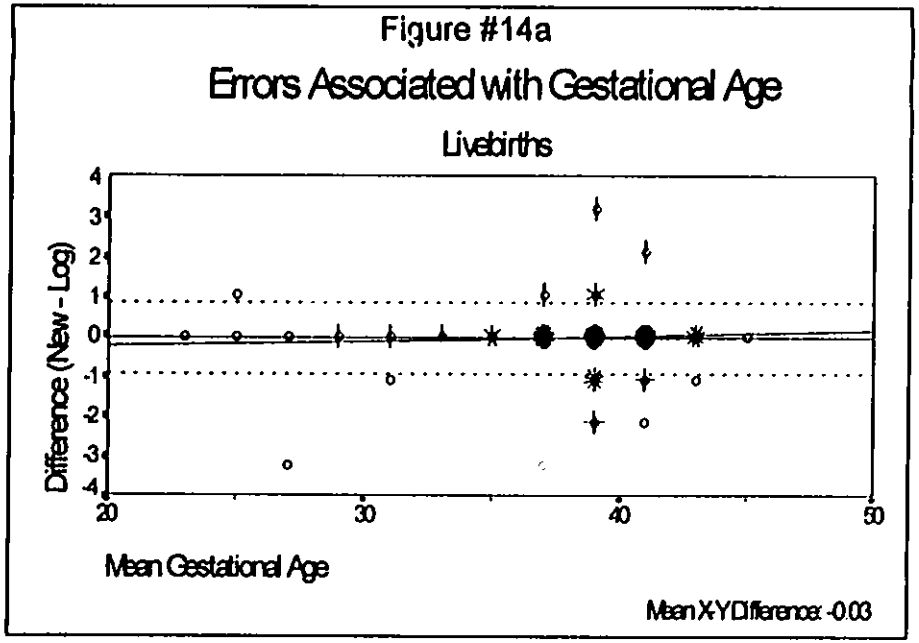


Figure #14c

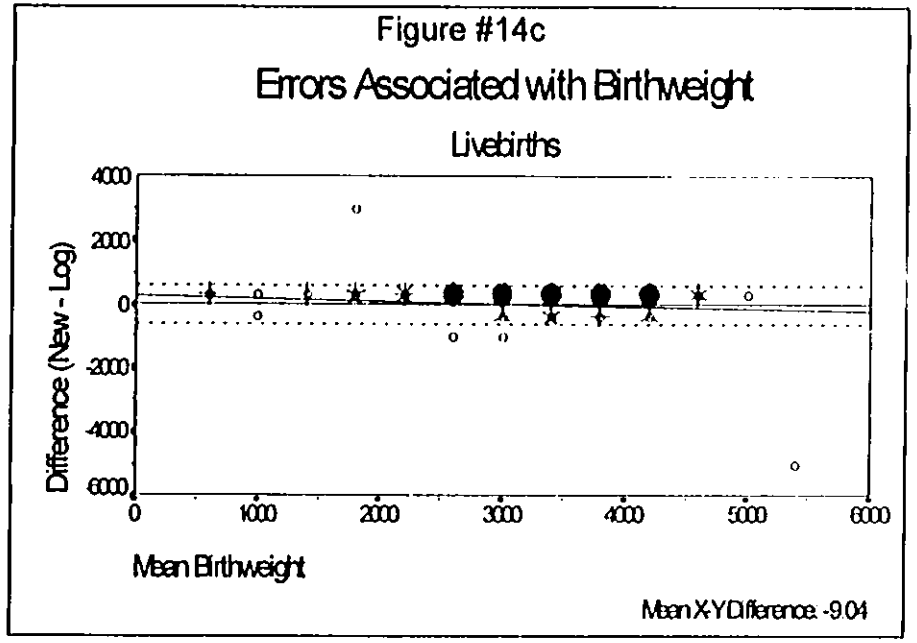


Figure # 14b

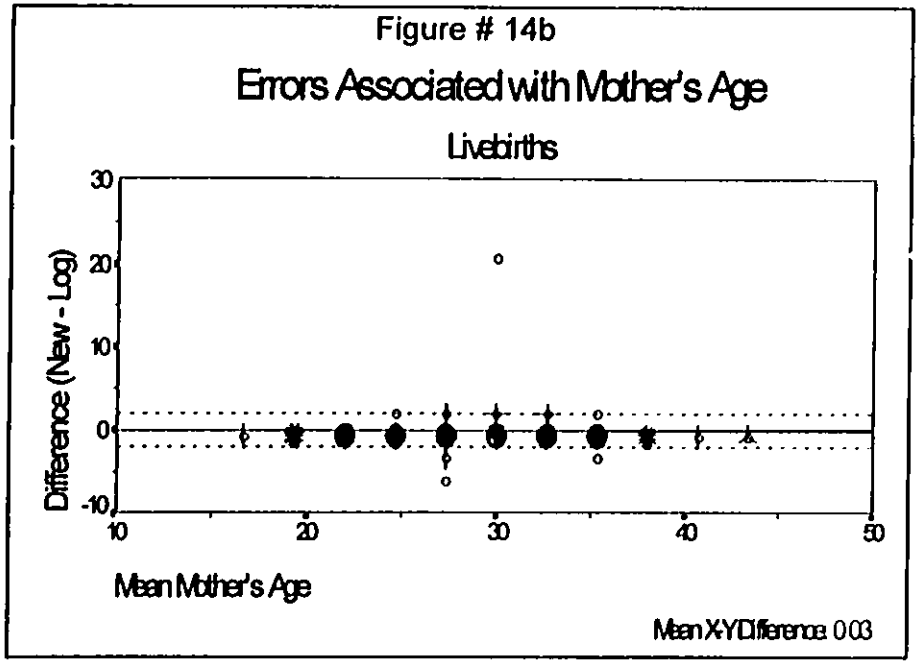
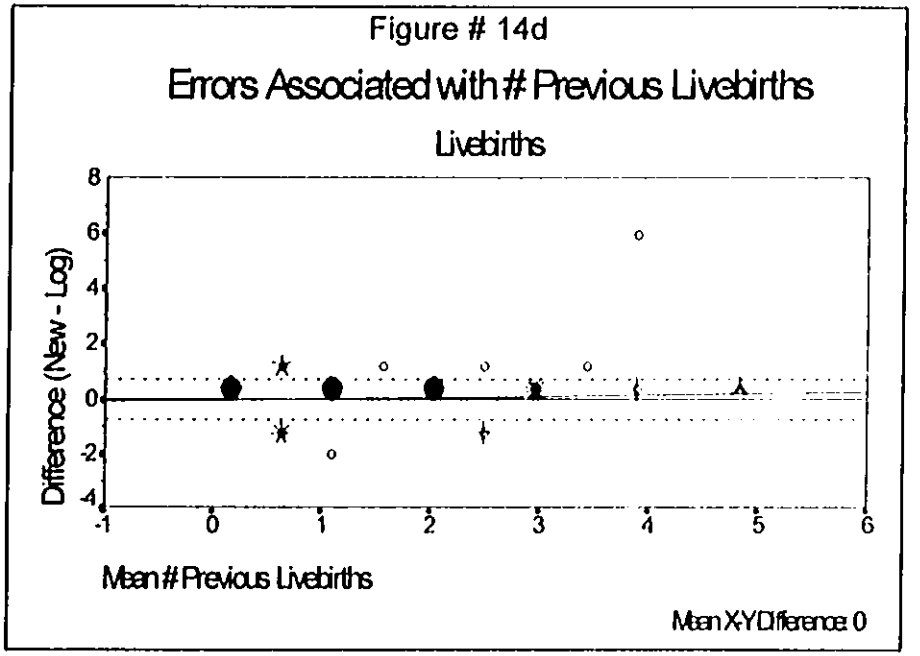


Figure # 14d



Agreement for sex of the baby was extremely high with kappa equal to .973 based on 450 data pairs. Agreement for postal code is shown below in (Table 24, row #1). For comparisons between the newly-entered database and the logbook, errors occurring in the first three digits of the postal code were present in 3.3% of cases. Errors in the last three digits were more common, occurring in 5.8% of cases. Only 6 (1.1%) had errors in both the first and last halves of the postal code. Occasionally in the hospital caseroom logbooks only the first three characters of the postal code were listed rather than the full postal code. In these cases, if the first three characters were in agreement with the other data source of the pair, that pair was included in Column #6; if the first three characters were not the same, it was included in Column #3. Among livebirths, postal codes were missing the last three digits in 11.1% of all cases and in a further 3.6% postal code was missing entirely.

Table 24
Comparison of Postal Codes
Sample A: Livebirths

Sources	N	Wrong First 3 Digits	Wrong Last 3 Digits	Wrong First & Last 3 Digits	Missing Last 3 Digits	Missing
New Database vs. Logbook	450	15 (3.3%)	26 (5.8%)	5 (1.1%)	49 (11.1%)	16 (3.6%)
Old Database vs. Logbook	443[a]	18 (4.1%)	27 (6.1%)	5 (1.1%)	49 (11.1%)	12 (2.7%)

[a] Excluding 1989 births for which postal code was not entered into the existing database

5.7.1.2 Existing PNOB Database

Overall agreement involving the existing PNOB database and the logbook (Table 25) was lower than agreement between the new PNOB database and the logbook with a weighted mean ICC of .916. Agreement for gestational age and mother's age was extremely high (ICC > 0.97) and was consistent with agreement obtained for the newly-entered database. Agreement for birthweight (ICC=0.89) and number of previous livebirths was lower (ICC=0.81). The lower limits of the 95% confidence intervals dipped to .85 and .74 for the latter two variables.

Table 25
Comparison of Existing Database & Logbook
Sample A: Livebirths

Variable	n	ICC	LL 95% CI	UL 95% CI
Gestational Age	441	.9833	.9761	.9883
Mother's Age	437	.9775	.9679	.9842
Birthweight	439	.8944	.8524	.9250
# Previous Livebirths	445	.8116	.7411	.8644
Weighted Mean ICC		.9163		

The difference versus mean scatterplots for gestational age (Figure 15a), mother's age (Figure 15b) and birthweight (Figure 15c) show some random error but little or no systematic bias or trend. There is one highly discrepant pair for mother's age with a difference of 20 years (Figure 15b) which was seen previously for the comparison with the newly-entered PNOB database. The scatterplot for

Figure # 15a

Errors Associated with Gestational Age

Livebirths

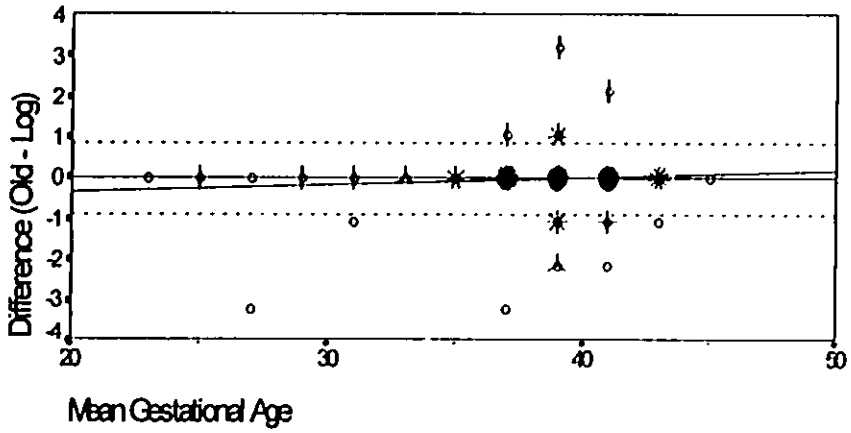


Figure # 15c

Errors Associated with Birthweight

Livebirths

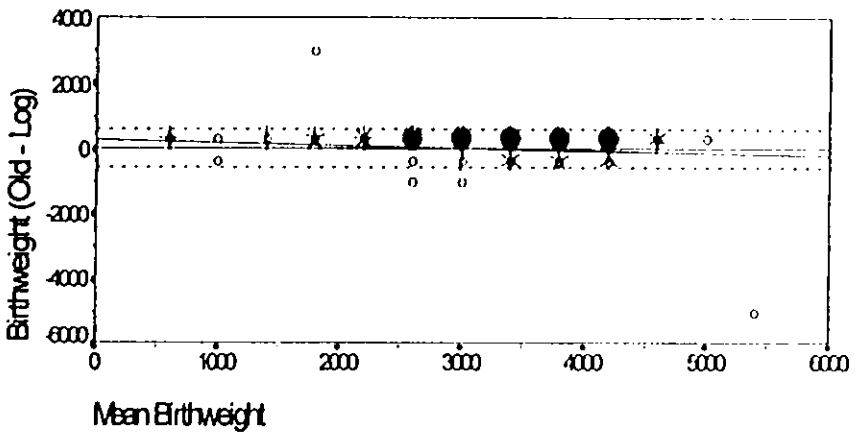


Figure # 15b

Errors Associated with Mother's Age

Livebirths

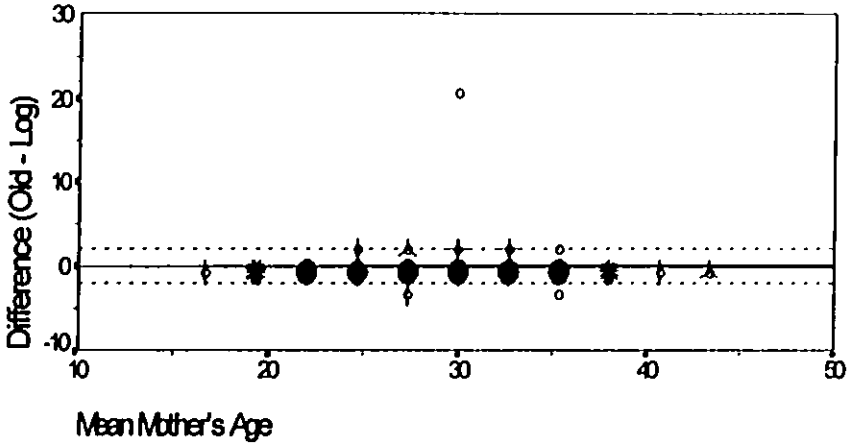
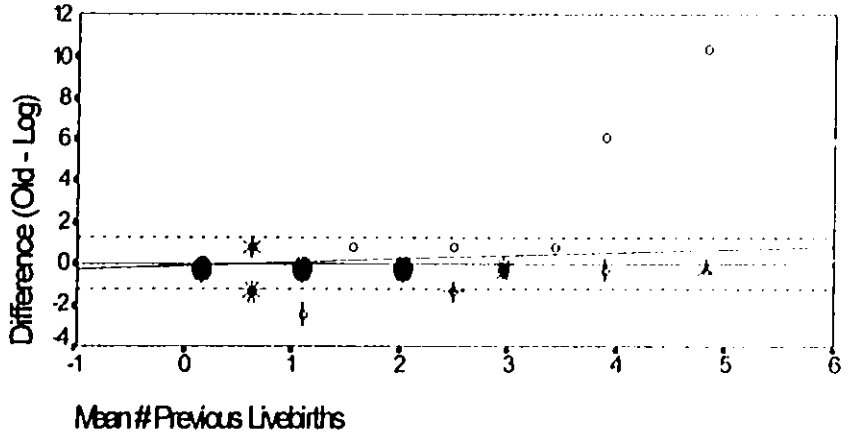


Figure #15d

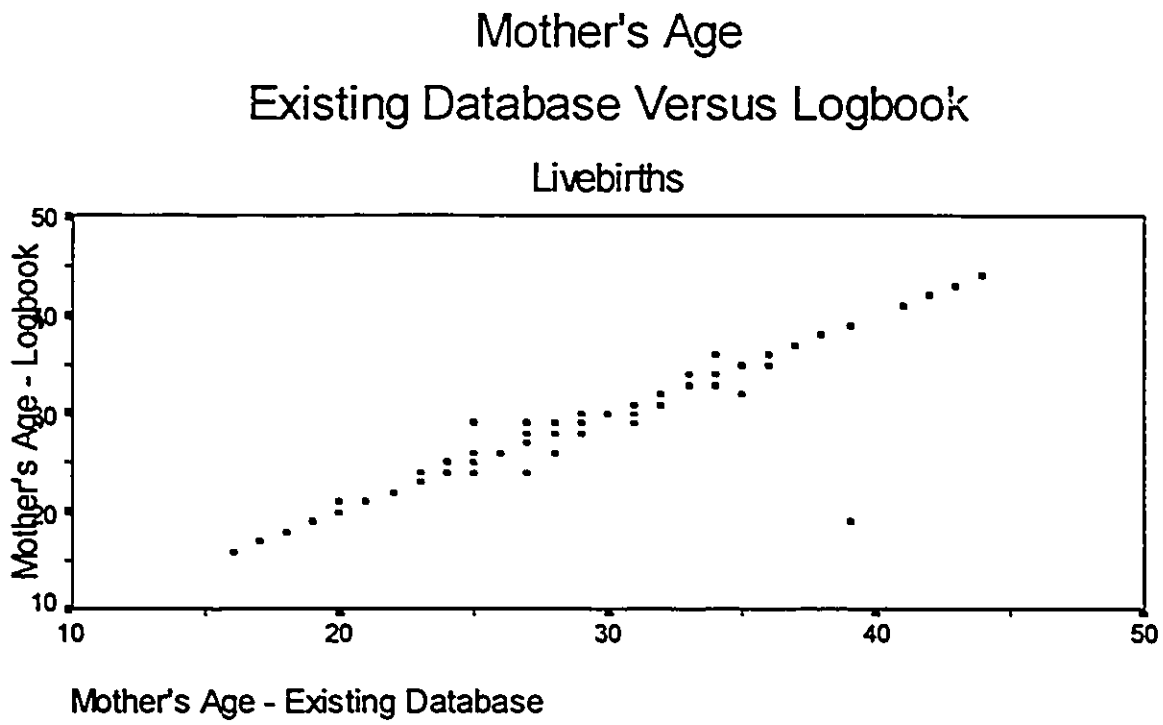
Errors Associated with # Previous Livebirths

Livebirths



number of previous livebirths (Figure 15d) shows a trend toward increasing differences with increasing parity. The discrepant pair for mother's age also appears in the raw-data scatterplot in Figure 16.

Figure 16



Agreement for the variable 'sex' was extremely high (k=.973), which is identical to the kappa obtained for the new database and the logbook. Errors were present in the first three digits of postal code in 4% of cases and 6.1% had errors in the last half of the postal code (Table 24, row #2). A further 1% had errors in both the first and last half of postal code. The last three digits of the postal code were missing in 10.6% of cases and a further 3.3% were missing the entire postal code.

5.7.2 Sample B: Presumed Stillbirths Compared to the Logbook

5.7.2.1 Newly-Entered PNOB Database

The analysis of presumed stillbirths includes data from 270 cases in which the existing database indicated that a baby was stillborn. The weighted mean ICC for all variables was .9791 and all point estimates of ICC were above .94 (Table 26). Agreement for gestational age, mother's age and birthweight all exceeded .99 while agreement for number of previous livebirths was .94.

Table 26
Comparison of Newly Entered Database & Logbook
Sample B: Presumed Stillbirths

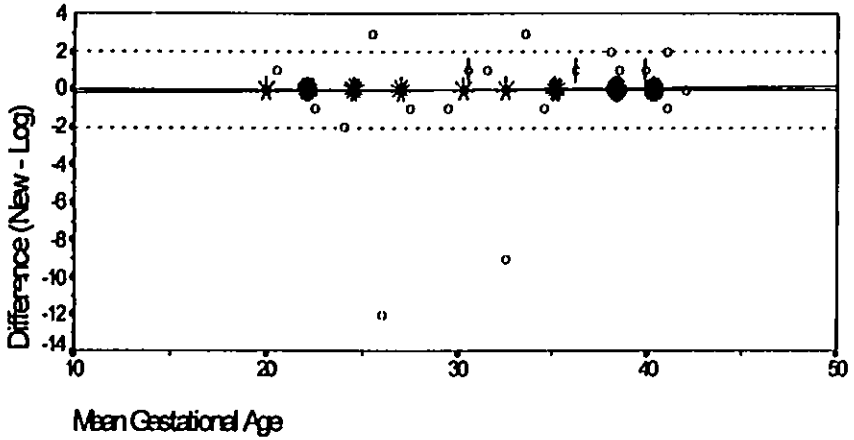
Variable	n	ICC	LL 95% CI	UL 95% CI
Gestational Age	265	.9902	.9860	.9931
Mother's Age	265	.9921	.9888	.9945
Birthweight	263	.9942	.9918	.9960
# Previous Livebirths	263	.9396	.9148	.9574
Weighted Mean ICC		.9791		

The difference versus mean scatterplots for gestational age, mother's age and birthweight all showed some random error, but little evidence of trend or systematic bias (Figures 17a-c). For the plot of number of previous livebirths (Figure 17d), there is a slight trend toward overestimation of parity in the new database as parity increases.

Figure #17a

Errors Associated with Gestational Age

Presumed Stillbirths

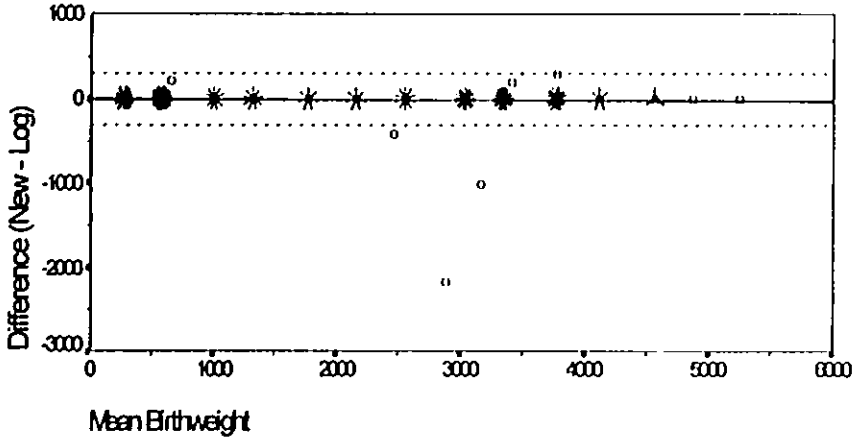


Mean X - Y Difference: -0.03

Figure # 17c

Errors Associated with Birthweight

Presumed Stillbirths

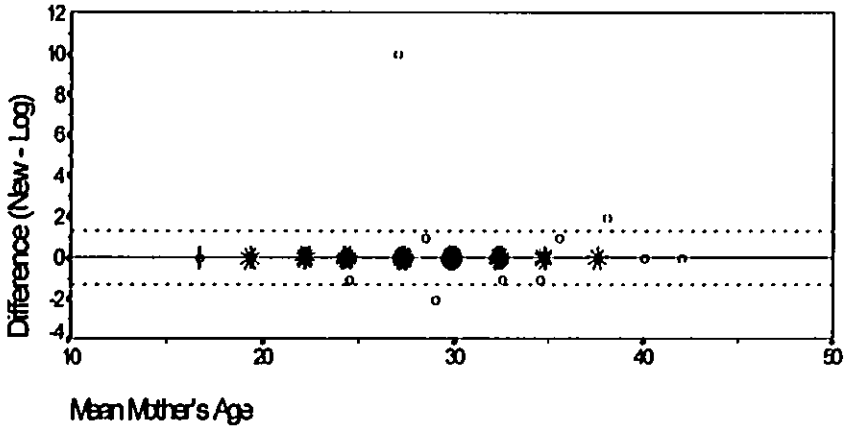


Mean X - Y Difference: -10.19

Figure # 17b

Errors Associated with Mother's Age

Presumed Stillbirths

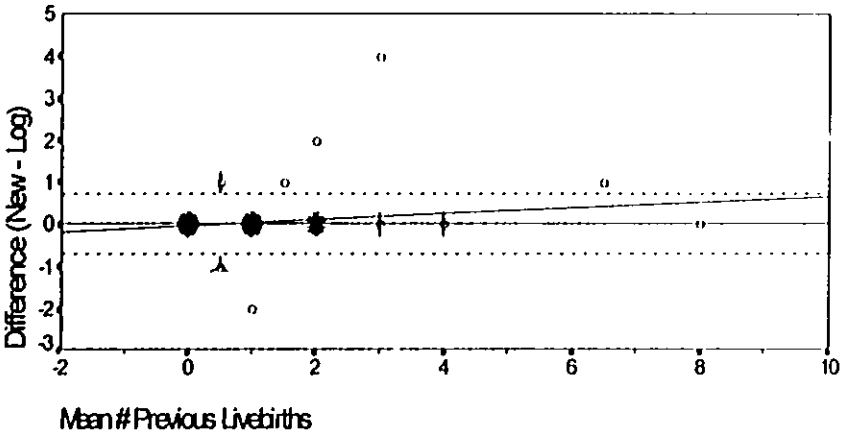


Mean X - Y Difference: 0.03

Figure # 17d

Errors Associated with # Previous Livebirths

Presumed Stillbirths



Mean X - Y Difference: 0.02

Agreement between sources for the nominal scale variable is, again, expressed as Kappa coefficients. Agreement for sex of baby was very high (k=.9612) between the existing database and the logbook based on 265 data pairs. Errors were present among the first 3 digits of postal code in 3.4% of cases and among the last three digits in 7.1% of cases (Table 27, row #1). A further 2.6% had errors in both the first and last half of the postal code. The first three digits of postal code only were present and correct in 10.8% of cases. Postal code was missing entirely in 18.7% of cases.

Table 27
Comparison of Postal Codes
Sample B: Presumed Stillbirths

Sources	N	Wrong First 3 Digits	Wrong Last 3 Digits	Wrong First & Last 3 Digits	Missing Last 3 Digits	Missing
New Database vs. Logbook	268	9 (3.4%)	19 (7.1%)	7 (2.6%)	29 (10.8%)	50 (18.7%)
Old Database vs. Logbook	225[a]	7 (3.1%)	16 (7.1%)	7 (3.1%)	27 (12.0%)	19 (8.4%)

[a] Excluding 1989 births for which postal code was not entered into the existing database

5.7.2.2 Existing PNOB Database

Once again, there was outstanding agreement for all variables, with a weighted mean ICC of .97 (Table 28). Gestational age (ICC=.986) and mother's age (.992) had near-perfect agreement while birthweight (ICC=.96) and number of

previous livebirths (ICC=.937) were slightly lower. All lower limits of 95% confidence intervals exceeded .91.

Table 28
Comparison of Existing Database & Logbook
Sample B: Presumed Stillbirths

Variable	n	ICC	LL 95% CI	UL 95% CI
Gestational Age	264	.9856	.9795	.9899
Mother's Age	265	.9917	.9882	.9942
Birthweight	264	.9605	.9439	.9722
# Previous Livebirths	265	.9368	.9109	.9554
Weighted Mean ICC		.9686		

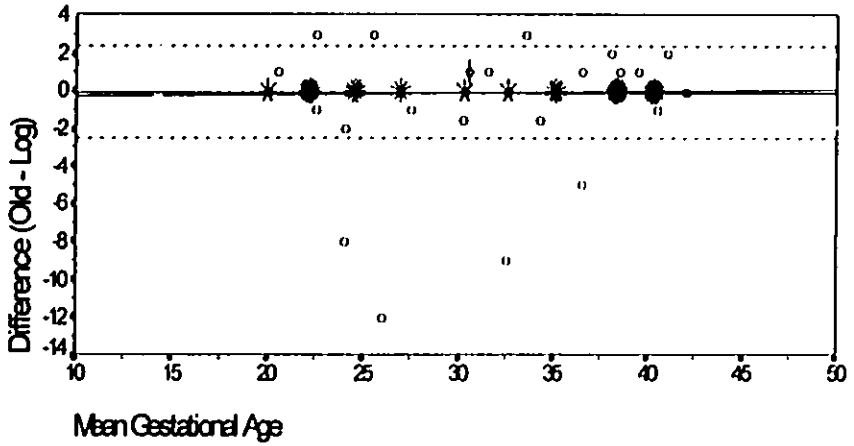
Scatterplots of gestational age, mother's age and birthweight (Figures 18a-c) show minimal systematic bias and trend. For each plot there is at least one example of a highly discrepant pair. The plot of number of previous livebirths shows some positive trend, but still little systematic bias (Figure 18d).

Agreement for sex of baby was excellent at .9377 based on 265 data pairs. Errors in postal code were present in 3.1% and 7.1% of cases for the first and last half, respectively (Table 27, row #2). An additional 3.1% of cases contained errors in both the first and last half of postal code. When only the first three digits were given, they were correct in 12.0% of cases. In 8.4% of cases the entire postal code was missing.

Figure #18a

Errors Associated with Gestational Age

Presumed Stillbirths

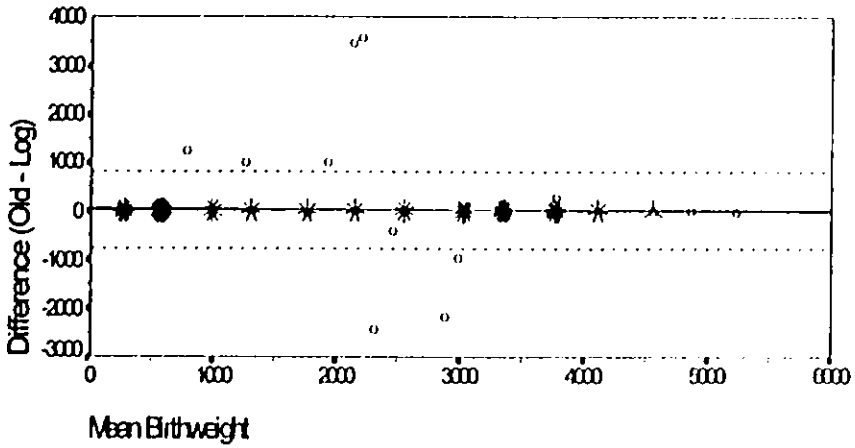


Mean X - Y Difference: -0.10

Figure #18c

Errors Associated with Birthweight

Presumed Stillbirths

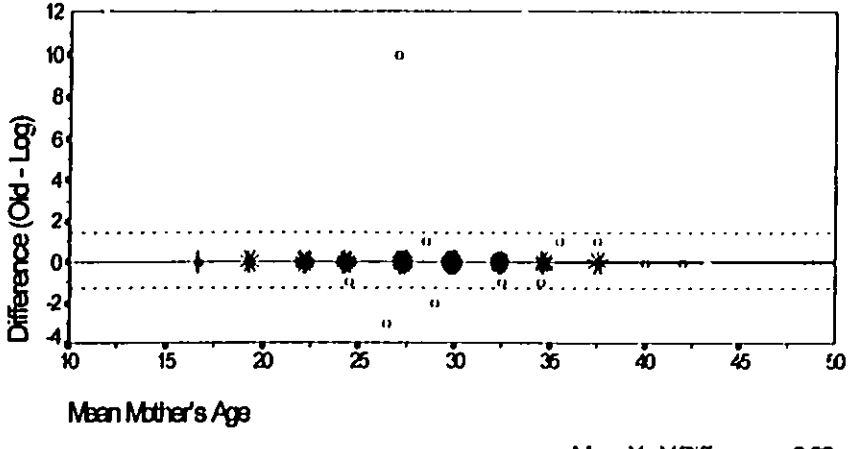


Mean X - Y Difference: 16.19

Figure #18b

Errors Associated with Mother's Age

Presumed Stillbirths

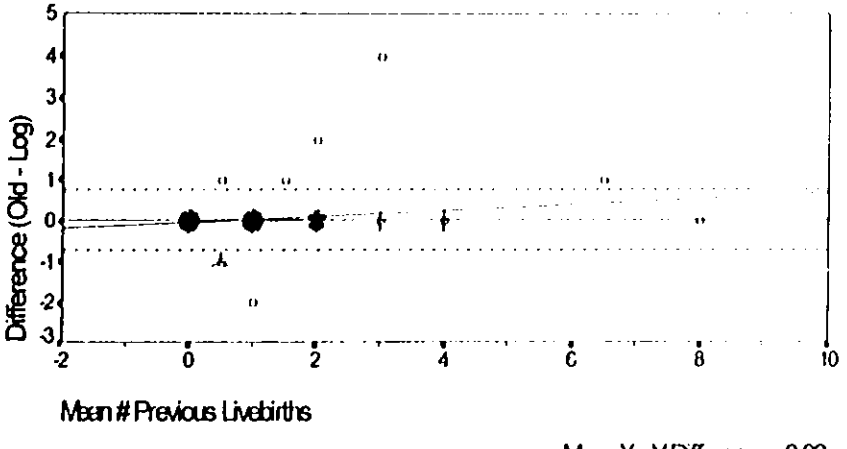


Mean X - Y Difference: 0.02

Figure #18d

Errors Associated with # Previous Livebirths

Presumed Stillbirths



Mean X - Y Difference: 0.02

5.7.3 Sample C: Missing Vital Status in PNOB Database Compared to Logbook

5.7.3.1 Newly-Entered PNOB Database

Agreement between the newly-entered PNOB database and the logbook was, overall, quite high (Table 29). The weighted mean ICC was .959. Agreement for mother's age and birthweight were near-perfect at .9955 and .9973, respectively. Gestational age had slightly lower agreement with ICC=.95 and agreement for number of previous livebirths was lower still at .89 with 95% confidence intervals of .85 and .92.

Table 29
Comparison of New Database & Logbook
Sample C: Missing Vital Status

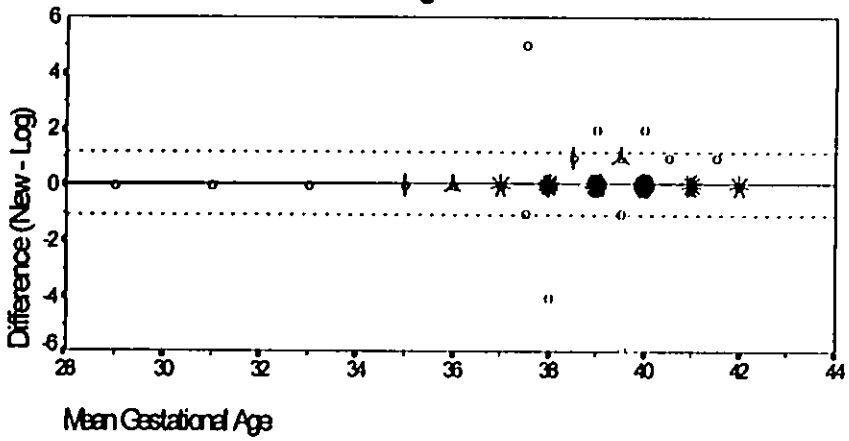
Variable	n	ICC	LL 95% CI	UL 95% CI
Gestational Age	205	.9534	.9341	.9672
Mother's Age	211	.9955	.9935	.9968
Birthweight	207	.9973	.9961	.9981
# Previous Livebirths	210	.8902	.8466	.9219
Weighted Mean ICC		.9590		

Mean versus difference scatterplots are presented in Figures 19a-d. Plots for gestational age, mother's age and birthweight show very little random error, systematic bias and trend. The plot of number of previous livebirths shows some evidence of trend but little systematic bias and random error.

Figure #19a

Errors Associated with Gestational Age

Missing Vital Status

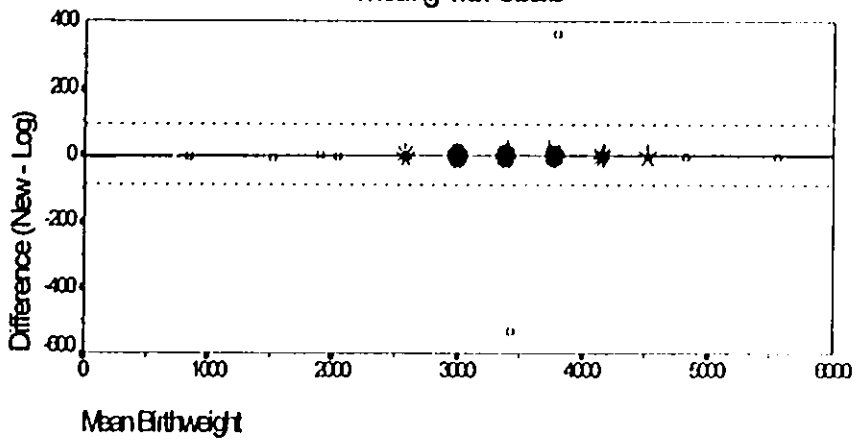


Mean X - Y Difference: 0.06

Figure #19c

Errors Associated with Birthweight

Missing Vital Status

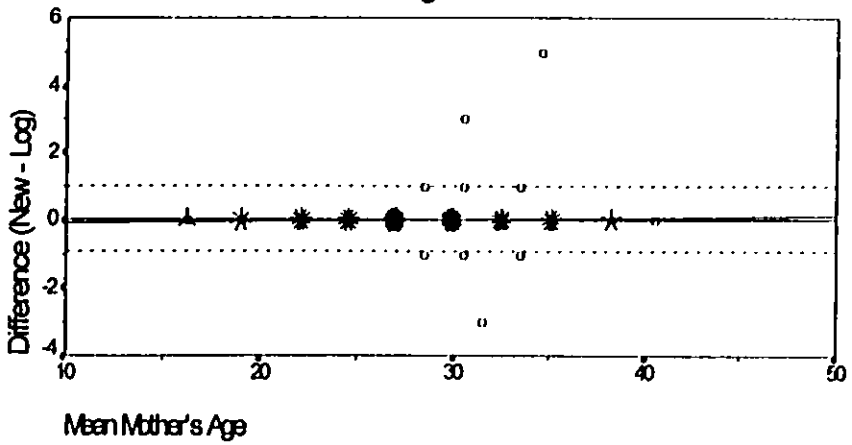


Mean X - Y Difference: -0.62

Figure # 19b

Errors Associated with Mother's Age

Missing Vital Status

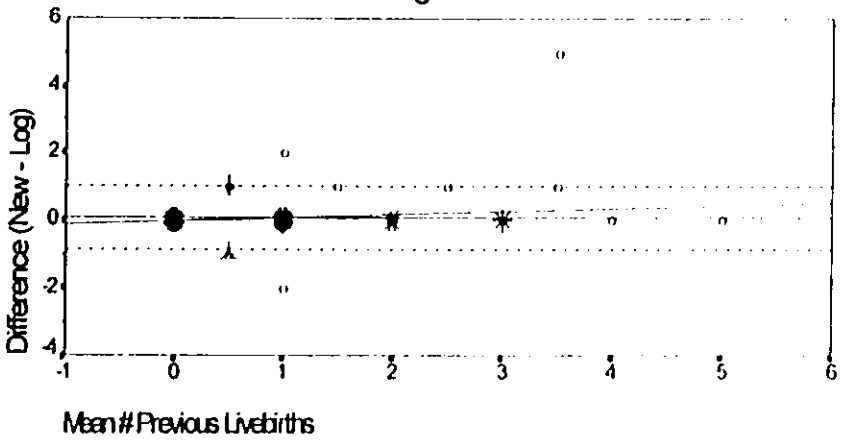


Mean X - Y Difference: 0.02

Figure #19d

Errors Associated with # Previous Livebirths

Missing Vital Status



Mean X - Y Difference: 0.05

Kappa for agreement for sex of baby was .966 based on 177 data pairs. Postal code was compared for 219 data pairs (Table 30, row #1) and contained an error in the first three digits in 3.2% of cases. Errors in the last three digits occurred in 3.7% of cases and errors in both the first and last half of postal code was present in an additional 1.4% of cases. Nearly one quarter of cases had only the postal code sorting area present in the logbook and a further 28.8% were missing the entire postal code.

Table 30
Comparison of Postal Codes
Sample C: Missing Vital Status

Sources	N	Wrong First 3 Digits	Wrong Last 3 Digits	Wrong First & Last 3 Digits	Missing Last 3 Digits	Missing
New Database vs. Logbook	219	7 (3.2%)	8 (3.7%)	3 (1.4%)	54 (24.7%)	63 (28.8%)
Old Database vs. Logbook	146[a]	0	5 (3.4%)	1 (0.7%)	35 (24.0%)	36 (24.7%)

[a] Excluding 1989 births for which postal code was not entered into the existing database

5.7.3.2 Existing PNOB Database

Mean weighted agreement for the existing PNOB database (ICC=.946) was slightly lower than that of the newly-entered PNOB database. Agreement for mother's age and birthweight was extremely high and mirrored that obtained with

the newly-entered PNOB database. Agreement for gestational age and number of previous livebirths were lower at .90 and .89, respectively (Table 31).

Table 31
Comparison of Existing Database & Logbook
Sample C: Missing Vital Status

Variable	n	ICC	LL 95% CI	UL 95% CI
Gestational Age	205	.8989	.8584	.9282
Mother's Age	211	.9955	.9935	.9968
Birthweight	207	.9971	.9958	.9979
# Previous Livebirths	209	.8918	.8489	.9231
Weighted Mean ICC		.9460		

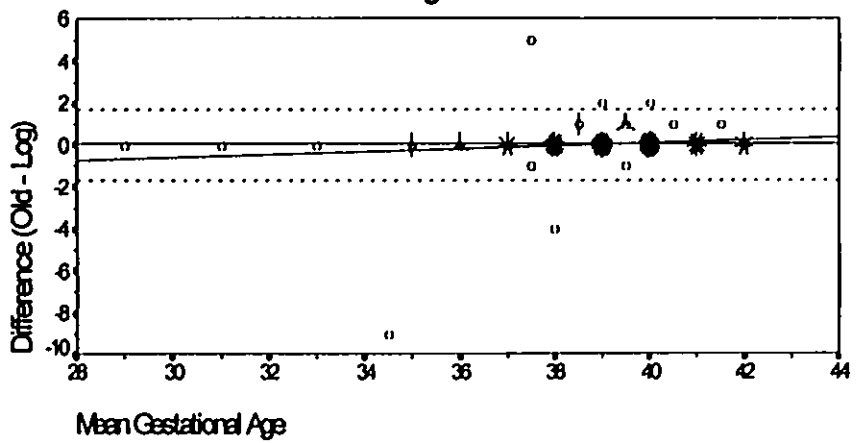
Scatterplots are shown in Figures 20a-d. Plots of mother's age and birthweight show virtually perfect agreement with nearly no random error, systematic bias and trend. Plots of gestational age and number of previous livebirths both show some evidence of a positive trend along with some random error, but little or no systematic bias.

Kappa for sex of baby was high at .955 based on 178 data pairs. Agreement for postal code is shown in row #2 of Table 30. There were no cases in which the first three digits of the postal code alone were not in agreement, although 1 case (0.7%) had errors in both the first and second halves of postal code. Five cases (3.4%) had errors in the last half of postal code. The first three digits of postal code only were given in 24%. A further 24.7% of cases had a missing postal code.

Figure #20a

Errors Associated with Gestational Age

Missing Vital Status

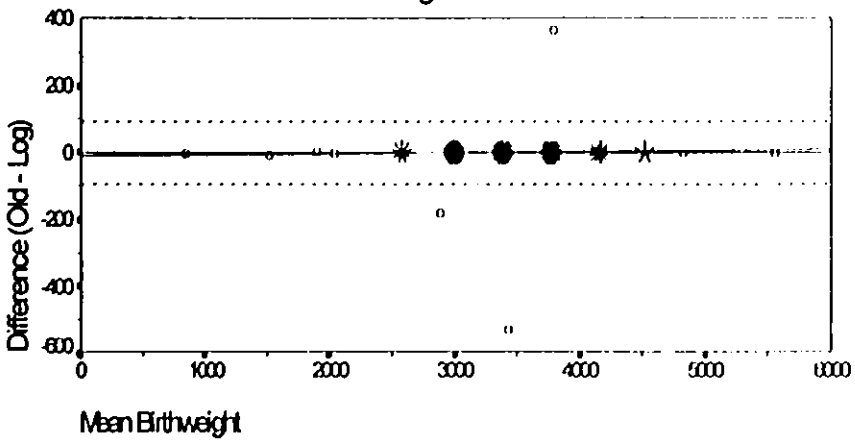


Mean X - Y Difference: 0.01

Figure # 20c

Errors Associated with Birthweight

Missing Vital Status

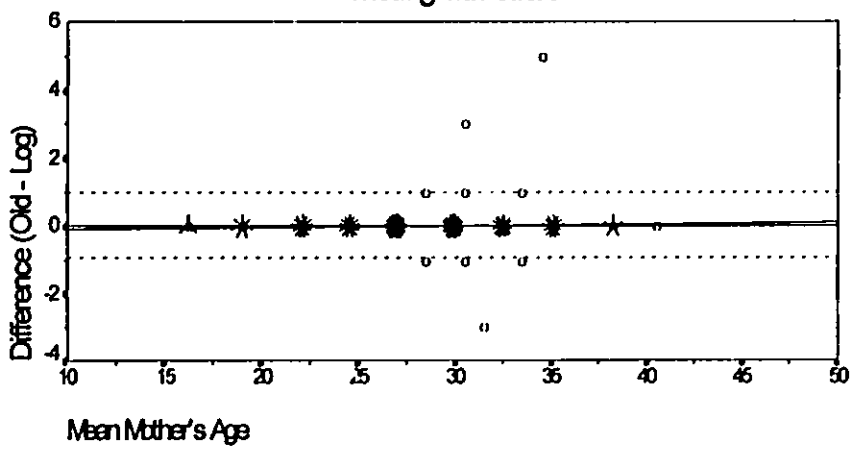


Mean X - Y Difference: -1.66

Figure # 20b

Errors Associated with Mother's Age

Missing Vital Status

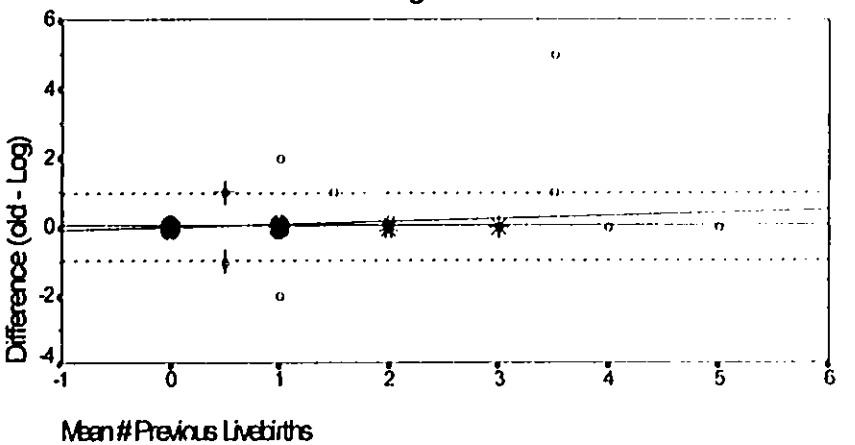


Mean X - Y Difference: 0.02

Figure # 20d

Errors Associated with # Previous Livebirths

Missing Vital Status



Mean X - Y Difference: 0.05

5.7.4 Sample D: False Negatives in PNOB Database Compared to Logbook

5.7.4.1 Newly-Entered PNOB Database

Agreement between the newly-entered PNOB database and the logbook was virtually perfect (Table 32) for gestational age (ICC=1.00), mother's age (ICC=0.9991), birthweight (ICC=1.00) and number of previous livebirths (ICC=1.00). All difference versus mean scatterplots demonstrate this near perfect agreement (Figures 21a-d).

Table 32
Comparison of New Database & Logbook
Sample D: False Negatives

Variable	n	ICC	LL 95% CI	UL 95% CI
Gestational Age	34	1.000	1.000	1.000
Mother's Age	36	.9991	.9982	.9995
Birthweight	34	1.000	1.000	1.000
# Previous Livebirths	36	1.000	1.000	1.000
Weighted Mean ICC		.9998		

Kappa for sex of baby was equal to 1. Agreement for postal code was quite high (Table 33; row 1). Discrepancies were found for one pair each from the first three, the last three and the first and last three digits. The last three digits were missing in three cases (8.3%) and postal code was missing entirely in 6 cases (16.7%).

Figure #21a

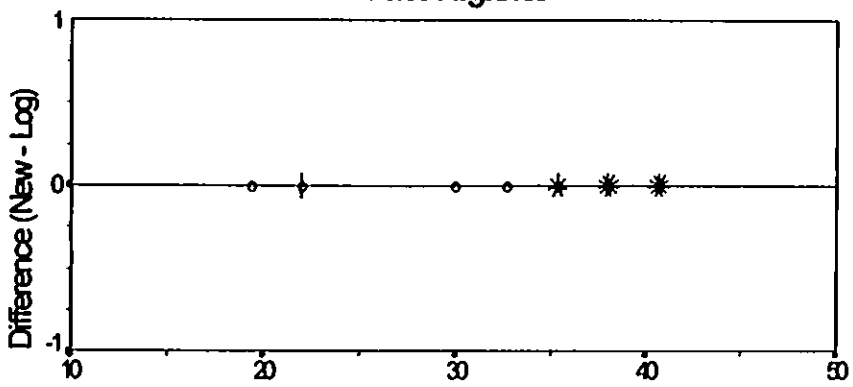


Figure #21c

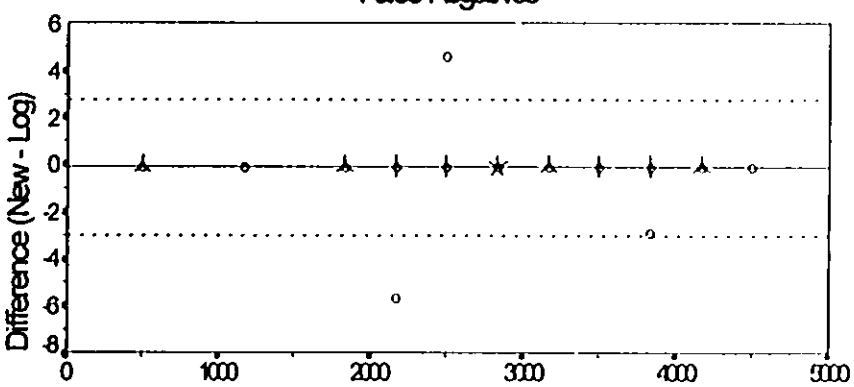


Figure # 21b

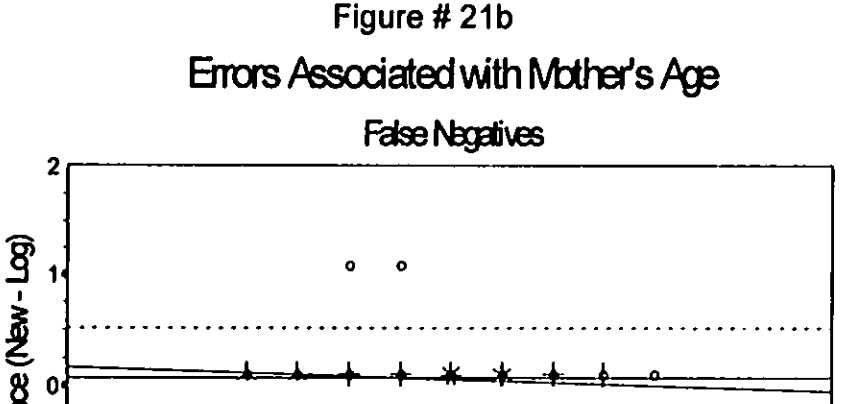


Figure # 21d

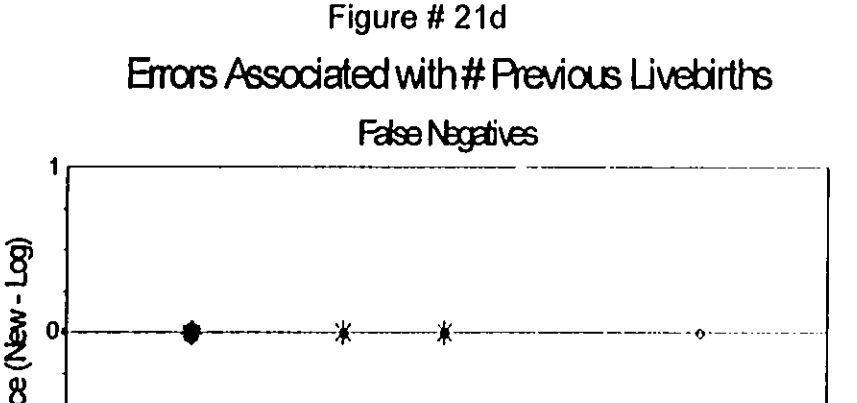


Table 33
Comparison of Postal Codes
Sample D: False Negatives

Sources	N	Wrong First 3 Digits	Wrong Last 3 Digits	Wrong First & Last 3 Digits	Missing Last 3 Digits	Missing
New Database vs. Logbook	36	1 (2.8%)	1 (2.8%)	1 (2.8%)	3 (8.3%)	6 (16.7%)
Old Database vs. Logbook	29[a]	1 (3.4%)	2 (6.9%)	0	2 (6.9%)	4 (13.8%)

[a] Excluding 1989 births for which postal code was not entered into the existing database

5.7.4.2 Existing PNOB Database

Virtually perfect agreement was again obtained for gestational age (ICC=.9977), mother's age (ICC=.9991) and number of previous livebirths (ICC=1.0) (Table 34). For birthweight, unlike with the newly-entered PNOB database, only moderate agreement was obtained for the existing PNOB database (ICC=.71)

Table 34
Comparison of Existing Database & Logbook
Sample D: False Negatives

Variable	n	ICC	LL 95% CI	UL 95% CI
Gestational Age	35	.9977	.9955	.9988
Mother's Age	36	.9991	.9982	.9995
Birthweight	35	.7050	.4934	.8381
# Previous Livebirths	36	1.000	1.000	1.000
Weighted Mean ICC		.9265		

Figure # 22a

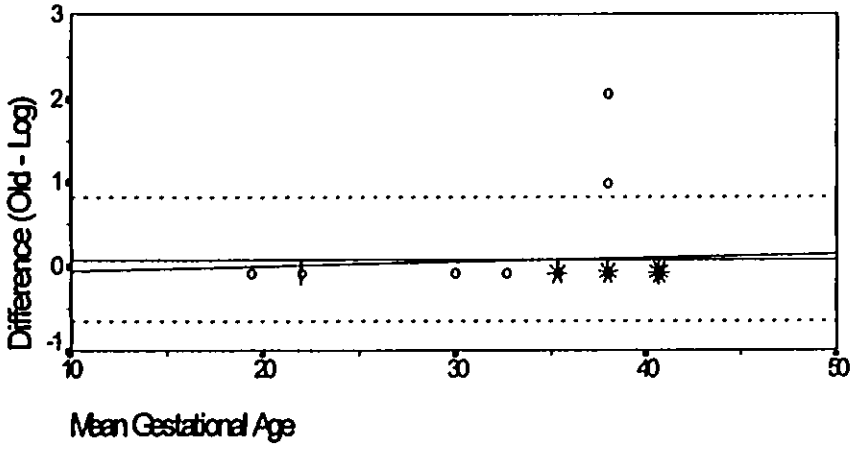


Figure # 22c

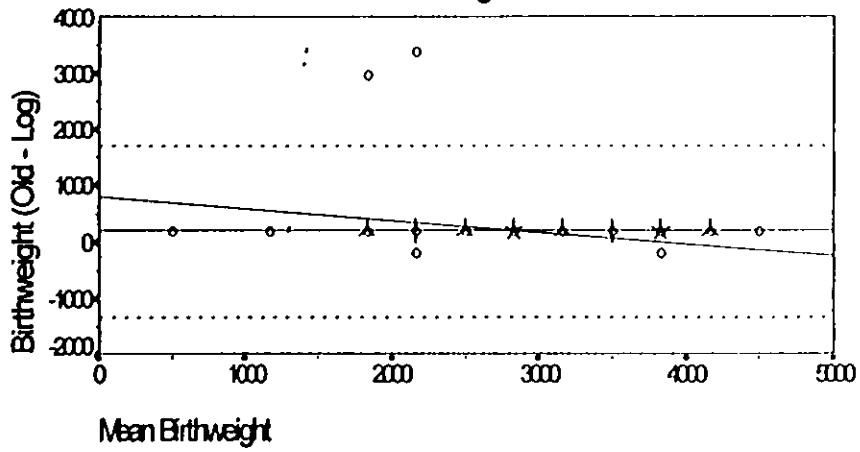


Figure # 22b

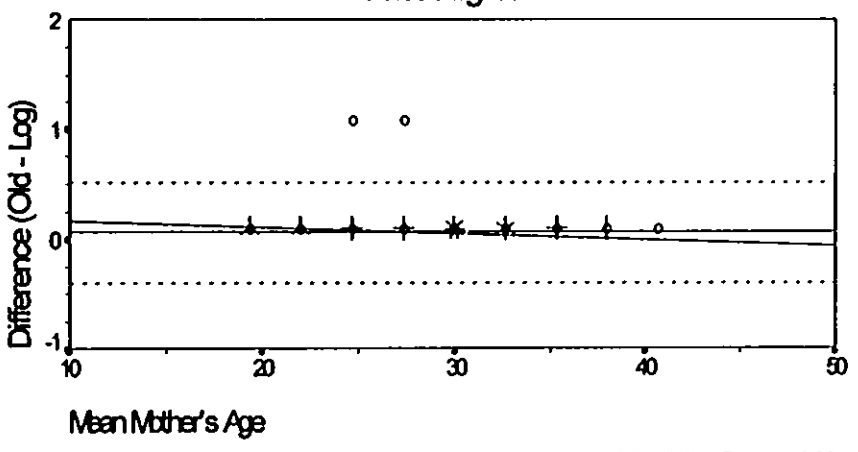
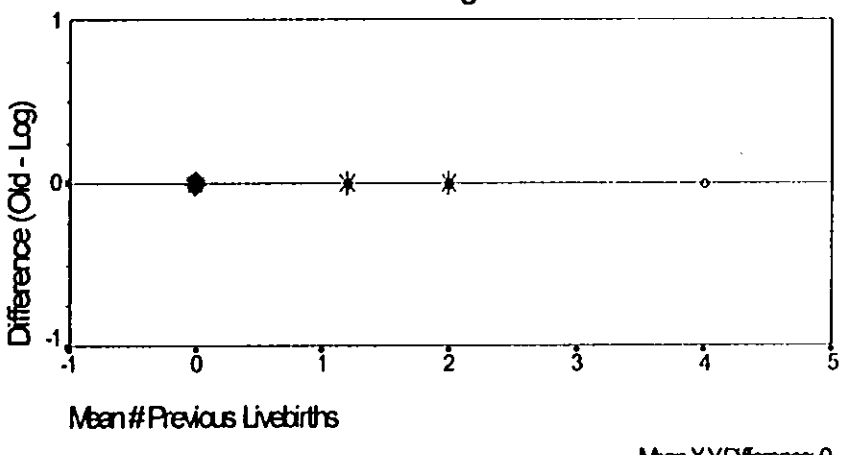


Figure #22d



Scatterplots of difference versus mean (Figures 22a-d) reinforce the near perfect agreement obtained for gestational age, mother's age and number of previous livebirths. The scatterplot of birthweight (Figure 22c) reveals, again, two highly discrepant data pairs. Kappa for sex of baby was perfect (1.0). There was also excellent agreement for postal code with only one discrepancy (3.4%) among the first three digits and 2 discrepancies (6.9%) among the last three digits. The last 3 digits of postal code were missing for 2 records (6.9%) and were missing entirely for 4 (13.8%).

5.7.5 Assessment of Records with a Discrepancy between PNOB Databases and Logbook

The assessment of agreement was again carried out for all babies with a discrepancy or missing value among the existing and new PNOB databases and the logbook, this time using the medical records as the gold standard. Due to the small number of records with discrepancies for any one variable, all four samples (livebirths, stillbirths, missing vital status and false negatives) will be combined. Agreement and variable-specific sample sizes are summarized in Table 35.

Table 35
Comparison of Databases & Medical Records
All Births with Discrepant Data

Variable	Source	N	ICC	LL 95% CI	UL 95% CI
Gestational Age	New Database	43	.9281	.8413	.9648
	Existing Database	43	.9244	.8175	.9644
Mother's Age	New Database	26	.4982	.1551	.7372
	Existing Database	26	.4614	.1114	.7111
Birthweight	New Database	32	.9903	.9799	.9953
	Existing Database	32	.4449	.1306	.6825
# Previous Livebirths	New Database	26	.7308	.3337	.8871
	Existing Database	26	.4320	.0544	.7003

5.7.5.1 Gestational Age

Even among data with discrepancies between the logbook and the existing and new PNOB databases, ICCs for agreement were both quite high at 0.92 - 0.93, indicating that when there was disagreement, it was likely attributable to an incorrect entry in the logbook. Mean versus difference scatterplots (Figure 23a,b) are very similar for the two databases. There is some random error, but little evidence of systematic bias or trend toward increasing differences with either increasing or decreasing gestational age.

5.7.5.2 Mother's Age

Poor agreement for the existing and new databases (ICC = .46 and .50, respectively) versus the medical records was seen among records with discrepancies between the three sources, indicating that when a discrepancy was encountered, the likely source of error was the PNOB form itself rather than the logbook. Among the scatterplots (Figure 23 c,d) there is a small degree of systematic bias (toward overestimation in databases) but no suggestion of increasing differences in either direction. There is one highly discrepant data point which would have greatly influenced the agreement measure, given the small sample size of 26.

5.7.5.3 Birthweight

A different picture emerges for the variable 'birthweight'. In cases with discrepancies between the three data sources, the existing PNOB database was in poor agreement with the medical records (ICC = .44) but the newly-entered PNOB database was in near-perfect agreement (ICC = .99). This finding indicates that when there is a discrepancy, the error is usually in the reading and entering of data at the health department level. The scatterplots highlight the difference between the databases. For the newly-entered database (Figure 24a) there is no systematic bias or trend toward increasing differences at either end of the graph. Random error is also kept to a minimum. The plot of the existing database (Figure 24b) demonstrates a great deal of systematic bias and random error, including several highly discrepant data points. There is suggestion of a trend toward increasing

Figure #23a

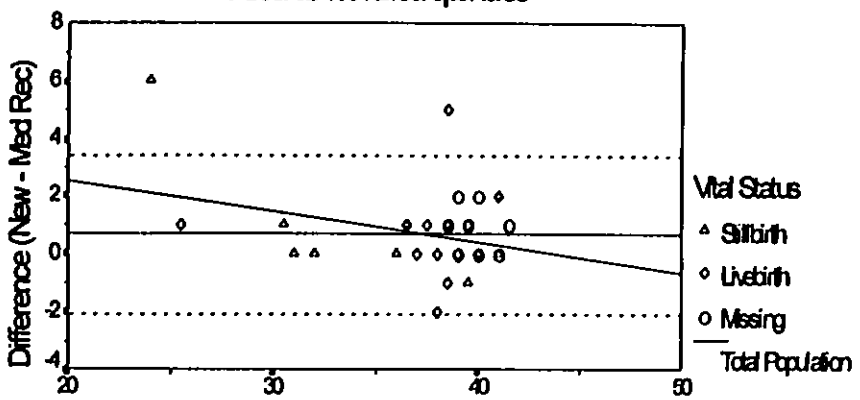


Figure #23b

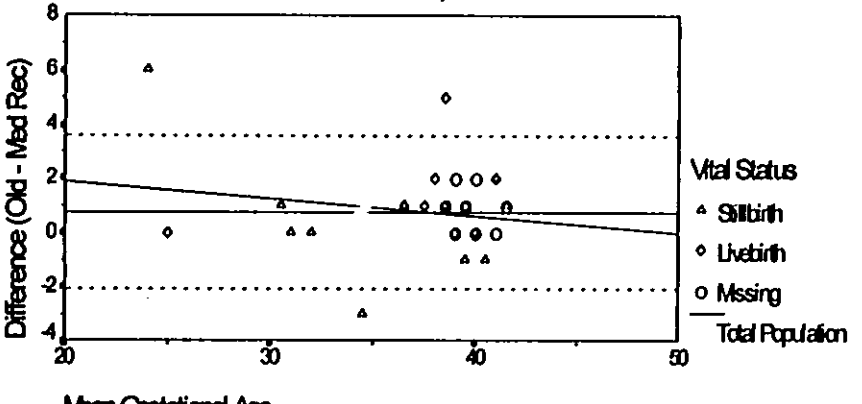


Figure #23c

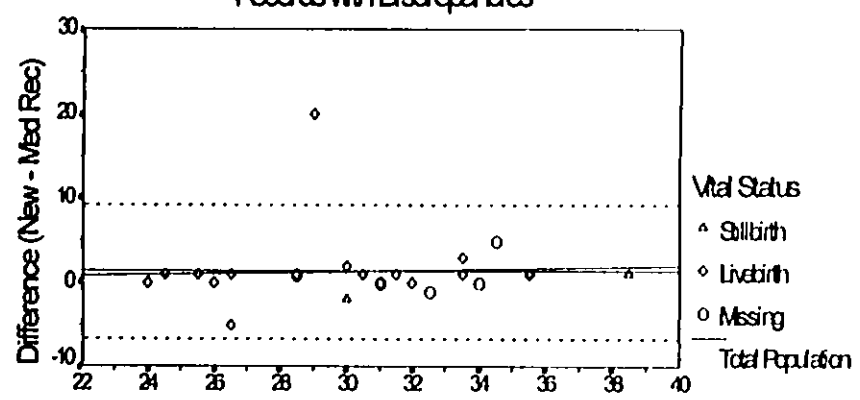
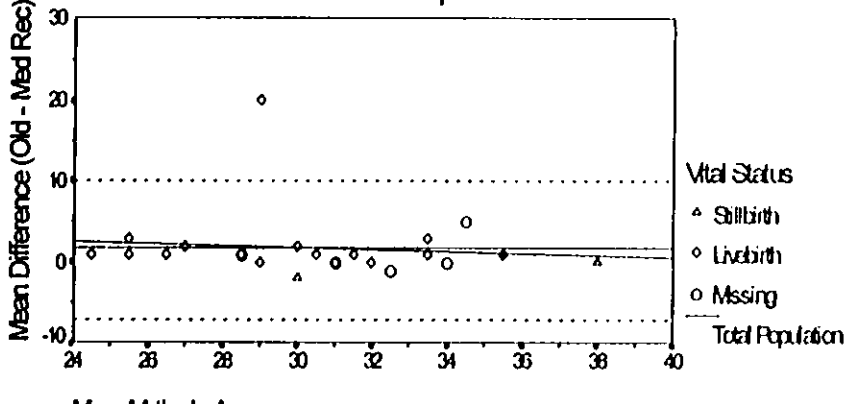


Figure #23d



differences with lower birthweight, which was expected given that many errors were caused by an added zero to the birthweights of very low birthweight babies (<1000g).

5.7.5.4 Number of Previous Livebirths

The pattern of agreement for number of previous livebirths is similar to that of birthweight. When a discrepancy existed, the existing PNOB database was in poor agreement with the medical records (ICC=.43). The newly-entered database, however, was in better but not outstanding agreement with the medical records (ICC=.73) indicating that only some of the problem of accuracy can be taken care of in the health department, the remainder requiring better transcription of data onto the PNOB form at the hospital level. The scatterplot of the newly-entered database and medical records (Figure 24c) shows some random error systematic bias and possibly some evidence of increasing differences with increasing parity. The plot of the existing database and medical records (Figure 24d) shows a great deal of random error, systematic bias and a clear trend toward increasing differences with increasing parity.

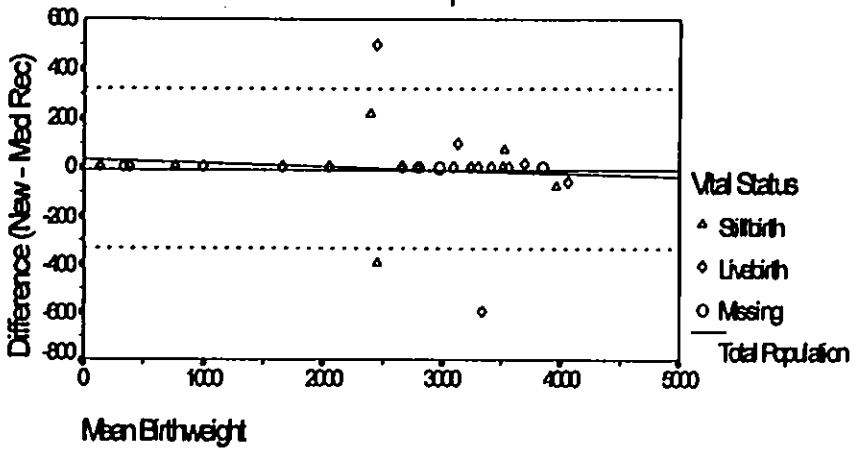
5.7.5.5 Sex of Baby

There were only eight cases with discrepancies for the variable 'sex of baby'. Kappa for both databases was only 0.09, indicating that incorrect information was recorded on the PNOB form at the hospital level.

Figure # 24a

Errors Associated with Birthweight

Records with Discrepancies

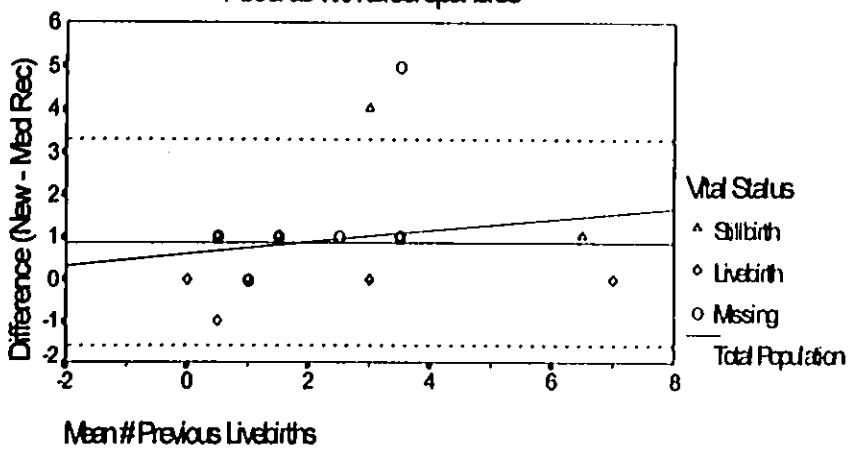


Mean X-Y Difference = -7.09

Figure # 24c

Errors Associated with # Previous Livebirths

Records with Discrepancies

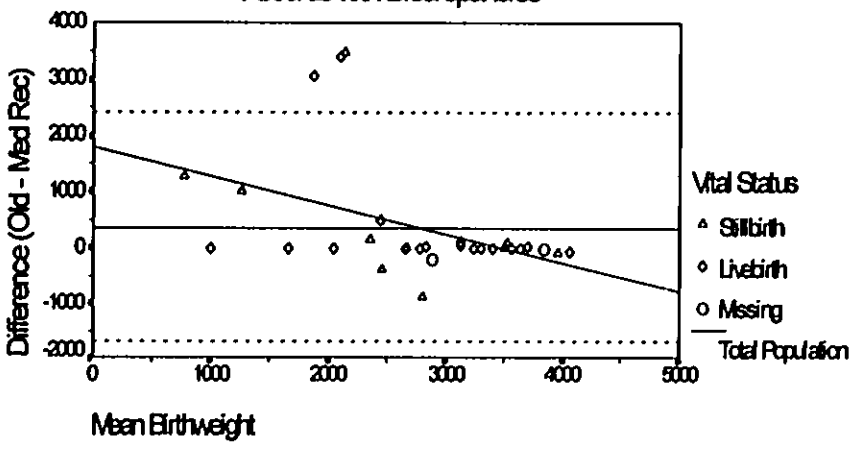


Mean X-Y Difference = 0.85

Figure # 24b

Errors Associated with Birthweight

Records with Discrepancies

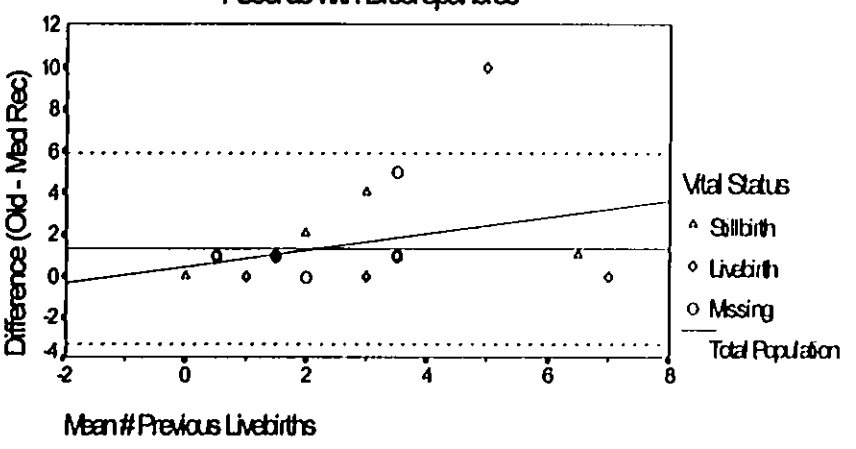


Mean X-Y Difference = 360.53

Figure # 24d

Errors Associated with # Previous Livebirths

Records with Discrepancies



Mean X-Y Difference = 1.31

5.7.6 Assessment of a Random Sample of Cases with No Discrepancy

A random sample was drawn from all cases with no discrepancy for any variables between the existing and newly-entered PNOB database and the caseroom logbooks. Livebirths, stillbirths and babies with missing vital status in the existing PNOB database were analyzed together because there was no reason to suspect any differences in reliability and validity of the data among perfect records of those babies born live or still, and to economize on sample size.

Agreement on all variables (Table 36), particularly gestational age, mother's age and birthweight were virtually perfect (ICC > .99). Agreement for number of previous livebirths was slightly lower (ICC=.97), but still exceptional.

Table 36
Comparison of Perfect Data & Medical Records
Sample E: All Samples Combined

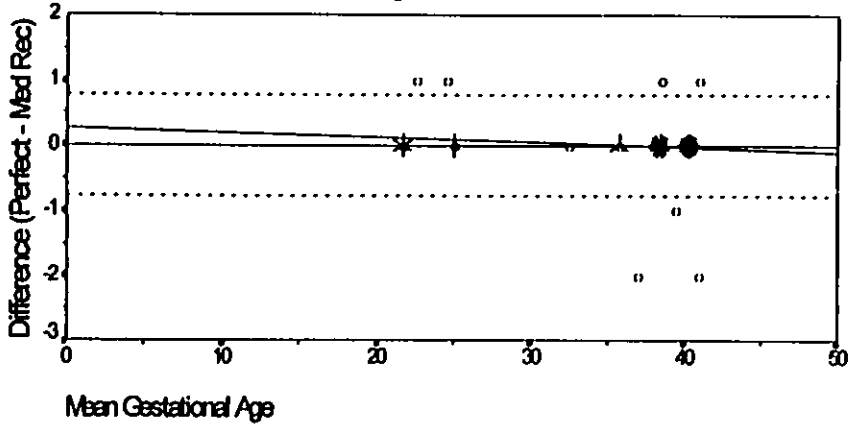
Variable	n	ICC	LL 95% CI	UL 95% CI
Gestational Age	92	.9982	.9973	.9988
Mother's Age	95	.9946	.9919	.9964
Birthweight	92	.9918	.9877	.9945
# Previous Livebirths	95	.9727	.9593	.9817
Weighted Mean ICC		.9892		

Scatterplots of gestational age, mother's age and birthweight all show no evidence of systematic bias or trend (Figure 25a-d). There is a slight trend present in the scatterplot of number of previous livebirths (Figure 25d) showing

Figure # 25a

Errors Associated with Gestational Age

Perfect Agreement - 3 Sources

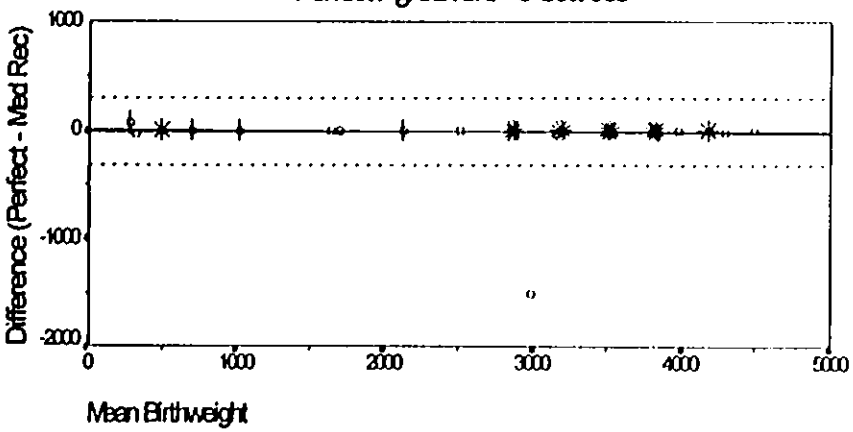


Mean X - Y Difference: 0

Figure # 25c

Errors Associated with Birthweight

Perfect Agreement - 3 Sources

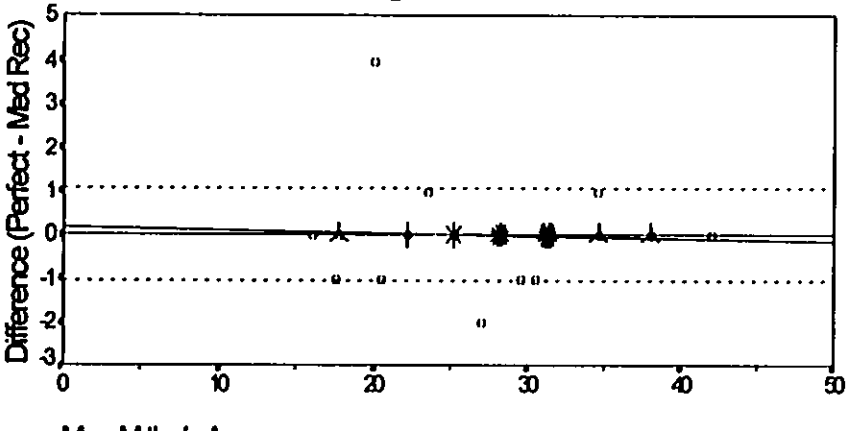


Mean X - Y Difference: -15.05

Figure # 25b

Errors Associated with Mother's Age

Perfect Agreement - 3 Sources

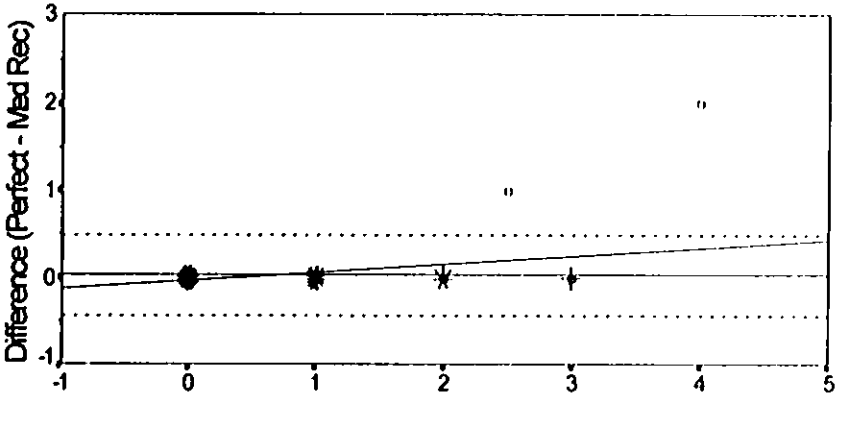


Mean X - Y Difference: 0

Figure # 25d

Errors Associated with # Previous Livebirths

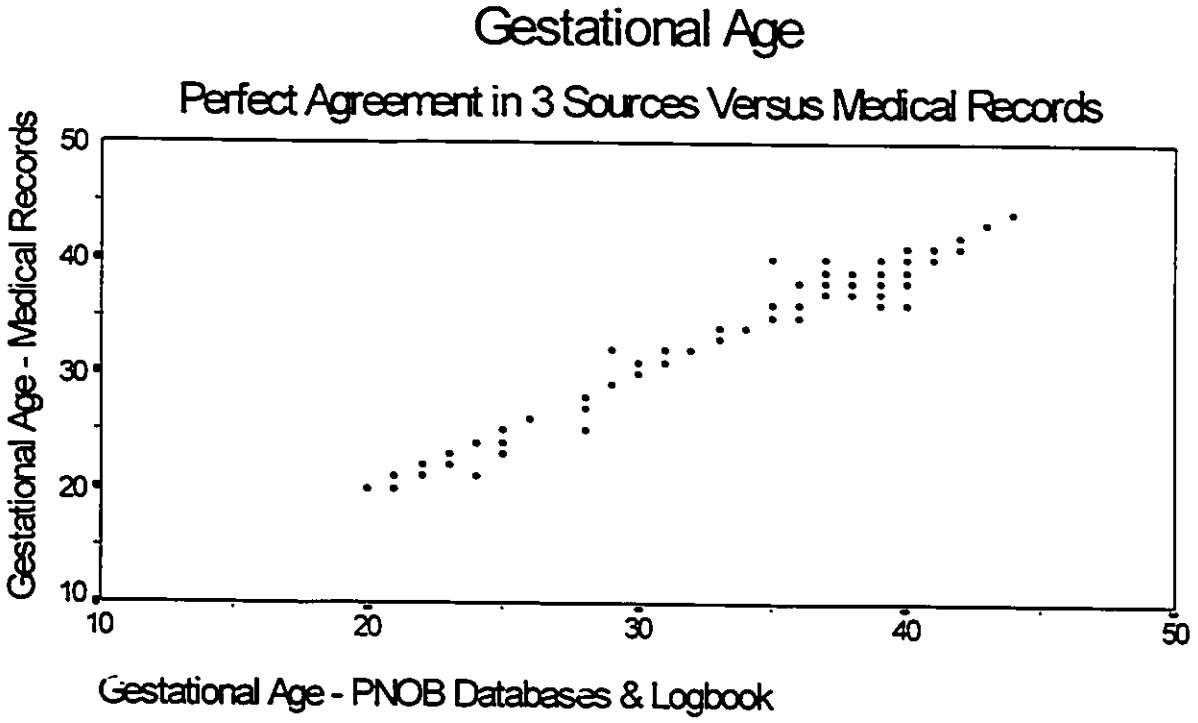
Perfect Agreement - 3 Sources



Mean X - Y Difference: 0.03

overestimation of parity in the databases and logbook with an increasing number of previous livebirths. A raw-data scatterplot of gestational age is provided to demonstrate the near-perfect agreement between the PNOB databases and logbook, and the medical records (Figure. *Kappa* for the variable sex was quite high at 0.9774.

Figure 26



5.8 Assessment of Liveborn/Stillborn Status

Two hundred nineteen stillbirths were confirmed in this study. The rates of stillbirth among the hospitals are shown in Table 37.

Table 37
Rates of Stillbirth Among Hospitals

Hospital	Frequency	Stillbirth Rate
A	74	.644%
B	79	.901%
C	24	.284%
D	29	.358%
E	13	.709%
TOTAL	219	

5.8.1 Existing PNOB Database

According to the existing PNOB database, 270 babies were stillborn. The status of three of these apparent stillbirths could not be confirmed in either the logbook or the mother's medical chart. There were 85 false positives (livebirths incorrectly identified as stillbirths in the PNOB database) and 29 false negatives (stillbirths erroneously identified as livebirths in the database). Two hundred twenty-four babies had a missing value for liveborn/stillborn in the existing PNOB database. Most of these 224 babies were confirmed to be liveborn, except for eight who were confirmed in the logbooks/medical records to be stillborn (Table 38).

Table 38
Existing PNOB Database Information Versus Confirmed Status

Source		Confirmed Status			Total
		Stillborn	Liveborn	Missing	
Existing PNOB Database	Stillborn	182	85	3	270
	Liveborn	29			
	Missing	8	214	2	224
Total		219			

After excluding the eight babies for whom the existing PNOB database contained missing data for liveborn/stillborn status, sensitivity was moderately high at 86.3% (182/211). When the eight stillbirths that were overlooked are included in the denominator, sensitivity drops slightly to 83.1% (182/219). After excluding the three babies whose outcomes could not be verified, positive predictive value of the existing PNOB database was only 68.2%. When those three babies are included in the denominator, positive predictive value decreases slightly to 67.4%.

As a result of the finding of a large number of babies with discrepant live/stillborn status, a *post-hoc* analysis was done to explore the similarities and dissimilarities of babies incorrectly identified and the true population to which they belong (i.e. true livebirths and apparent but incorrectly identified stillbirths; and true stillbirths and apparent livebirths). Independent sample t-tests were performed on the variables 'birthweight' and 'gestational age' and the results are shown below in Table 39 for comparisons between true livebirths and livebirths incorrectly reported as stillbirths.

Table 39
Comparison of True Livebirths
and Incorrectly Identified Livebirths - Existing Database

		N	Mean	Standard Deviation	Standard Error of Mean
Birthweight	True Livebirths	450	3382	609	28.7
	Livebirths Incorrectly Identified as Stillbirths	84	3365	750	81.8
Gestational Age	True Livebirths	452	38.96	2.345	0.111
	Livebirths Incorrectly Identified as Stillbirths	84	38.38	3.127	0.341

For birthweight, the Levene's test for equal variances yields an F of 2.801 (p=0.095), indicating that a t-test for equal variances can be used. The t-test produces a t of 0.23 (p=0.819). The standard error of the difference is 75 and the 95% confidence interval of the difference ranges from -130.5 to +165.

For gestational age, the Levene's test for equal variances yields an F of 4.439 (p=0.036) indicating unequal variances unlikely to be produced by chance alone. Therefore, a t-test with unequal variances was performed, yielding a t of 1.62 (p=0.108) and a standard error of the difference of 0.359. The 95% confidence interval for the difference ranges from -0.130 to +1.293.

Differences in birthweight and gestational age between stillbirths incorrectly reported as livebirths and true stillbirths are more apparent, as shown in Table 40.

Table 40
Comparison of True Stillbirths and
Incorrectly Identified Stillbirths - Existing Database

		N	Mean	Standard Deviation	Standard Error of Mean
Birthweight	True Stillbirths	180	1491	1216	90.7
	Stillbirths Incorrectly Identified as Livebirths	28	2690	961	181.7
Ge[illegible]	True Stillbirths	181	29.5	7.040	0.523
	[illegible]tillbirths Incorrectly [illegible]ntified as Livebirths	29	35.2	5.960	1.107

For b[illegible]rthweight, the Levene's test showed differences in variance and the unequal variances independent samples t-test produced a t of 5.90 ($p<0.001$). The standard error of the difference was 203, resulting in 95% confidence intervals from +788.8 to +1608.

For gestational age, the Levene's test showed unequal variances ($p=0.001$), indicating that an unequal variance independent sample t-test should be used, yielding a t of 4.66 ($p<0.001$). The standard error of the difference was 1.224 and the 95% confidence intervals were bounded by 3.233 and 8.176.

5.8.2 Newly Entered PNOB Database

Sensitivity was calculated for the newly-entered PNOB database (Table 41) to see how much of the problem could be attributed to faulty data-entry at the health department level and how much could be attributed to incorrect information on the PNOB form, itself. After excluding the missing data, sensitivity was 99.1%

(209/211). By including the missed stillbirths in the denominator, sensitivity dropped slightly to 95.4% (209/219). This value of sensitivity for the newly-entered PNOB database is much higher than that calculated with the existing database. Positive predictive value can be estimated at 73.3% when the three babies with unknown status are excluded. When included, positive predictive value drops slightly to 72.6%.

Table 41
Newly Entered PNOB Database Information Versus Confirmed Status

Source		Confirmed Status			Total
		Stillborn	Liveborn	Missing	
New PNOB Database	Stillborn	209	76	3	288
	Liveborn				
	Missing	8	223	2	233
Total		219	753	8	

A comparison of birthweight and gestational age for true livebirths and livebirths incorrectly identified as stillbirths appears inTable 42.

Table 42
Comparison of True Livebirths and
Incorrectly Identified Livebirths - Newly Entered Database

		N	Mean	Standard Deviation	Standard Error of Mean
Birthweight	True Livebirths	447	3362	641	30.3
	Livebirths Incorrectly Identified as Stillbirths	75	3416	682	78.8
Gestational Age	True Livebirths	451	38.89	2.510	0.118
	Livebirths Incorrectly Identified as Stillbirths	75	38.76	2.241	0.259

For birthweight, Levene's test showed equal variances (p=0.437) and, as a result, a t-test with equal variances was used and yielded a t of -0.67 (p=0.502) for birthweight. The standard error of the difference was 81 and the 95% confidence intervals ranged from -213 to +104, indicating no difference in birthweight between true livebirths and livebirths incorrectly reported as stillbirths.

For gestational age, the Levene's test showed equal variances (p=0.953). The equal variance t-test yielded a t of 0.42 (p=0.676). The standard error of the difference was 0.308 and the 95% confidence intervals included the values of -0.477 and +0.735, indicating no difference in gestational age between true livebirths and livebirths incorrectly reported as stillbirths.

No analysis could be done on the true stillbirths and stillbirths incorrectly identified as livebirths because there were only two babies identified in the newly-entered PNOB database as livebirths who were truly stillborn. These two stillbirths weighed 1461g and 1000g and were born at 33 and 29 weeks of gestation, respectively.

6. DISCUSSION

6.1 Quality of This Study

It was important that the database used in this study to assess the information on the PNOB form was of high quality. The method chosen to capture the actual information on this form was to enter this information into another separate database. Great care was taken in entering the data. The intra-rater reliability (double-entry) of the newly-entered PNOB database was high (99%). An initial comparison with the existing PNOB database identified a few errors which were corrected in the new database. Thus, the newly entered PNOB database was of high quality. Double-entering all data rather than just 10% would have further improved the quality of the data in this study. By double-entering all data and verifying any discrepancies from the original sources, one could have had complete assurance of the accuracy of the raw data, except for the possibility of entering the same incorrect value twice. Time and resource limitations precluded this step.

The comprehensiveness and scope of this study could have been broadened by gathering at least part of the data backwards, that is, by randomly selecting some births from the hospital medical records and tracing them back through the logbook to the existing PNOB database. Had this strategy been incorporated into the present study, it would have been possible to comment on the completeness of the existing PNOB database. Without this design detail, it is difficult to comment on

the completeness of reporting of all births which Klaucke lists as a critical attribute of a surveillance system [3].

However, one design element did allow me to assess completeness of coverage of stillbirths alone. Since every record in the caseroom logbooks was examined, it was possible to determine that eight stillborn babies were not included in the existing PNOB database. Many of these babies were extremely premature, with four barely meeting the Vital Statistics Act criteria for distinguishing a stillborn baby from a miscarriage. Strict adherence to the Vital Statistics Act definition of a stillborn baby having attained a gestational age of 20 completed weeks *or* 500 grams is required. This underreporting of extremely premature or extremely low birthweight stillbirths is not surprising and has been reported elsewhere as being potentially "substantial" [4]. The extent of this type of underreporting in Ottawa-Carleton was not addressed in this study.

The sensitivity and positive predictive values of the two PNOB databases for stillbirth are dependent on the accurate identification of *every* stillbirth in Ottawa-Carleton. Although the existing database and the logbook were not always in agreement regarding vital status, there was no instance in which both the existing database *and* the logbook indicated that a baby was born alive that was later confirmed by the medical chart to have been stillborn. Even when an entry in the existing database had a missing value for whether the baby was born alive, there was still no instance in which the logbook incorrectly indicated that a baby was stillborn. There were, however, two instances in which both the existing database

and the logbook indicated that a baby was stillborn that were later confirmed by the mothers' medical charts to have been liveborn. Given this information, there is sound evidence to believe that every stillbirth was identified within a small margin of error.

The calculation of the sensitivity of the newly-entered PNOB database is an accurate measure since all, or very nearly all stillbirths were identified. Unlike the existing PNOB database, the positive predictive value of the newly-entered PNOB database represents a "best estimate" since not every PNOB form was physically examined by this investigator. The 'a' cell of true positives is complete, but the 'b' cell may not be entirely complete. To have missed some babies in the 'b' cell would require that a baby truly liveborn have a physical PNOB form reading 'stillborn' but incorrectly entered as 'liveborn' by the Health Department's data entry clerk. The likelihood of such a compounded error is low. Nonetheless, the positive predictive value of the newly-entered PNOB database should be viewed with an appropriate degree of caution.

There are concerns expressed in the literature about the quality of hospital medical records as a gold standard. For this thesis I have chosen to use a combination of hospital caseroom logbooks and maternal medical records as the gold standard. For all studies of validation of birth records identified from the literature, the gold standard used for comparison was the mother's or newborn's medical chart. One study was located, however, which used operating room logbooks as the gold standard for a study to determine the sensitivity and positive

predictive value of a database [10] of surgical procedures. In cases of discrepancy between the hospital database and the operating room logbooks, the medical records were examined. The study's authors did not provide any clues as to which source (i.e. logbook or medical records) would be a better gold standard.

Most of the concerns about the degree of "goldness" of hospital medical records pertain to whether the relevant information is captured within it. In the present study, this issue was not a great concern since the study variables were objective ones which were nearly always present in the chart.

6.2 Interpretation of Agreement Measures

The literature provides only limited guidance for the interpretation of agreement measures. Nearly twenty years ago Landis and Koch [32] proposed the following simple classification scheme in Table 43 for kappa (recall the equivalency of kappa and the ICC):

Table 43
Kappa Classification Scheme

Kappa	Strength of Agreement
0.00	poor
0.01 - 0.20	slight
0.21 - 0.40	fair
0.41 - 0.60	moderate
0.61 - 0.80	substantial
0.81 - 1.00	almost perfect

While this scheme for interpreting agreement coefficients closely resembles that already widely used for interpreting the Pearson correlation coefficient, this guide to interpretation is not above criticism. This guide makes no distinction between different applications of the data for which the agreement coefficient was calculated. Suppose the ultimate goal of the study were to demonstrate sufficient agreement between judges to justify combining the scores of several raters to reach an average score. In this case, perfect agreement is probably not necessary and 'reasonable' agreement is all that is needed. On the other hand, suppose the goal of a study was to demonstrate sufficient agreement between judges to allow for one judge's score to substitute for that of another judge. In this case, it is clear that the standard of agreement required to substitute one for the other should be much higher than that required to pool results with reasonable confidence. This latter situation of substitution is the outcome with which this study is concerned. Dunn, in a book solely devoted to the topic of reliability studies has suggested that a more stringent interpretation of kappa is needed when the goal of a study is to substitute one measurement for another [33].

For this study I have chosen to call values of ICC in excess of 0.90 excellent. Values greater than 0.80 are likely acceptable, but need to be used with caution. This classification is much more stringent than that proposed by Landis and Koch [32], but I believe it is justified by the argument that the ultimate goal of this study is to demonstrate that one source can be safely substituted for another. Another reason for adopting such a rigorous classification concerns the objective nature of

the data being compared. Unlike other recent studies of inter-rater reliability measuring complex constructs such as dementia [34] and tumour characteristics of malignant melanoma [35], this study measures agreement for very easy-to-measure variables (with the possible exception of gestational age) and, as a result, extremely high inter-rater agreement should be expected.

6.3 PNOB Record Completion in Each Hospital

There was consistency between hospitals with respect to completion of PNOB forms. Timely completion of the PNOB form could potentially be an important factor determining the accuracy of the information contained within it. All hospitals reported that forms were usually completed within an hour, most within the caseroom. Hospitals reported differing sources of information for the actual PNOB form; two used information from the mother's chart, two completed the PNOB form in parallel to the mother's chart and one completed the form from the caseroom nurses' notes. Only two hospitals had systems in place for checking the accuracy of the data on the PNOB form. Unfortunately, this study was unable to address the superiority of different combinations of quality assurance measures. The sample size, limited by the number permitted by the hospitals, was insufficient to demonstrate any differences.

6.4 *Quality of the Ottawa-Carleton Health Department's PNOB Database*

The descriptive statistical analysis failed to demonstrate any statistically significant differences between mean values for gestational age, mother's age and birthweight or for proportions for sex and number of previous livebirths obtained by the three data sources. This finding of no mean difference does not imply agreement for individual records, but does indicate that the existing PNOB database is presently able to provide accurate and complete population-based descriptive statistics

The reliability of data entry from the PNOB form as measured by comparing the existing and newly-entered PNOB databases was quite high for all variables, with most ICCs and kappas in excess of 0.95. While agreement for vital status (k=0.90) was nearly perfect by the standards of Landis and Koch [32], put into to proper context in which the outcome (live or stillborn) is simply read off a form and entered into a database, this measure of agreement seems very low. There were 27 babies identified in the logbook which the existing PNOB database failed to identify as stillborn, even when the PNOB form (new PNOB database) clearly indicated "stillbirth". There were a further 9 stillbirths later discovered in the logbook which the existing PNOB database indicated were liveborn but upon examining the PNOB form, I could not confidently ascribe to either "liveborn" or "stillborn". This is a very objective and critical measure with critical implications where nothing short of 100% agreement should be expected.

or alive. Most continuous variables had intraclass correlation coefficients greater than 0.90. Gestational age, maternal age, birthweight and sex of baby were consistently extremely accurate across all pairwise comparisons and for livebirths, stillbirths and babies with missing vital status. Very minor problems were encountered with number of previous livebirths for livebirths (ICC = 0.89) and 'postal code'. The rate of 'any error' for postal code was consistently between 10 and 15% for livebirths and stillbirths. For babies with missing vital status the error rate was much lower (< 10%) but many more PNOB forms with postal code missing entirely. As a rule, mistakes were always more common among the last three rather than the first three digits, indicating that while the entire postal code may not be extremely accurate, the first three digits can be used accurately about 95% of the time. Even if agreement isn't perfect, if one is willing to tolerate a small amount of possible error, the health department's existing PNOB database can provide a reasonably accurate account of each baby's critical information.

Comparisons made between medical records and the PNOB databases were split into those records with no discrepancies among the existing and new databases and logbook and those with discrepant records. For babies with discrepancies between the logbook, existing and new databases, the medical records were in agreement with the existing and newly-entered databases equally well for gestational age (ICC = 0.92) indicating that the logbook was the major contributor of discrepancies. Equally poor agreement was obtained with the existing and

newly-entered databases for mother's age (ICC < 0.50) indicating that in cases of discrepancy it is usually the PNOB form which is wrong.

Particularly large differences in agreement between the two PNOB databases were obtained for birthweight (among stillbirths) and number of previous livebirths (among livebirths) indicating that a large portion of the discrepancy of the existing PNOB database and the gold standards can be attributed to the health department's interpretation and entry of the data. Intervention at the health department level would virtually correct all errors with birthweight, and would reduce some of the errors associated with number of previous livebirths. This finding can be encouraging since it points to the fact that large gains in the accuracy of the data can be expected and achieved through changes within the regional health department's control.

Medical records charts were examined for a random sample of births with perfect agreement between the PNOB databases and the logbook. It is reassuring that among these records with no discrepancies, agreement for most variables was exceptional. The finding indicates that when the logbook and the PNOB form are in agreement for all variables, one can be extremely confident that they will also be in agreement with the medical records.

One important qualification must be noted for the accuracy of the variable "number of previous livebirths". Roughly 85% of all women had had either zero or one previous livebirths, with a further 10% having had two. This is an example of a homogenous setting, in which the full measurement scale is not being used.

Measures of reliability agreement are variance-dependent. For this reason, it is not surprising to find a lower degree of inter-rater reliability by virtue of the fact that the between-subject variance is very low. For the same degree of within-subject variance, the lower the between-subject variance the lower the ICC will be [16].

The ICC of a very homogenous dataset, like number of previous livebirths, will naturally be low. This doesn't mean that the ICC as a measure of agreement is inappropriate. It simply means that the ICC can only be generalized to a population with a similar range for the variable in question [36]. If, for some reason, the population changed and women were routinely having many babies (i.e. greater variability of number of previous livebirths), the estimate of inter-rater agreement for number of previous livebirths arrived at in this paper would not apply. Since the population to which these results will be generalized is very similar to that studied, this ICC does apply. Given this low between-subject variance inherent in the population of Ottawa-Carleton, extremely high estimates of intraclass correlation will be difficult to attain.

Despite excellent agreement for virtually all variables, this study confirmed the hypothesis that the existing PNOB database does not have good sensitivity or positive predictive value in detecting stillborn babies. The problem of liveborn babies being incorrectly coded as stillborn rests with the miscoding of the PNOB forms by the labour and delivery staff rather than the incorrect entering of vital status into the existing PNOB database by health department staff. The problem of false negatives (calling stillbirths livebirths), however, appears to lie with the

incorrect entering of vital status information into the existing PNOB database by the health department staff.

While a number of stillbirths are being mistakenly coded on the PNOB forms as livebirths, in the context of the entire database, this will likely have little effect on studies of livebirths. Furthermore, the livebirths that are being mistakenly coded as stillbirths do not appear to be smaller or less mature than other livebirths. (There is a small, but nonsignificant trend on a *post-hoc* analysis toward false-positives having a lower gestational age: p=0.108.) This finding provides assurance that the smaller babies at higher risk for neonatal death are not being systematically miscoded as stillbirths and, therefore, not biasing the livebirth pool to bigger, more mature babies. The newly-entered database provides sound evidence that the false positives are not systematically different than true livebirths with respect to gestational age (p=0.502) and birthweight (p=0.676), indicating that with care at the data entry stage, the PNOB form can be relied upon to provide high quality data for livebirths.

Having a database which can provide accurate data for livebirths is an important accomplishment for the health department since many public health activities involve solely the study of livebirths for issues related to service and community health. One of the public health activities for which this system would not be appropriate is a sentinel surveillance system for stillbirths. It is also not useful in the study of adverse fetal outcomes and their risk factors because the smallest and most immature stillbirths are being underreported. Stillbirths that

are missed are vastly different from true stillbirths (according to *post-hoc* tests). The 95% confidence intervals indicate that the mean gestational age of stillborn babies reported to be livebirths is between 3.2 and 8.2 weeks greater than the mean gestational age of the true stillbirths. The confidence intervals also indicate that the mean birthweight of the overlooked stillbirths is between 789 and 1608 grams greater than the mean birthweight of true stillbirths. Unintentionally excluding these older, heavier stillbirths would bias the stillbirth population to those who were a little smaller and less mature at the time of their demise. Studies of stillbirths would be better suited to the provincial level where the PNOB can be corroborated with the "Report of Stillbirth or Neonatal Death" and the "Medical Certificate of Stillbirth".

In summary, the existing database provides an adequate source of information on a number of variables relevant to the needs of the Ottawa-Carleton Health Department, including gestational age, mother's age, birthweight, number of previous livebirths, sex of baby and the first half of the postal code. Even though the newly-entered database provides more accurate information than the existing database regarding whether the baby was liveborn or stillborn, it is discouraging that positive predictive value could only be improved from 71% to 76%. Adequate data cannot be obtained for whether the baby was born alive.

Two hundred twenty four records (0.58%) in the existing PNOB database were missing whether the baby was born alive or still. All but two of these records belonged to babies born in 1989 and 1990. Care will need to be taken to ensure that

this trend toward more complete reporting of birth outcomes continues in the future.

6.5 Suggestions for Improvement of the PNOB Database

Based upon qualitative observations made during the data collection phase and the quantitative conclusions of this study, several recommendations to improve the quality of the PNOB database can be made. On the PNOB form, itself, the checkboxes "yes" and "no" for the question "was baby born alive" are only 4 millimetres square with a space of 6 millimetres between them. During data collection, I often found it difficult to decide to which box the "X" belonged. If the boxes were bigger and further apart, some of the problem may be solved. A system of double-checks could be put into place by requiring that the one and five minute Apgar scores be reported; if a baby is truly a stillborn, the Apgar scores would, by definition, be "0" and "0". This additional information would represent a negligible increase in the workload for the hospitals since Apgar scores are routinely assessed and recorded in the caseroom logbooks. A second type of double-check could involve actually printing the word "stillbirth" on a given line on the PNOB form *in addition* to the standard checkboxes. This action is more deliberate than just checking off a box and may prompt greater accuracy in reporting whether the baby was stillborn. I do not believe that having a line upon which to print "live" or "still" exclusively would be wise, but rather using both formats simultaneously with each being the other's double-check.

Although the reporting of birthweight was generally quite reliable, this study did identify a problem with the health department's data-entry of the weight of tiny babies. The study shows that the weight of these very low birthweight babies is being entered as though it was recorded on the PNOB form in kilograms rather than grams; that is to say, for instance, that if a baby's true birthweight was 320g, it was frequently entered incorrectly as 3.20kg. Fortunately, the solution to this problem is fairly straight-forward. By using the Canadian birthweight percentile data already published by Arbuckle, Wilkins and Sherman [37], the data-entry package could be programmed to prompt for verification of birthweights beyond the 95th, 97th or 99th percentile based on gestational age. (Although the data of Arbuckle, Wilkins and Sherman is presented according to sex of the baby and multiplicity of the birth, these would not necessary for the health department's requirements.) For example, with a fetus of 28 weeks gestation, a system such as this based on the 99th percentile would detect any birthweight in excess of 1900g and prompt for verification. Empirically, most mistakes in this study were made among babies under 28 weeks gestation, and the discrepancy between a birthweight with a misplaced decimal and the 99th percentile would be even more obvious. A system such as this could be particularly important in the case of a data-entry clerk who may be unfamiliar with expected gestational age and birthweight patterns.

Further suggestions for improving data quality at the health department level include double-entering a sample of data to monitor the accuracy of data-entry. One interesting strategy adopted by one Ontario health department

epidemiologist is to routinely send PNOB forms with missing or illegible data back to the hospital from which it came to have the information completed [38].

Even though some of the inaccuracy was due to clerical error at the regional health department level, a sizeable proportion was not, implying that at least some of the responsibility for this problem lies with the hospitals. Regular and timely feedback regarding information derived from the PNOB database with the nursing and clerical staff may lead to enhanced data quality. Regular communication with researchers utilizing the database may demonstrate to staff that worthwhile research is done with the PNOB forms and at the same time encourage the accurate completion of the PNOB forms by the labour and delivery staff.

6.6 Further Research

Further research is needed to enable us to fine-tune the reliability of the PNOB database since its worth has already been demonstrated. I believe the next step lies in randomized controlled trials of strategies to improve data quality, such as PNOB form design, different methods of feedback to labour and delivery staff (e.g. semi-annual catchy newsletter vs. inservice seminar vs. no intervention). A wise outcome measure for studies of this type would be intraclass correlation coefficient, sensitivity and positive predictive value.

6.7 Generalization to Other Health Departments

To date this research has been concerned with the reliability of a database derived from the Physician Notices of Live Birth or Still Birth in the Regional

Municipality of Ottawa-Carleton. Ottawa-Carleton is only one of a number of provincial health departments to make use of the information for research purposes. At the moment, the extent of the problem with the recording of vital status is unknown. Other health departments that routinely enter and use data contained on the PNOB form do have reason to be concerned by the findings of this study and an assessment of their own databases, in particular, the assessment of vital status, is likely a worthwhile endeavour. Any departure from perfect agreement regarding whether the baby was born alive should merit a closer investigation within the given health unit.

7. BIBLIOGRAPHY

1. Wadhera S, Millar WJ, Nimrod C. Trends in Canadian birth weights, 1971 to 1989. *Can Fam Physician* 1992;38:1632-1640.

2. Millar WJ, Strachan J, Wadhera S. Trends in low birthweight Canada 1971 to 1989. *Health Reports* 1991;3:311-325.

3. Klaucke DN. Evaluating public health surveillance systems. In: Halperin W, Baker ELJ. eds. *Public health surveillance*. New York: Van Nostrand Reinhold, 1992:26-41.

4. Campbell K. Neonatal mortality and other measures. 1993;(Abstract)

5. Floyd L, Lavoie M, Terry JS. The status of birth certificate information in Georgia. *Journal of the Medical Association of Georgia* 1981;70:871-873.

6. Ferrara A, Atakent YS, Levinson B. Discrepancies in birth weights between hospital records and health department data for low birth weight infants in New York City. *Public Health Reports* 1988;103:472-478.

7. Teperi J. Multi method approach to the assessment of data quality in the Finnish medical birth registry. *Journal of Epidemiology and Community Health* 1993;47:242-247.

8. David RJ. The quality and completeness of birthweight and gestational age data in computerized birth files. *American Journal of Public Health* 1994;70:964-973.

9. Buescher PA, Taylor KP, Davis MH, et al. The quality of the new birth certificate data: a validation study in North Carolina. *American Journal of Public Health* 1993;83:1163-1165.

10. Pipel D, Fraser GM, Kosecoff J, et al. Validation of a centrally maintained computerized hospital database: comparison with operating room logbooks. *Isr J Med Sci* 1993;29:287-291.

11. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *The Statistician* 1983;32:307-317.

12. Kramer MS, Feinstein AR. Clinical biostatistics: LIV. the biostatistics of concordance. *Clin Pharmacol Ther* 1981;29:111-123.

13. Cohen JA. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20:37-46.

14. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;70:213-229.

15. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 1973;33:613-619.

16. Cicchetti DV, Allison T. Assessing the reliability of scoring EEG sleep records: an improved method. *Proceedings and Journal of the Electrophysiological Technologists' Association* 1973;20:92-102.

17. Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol* 1987;126:161-169.

18. Bartko JJ. General methodology II: measures of agreement: a single procedure. *Statistics in Medicine* 1994;13:737-745.

19. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 1979;86:420-427.

20. Spence-Laschinger HK. Intraclass correlations as estimates of interrater reliability in nursing research. *Western Journal of Nursing Research* 1992;14:246-251.

21. Fleiss JL. Reliability of measurement. In: Fleiss JL. ed. *The design and analysis of clinical experiments*. New York: John Wiley and Sons, 1986:1-32.

22. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures -- statistics and strategies for evaluation. *Controlled Clinical Trials* 1991;12:142S-158S.

23. Coughlin SS, Pickle LW. Sensitivity and specificity-like measures of the validity of a diagnostic test that are corrected for chance agreement. *Epidemiol* 1992;3:178-181.

24. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Computers in Biology and Medicine* 1990;20:337-340.

25. Tanner MA, Young MA. Modeling agreement among raters. *J Am Stat Assoc* 1985;80:175-181.

26. Agresti A. A model for agreement between ratings on an ordinal scale. *Biometrics* 1988;44:539-548.

27. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45:255-268.

28. Bartko JJ. The intraclass correlation coefficient as a measure of Reliability. *Psychol Rep* 1966;19:3-11.

29. Kanji GK. *100 Statistical tests*. London: SAGE Publications. 1993:

30. Brown MB, Forsythe AB. Robust tests for the equality of variances. *JASA* 1974;69:364-367.

31. Levene H. Robust tests for the equality of variance. In: Olkin I. ed. *Contributions to probability and statistics*. Palo Alto: Stanford University Press. 1960:

32. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.

33. Dunn G. *Design and analysis of reliability studies: the statistical evaluation of measurement errors*. New York: Oxford University Press. 1989:

34. Baldereschi M, Amato MP, Nencini P, et al. Cross-national interrater agreement on the clinical diagnostic criteria for dementia. *Neurology* 1994;44:239-242.

35. Krieger N, Hiatt RA, Sagebiel RW, et al. Inter-observer variability among pathologists' evaluation of malignant melanoma: effects upon an analytic study. *J Clin Epidemiol* 1994;47:897-902.

36. Dedrick RF, Davis WK. Issues in use of intraclass correlation. *Diabet Care* 1990;13:1133-1134.

37. Arbuckle TE, Wilkins R, Sherman GJ. Birth weight percentiles by gestational age in Canada. *Obstet Gynecol* 1993;81:39-48.

38. McGurran J. 1993; (Personal Communication)

8. APPENDIX

WILLIAM AND CATHERINE BOOTH
FOUNDERS

EVA BURROWS
GENERAL

WESLEY HARRIS
TERRITORIAL COMMANDER

The Salvation Army
Grace General Hospital

1156 WELLINGTON ST., OTTAWA, ONTARIO, CANADA K1Y 2Z4

PHONE (613) 728-4611
FAX (613) 724-4644

February 19, 1993

Beth A. Henderson
M.Sc. Candidate
Department of Epidemiology & Community Medicine
University of Ottawa
451 Smyth Road
Ottawa, Ontario
K1H 8M5

Re: Physician Notices of Birth Database Study

Dear Ms. Henderson:

I would like to thank you for sending us a summary of your findings with respect to the study entitled "Investigation into the Distribution and Causes of Excess Stillbirths in the Regional Municipality of Ottawa-Carleton 1991".

With respect to your request to include the Grace General Hospital in your new study on the accuracy of the PNOB database, I would like to confirm that we would be pleased to assist you with this project. I have spoken with the persons responsible for the areas involved and they are in agreement.

If you have any further questions, please contact Ruth Benwell, Nursing Director/Maternity Care, whom you will find most helpful.

Thank you.

Sincerely,

Malcolm D. Robinson
(Captain)
Executive Director

em

HÔPITAL GÉNÉRAL D'OTTAWA OTTAWA GENERAL HOSPITAL

April 16, 1993

Dr. Paula J. Stewart
Assistant Clinical Professor
Department of Epidemiology and Community Medicine
Health Sciences Building

Dear Dr. Stewart,

RE: OGH-93-029 - Validation Study of the Physician Notices of Birth Database in the Regional Municipality of Ottawa-Carleton

The above protocol has been reviewed by the Human Experimental Procedures Committee at the Ottawa General Hospital. Basically, this is a review of information present in a hospital database.

This type of research generally does not need to be reviewed by the Committee and the Committee approves the conduct of this study.

A member of staff of the Ottawa General Hospital should assume responsibility for the confidentiality of the information and for supervising the student when working at the hospital. It is assumed that this will be Dr. Carl Nimrod. The principles of confidentiality should apply to all information obtained in this way.

Yours sincerely,

G.D. Goss, M.D., F.C.P.(SA), F.R.C.P.C.
Chairman
Human Experimental Procedures Committee

GG/lb

cc: Beth Henderson

LOEB MEDICAL RESEARCH INSTITUTE
INSTITUT DE RECHERCHE MÉDICALE LOEB

OTTAWA CIVIC HOSPITAL
HÔPITAL CIVIC D'OTTAWA

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

February 22, 1993

Ms. B. Henderson
Department of Epidemiology
& Community Medicine
University of Ottawa
451 Smyth Road
Ottawa, Ontario
K1H 8M5

Dear Ms. Henderson:

Re: Protocol#93-028C A Validation Study of the Physician Notices of Birth Database in the Regional Municipality of Ottawa-Carleton

Thank you for your letter of February 8, 1993. Chart reviews as such do not require formal Research Ethics Committee review. The only concern in this study is patient confidentiality, particularly in that you propose to use the mother's name.

Piease work out with Ms. Shirley Pilon, the Director of Health Records, the maintenance of patient confidentiality, and report back to me. Dr. Brian Ivey, as a member of the Ottawa Civic Hospital staff will be responsible for the conduct of the study in this Hospital.

Yours sincerely,

Raphael Saginur, M.D.
Chairman
Research Ethics Committee

RS/mal

c.c. Ms. Shirley Pilon
Dr. Brian Ivey

1053 Carling Avenue, Ottawa, Ontario K1Y 4E9
Tel: (613) 761-4395 Fax: (613) 761-9151

THE RIVERSIDE HOSPITAL
OF OTTAWA
1967 Riverside Drive
Ottawa, Ontario K1H 7W9
(613) 738-7100
Fax (613) 738-8522

February 12, 1993

Ms. Beth A. Henderson
Department of Epidemiology & Community Medicine
University of Ottawa
451 Smyth Road
Ottawa, Ontario
K1H 8M5

Dear Ms. Henderson:

Thank you for your letter of February 8th in which you report the results of your study into the distribution and causes of excess stillbirths in the Regional Municipality of Ottawa-Carleton in 1991. I am pleased to hear that there was an error in the numbers and that the actual stillbirth rate is considerably lower than what was originally reported.

We are agreeable to your alterating the study now to determine the accuracy of the physician notice of birth database with respect to other pieces of information such as the hospital case room log book and the mother's medical record.

Best wishes for a successful study.

Yours sincerely

G. Brian Doyle
Executive Director

/bp

c.c. Mrs. S. Timmins, Nursing Director - Obstetrics
Ms. D. Frenette, Director, Health Records
Dr. G.L. Liberty, Chief of Staff