

The final version of this paper was published in NSW Public Health Bulletin 2013;
24(2):65-69

Are pregnancy outcomes associated with risk factor reporting in routinely collected perinatal data?

Amanda J. Ampt^{A,C}, Jane B. Ford^A, Lee K. Taylor^B and Christine L. Roberts^A

^AKolling Institute of Medical Research, The University of Sydney

^BCentre for Epidemiology and Evidence, NSW Ministry of Health

^CCorresponding author. Email: amanda.ampt@sydney.edu.au

Abstract

Aim: The aims of our study were to assess reporting characteristics of commonly dichotomised pregnancy outcomes (e.g. preterm/term birth); and to investigate whether behaviours (e.g. smoking), medical conditions (e.g. diabetes) or interventions (e.g. induction) were reported differently by pregnancy outcomes.

Methods: Further analysis of a previous validation study was undertaken, in which 1680 perinatal records were compared with data extracted from medical records.

Continuous and polytomous variables were dichotomised, and risk factor reporting was assessed within the dichotomised outcome groups. Agreement, Kappa, sensitivity and positive predictive value calculations were undertaken. **Results:**

Gestational age, birthweight, Apgar scores, perineal trauma, regional analgesia and baby discharge status (live birth/stillbirth) were reported with high accuracy and reliability when dichotomised (Kappa values 0.95–1.00, sensitivities 94.7%–100.0%).

Although not statistically significant, there were trends for hypertension, infant resuscitation and instrumental birth to be more accurately reported among births with adverse outcomes. In contrast, smoking ascertainment tended to be poorer among preterm births or when babies were <2500 g. **Conclusion:** Dichotomising variables collected as continuous or polytomous variables in birth data results in accurate and well ascertained data items. There is no evidence of systematic differential reporting of risk factors.

Population level data are well suited to studies evaluating health care. With the risk of sampling bias removed, estimation of incidence and prevalence rates can be made, allowing for description of the total burden of a particular disease or outcome, analysis of risk factors and trends, as well as identification of health inequalities and estimation of health costs.¹⁻² Accurate conclusions from such analyses rely on high quality data that truly represent the population experience. Assessment of data quality (completeness and accuracy) is typically undertaken by a validation study, in which data from a sample of records from the population dataset are compared to a highly reliable and accurate source of data ('the gold standard') for the corresponding records. The accuracy and reliability of individual data items are typically reported.³⁻⁴

The variables in perinatal population data can be continuous (e.g. gestational age), nominal (e.g. mode of delivery) and ordinal (e.g. first, second, third or fourth degree perineal tears), with validation of such variables typically reporting percent agreement and Kappa statistics. However, these types of variables are frequently dichotomised in analyses (e.g. preterm birth, caesarean section, or third–fourth degree tears),⁵⁻⁶ but little assessment has been undertaken into the accuracy and reliability of their dichotomised form.

Differential reporting in population health data occurs when a variable is reported with different accuracy and reliability amongst different strata of another variable. This can introduce systematic bias, leading to under or over estimation of risk factor effects.⁷ For example, if smoking is more likely to be reported when an infant is growth restricted, this could result in the effect of smoking on growth restriction being over-estimated. Different accuracy and reliability statistics have been demonstrated for reporting of both pregnancy hypertension and induction depending on the mode of delivery,^{2,8} and for hypertension depending on the gestation.⁹ However, we are only aware of one other study that has investigated whether the occurrence of adverse infant or maternal outcomes might result in increased reporting of established risk factors for these outcomes.⁹

With little published research reporting on the dichotomised form of population data, the aims of our study were therefore twofold: a) to assess reporting characteristics of

commonly dichotomised pregnancy outcomes; and b) to investigate whether behaviours (e.g. smoking), medical conditions (e.g. diabetes) or interventions (e.g. induction) were reported differently by outcomes.

Methods

This study involved further analysis of data from a previous validation study of the New South Wales (NSW) Perinatal Data Collection (PDC). The PDC (formerly known as the NSW Midwives Data Collection) is a population-based statutory surveillance system and serves as a primary source of information about pregnancy and birth outcomes in NSW for all births ≥ 20 weeks gestation or ≥ 400 g birthweight. The original study is described in detail elsewhere.³ Briefly, randomly selected records from the PDC (referred to as the 'PDC sample') were compared with 'gold standard' data extracted from the corresponding patients' medical records (referred to as the 'validation data'). The PDC sample comprised 1680 records representing 2% of the state's births from 98 hospitals around NSW. Information from the medical records of the selected sample of women was extracted by experienced health managers without reference to information contained in the PDC sample. The data item with highest frequency of missing values was Apgar5, which was missing from six records in the PDC sample (0.36%), and from nine records in the validation data (0.54%).

We first assessed the accuracy and reliability of continuous and polytomous data items when examined as dichotomous outcomes. We chose data items that are commonly dichotomised including: gestational age (<37 weeks gestation, ≥ 37 weeks gestation); birthweight (<2500 g, ≥ 2500 g; <4000 g, ≥ 4000 g); Apgar score at 1 minute (Apgar1 <4, Apgar1 ≥ 4) and Apgar score at 5 minutes (Apgar5 <7, Apgar5 ≥ 7); epidural, caudal, pudendal or spinal analgesia (regional analgesia yes/no); second, third or fourth degree tears and/or episiotomy (perineal trauma yes/no); and baby discharge status (stillbirth/live birth).

Next we examined potential differential reporting of risk factors by determining the accuracy and reliability of risk factor reporting in the PDC sample for different pregnancy outcomes. Specifically, we hypothesised that the following established

risk factors may be more likely to be reported in the presence of an associated outcome:

- smoking when infants were small or preterm¹⁰
- maternal hypertension among preterm births¹¹
- maternal diabetes when infants were large¹²
- instrumental birth (forceps or vacuum) among women who experienced perineal trauma¹³
- induction among women who required regional analgesia¹⁴
- infant resuscitation (intermittent positive pressure respiration, bag and mask or intubation, or external cardiac massage and ventilation) when Apgar5 <7.

Analysis

Using the validation data as the 'gold standard', the reliability and accuracy of PDC reporting was determined by calculating the sensitivity, specificity, positive predictive value (PPV), negative predictive value, percent agreement and Cohen's Kappa statistic. These reporting characteristics were determined first for the commonly dichotomised variables and then for risk factors in the hypothesised outcome strata. When a record was missing a data item, it was excluded from the relevant analysis. We assessed the homogeneity of risk factor reporting across the dichotomised outcome strata by the Breslow-Day test, with Zelan adjustment where cell counts were less than five.

All analyses included the associated 95% exact binomial confidence intervals. These are not presented in the tables, but are available from the authors on request. All analyses were undertaken using SAS (version 9.2, SAS Institute, Cary, NC, USA).

Results

Of the 1680 records in the original validation study, 1678 were available for analysis. Characteristics of the PDC sample were representative of all births in NSW (Table 1).

Commonly dichotomised pregnancy outcomes (preterm birth, low and high birthweight, Apgar scores, perineal trauma, regional analgesia and stillbirth) as reported in the PDC had excellent levels of agreement, and high levels of ascertainment (sensitivities >94%) and accuracy (PPVs >96%) (Table 2).

The results of the investigation into differential reporting are presented in Table 3. PPVs were high, with 11 of 14 individual analyses $\geq 90\%$, but with inconsistencies in direction among outcome groups for each risk factor. There was more variability in the sensitivities, ranging from 66% for reporting of infant resuscitation amongst the group whose Apgar5 was ≥ 7 , to 99% for reporting of inductions with no regional analgesia. In total, six out of the 14 sensitivity measures were $\geq 90\%$. There was no overall pattern suggestive of significantly better reporting in the presence of an adverse outcome. Although there was a trend to higher ascertainment of infant resuscitation among infants with low Apgar5 (sensitivities of 86% vs 66%), of instrumental birth among women with perineal trauma (97% vs 88%), and of hypertension among preterm birth (77% vs 67%), the reverse was true for ascertainment of smoking both among preterm birth (82% vs 90%) and among small infants <2500 g (83% vs 90%). There were no statistically significant differences in reporting across strata, with Breslow-Day p values all >0.05.

Discussion

Our study demonstrated that dichotomising perinatal outcome data into categories that are typically reported in population health research^{5,6} resulted in high levels of ascertainment and accuracy. With all sensitivities $\geq 94.7\%$ and all PPVs $\geq 96.1\%$, reassurance is provided for the use of these data items in their dichotomised form where necessary for comparison to other findings or due to sample size constraints. There was no evidence of overall systematic bias in risk factor reporting across one strata of outcome (the adverse group) compared to the other. This study adds new information on dichotomised reporting characteristics and differential reporting. Strengths of this study include the highly representative nature of the PDC sample, the use of six measures of accuracy and reliability, and the small percentage of missing data. Limitations included small numbers in some outcome strata. Lack of statistical significance may thus have been a result of underpowering for some categories.

Most risk factors were fairly well ascertained regardless of outcome strata, with the exception of hypertension and infant resuscitation among the groups that did not have an adverse outcome. Reliability, as measured by PPV, was lowest amongst diabetes reporting for the adverse group, but numbers were small. There was a non-significant trend towards higher ascertainment of hypertension, instrumental birth and infant resuscitation in the adverse groups. It is recognised that these trends could become significant with larger sample sizes, and may introduce biases in research.

The non-significant trends in differential reporting were not always in the hypothesised direction. Ascertainment for behaviour (smoking) was lower amongst the adverse outcome group; while ascertainment and for some interventions (instrumental birth and infant resuscitation) for hypertension was higher in the adverse outcome groups. This latter finding is consistent with another study that identified a trend towards increased ascertainment of hypertension among women who delivered prematurely or suffered a morbidity.⁹ While it might be expected that some risk factors which may be reported earlier in pregnancy (e.g. smoking, hypertension) may not have the same impact on reporting as risks occurring closer to delivery (e.g. induction, infant resuscitation), there were no differences in ascertainment or accuracy for these factors. Overall our findings demonstrate the randomness of reporting errors and no evidence of systematic bias due to differential reporting by outcome.

This study used data collected in 1998 as this was the last time the PDC was validated against medical records. Some changes to the recording of information are likely to have occurred with the advent of electronic systems, but the majority of PDC recording still occurs at the time of the birth admission, and hence accuracy of variables once dichotomised and of maternal or infant outcome risk factor reporting are unlikely to have been affected.

Conclusion

Our findings demonstrate that dichotomised perinatal variables have high levels of accuracy and reliability when compared with medical records. In addition, ascertainment of risk factors show some non-significant differences within different

pregnancy outcome groups; however reporting errors are random in their direction, revealing that there is no evidence of systematic bias.

References

1. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttman A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol* 2011; 64(8): 821-9.
2. Roberts CL, Bell JC, Ford JB, Morris JM. Monitoring the quality of maternity care: How well are labour and delivery events reported in population health data? *Paediatr Perinat Epidemiol* 2009; 23(2): 144-52.
3. Taylor L, Pym M, Bajuk B, Sutton L, Travis S, Banks C. Validation study: NSW Midwives Data Collection 1998. *N S W Public Health Bull Supplementary Series* 2000; 9(1): 97-9.
4. Lain SJ, Hadfield RM, Raynes-Greenow CH, Ford JB, Mealing NM, Algert CS et al. Quality of data in perinatal population health databases: a systematic review. *Med Care* 2012; 50(4): e7-20.
5. Ford JB, Roberts CL, Simpson JM, Vaughan J, Cameron CA. Increased postpartum hemorrhage rates in Australia. *Int J Gynecol Obstet* 2007; 98(3): 237-43.
6. Roberts CL, Algert CS, Morris JM, Ford JB, Henderson-Smart DJ. Hypertensive disorders in pregnancy: A population-based study. *Med J Aust* 2005; 182(7): 332-5.
7. Schoendorf KC, Branum AM. The use of United States vital statistics in perinatal and obstetric research. *Am J Obstetr Gynecol* 2006; 194(4): 911-5.
8. Roberts C, Lain S, Hadfield R. Quality of population health data reporting by mode of delivery. *Birth* 2007; 34(3): 274-5.
9. Roberts CL, Bell JC, Ford JB, Hadfield RM, Algert CS, Morris JM. The accuracy of reporting of the hypertensive disorders of pregnancy in population health data. *Hypertens Pregnancy* 2008; 27(3): 285-97.
10. Cnattingius S. The epidemiology of smoking during pregnancy: smoking prevalence, maternal characteristics, and pregnancy outcomes. *Nicotine Tob Res* 2004; 6 Suppl 2: S125-40.

11. Rosenberg TJ, Garbers S, Lipkind H, Chiasson MA. Maternal obesity and diabetes as risk factors for adverse pregnancy outcomes: differences among 4 racial/ethnic groups. *Am J Public Health* 2005; 95(9): 1545-51.
12. Makgoba M, Savvidou MD, Steer PJ. The effect of maternal characteristics and gestational diabetes on birthweight. *BJOG* 2012; 119(9): 1091-7.
13. Mikolajczyk RT, Zhang J, Troendle J, Chan L. Risk factors for birth canal lacerations in primiparous women. *Am J Perinatol* 2008; 25(5): 259-64.
14. Maslow AS, Sweeny AL. Elective induction of labor as a risk factor for cesarean delivery among low-risk women at term. *Obstet Gynecol* 2000; 95(6 Pt I): 917-22.
15. Bell JC, Ford JB, Cameron CA, Roberts CL. The accuracy of population health data for monitoring trends and outcomes among women with diabetes in pregnancy. *Diabetes Res Clin Pract* 2008; 81(1): 105-9.

Table 1. Comparison of Perinatal Data Collection (PDC) sample with all NSW births, 1998

Characteristics	Sample of PDC data ^a <i>n</i> (%)	NSW births ^{b,c} <i>n</i> (%)
Labour onset		
Spontaneous	1067 (63.6)	56 283 (65.2)
Induced	446 (26.6)	20 898 (24.2)
No labour	164 (9.8)	9103 (10.6)
Mode of delivery		
Normal vaginal delivery	1174 (70.0)	59 398 (68.8)
Forceps	91 (5.4)	4545 (5.3)
Vacuum	81 (4.8)	4526 (5.2)
Vaginal breech	18 (1.1)	1050 (1.2)
Caesarean section – planned (no labour)	164 (9.8)	9103 (10.6)
Caesarean section – intra-partum	149 (8.9)	7654 (8.9)
Any smoking during pregnancy	333 (19.9)	17 066 (19.8)
Maternal medical conditions		
Gestational diabetes or diabetes mellitus	67 (4.0)	3451 (4.0)
Any hypertension	119 (7.1)	6202 (7.2)
Gestational age <37 completed weeks	102 (6.1)	5953 (6.9)
Infants with Apgar1 <4	55 (3.3)	2878 (3.3)
Infants with Apgar5 <7	37 (2.2)	2191 (2.6)
Infant birthweight <2500 g	90 (5.4)	5299 (6.1)
Infant birthweight ≥4000 g	201 (12.0)	10 404 (12.1)
Perineal status		

Intact	732 (43.6)	38 581 (44.7)
1 st degree tear or graze	335 (20.0)	16 300 (18.9)
2 nd degree tear	294 (17.5)	14 926 (17.3)
3 rd degree tear	10 (0.6)	608 (0.7)
4 th degree tear	1 (<0.1)	48 (<0.1)
Episiotomy	247 (14.7)	12 633 (14.6)
Both tear and episiotomy	24 (1.4)	1249 (1.5)
Regional analgesia	554 (33.0)	27 623 (32.0)
Induction	446 (26.6)	20 898 (24.2)
Infant resuscitation	127 (7.6)	6565 (7.6)
Baby discharge status		
Discharged	1558 (92.3)	80 517 (93.3)
Stillbirth	11 (0.7)	595 (0.7)
Neonatal death	2 (0.1)	200 (0.2)
Transferred	106 (6.3)	4859 (5.6)
Transferred and died	0 (0.0)	16 (<0.1)

^aSample size = 1678.

^bSample size = 86 305.

^cSource: 1998 Perinatal Data Collection.

Table 2. Agreement, ascertainment and accuracy of dichotomised pregnancy outcome variables reported in the Perinatal Data Collection (PDC) compared with validated data

Outcome	Cases in PDC sample	Cases in validation data	Agreement %	Kappa	Sensitivity %	Specificity %	PPV %
<37 weeks gestation	102	103	99.5	0.95	95.1	99.7	96.1
<2500 g birthweight	90	90	99.9	0.99	98.9	99.9	98.9
≥4000 g birthweight	201	202	99.9	0.99	99.5	100.0	100.0
Apgar1 <4	55	54	99.9	0.99	100.0	99.9	98.2
Apgar5 <7	37	38	99.8	0.96	94.7	99.9	97.3
Perineal trauma	576	575	97.8	0.95	96.9	98.3	96.7
Regional analgesia	554	561	98.2	0.96	96.6	98.9	97.8
Stillbirth	11	11	100.0	1.00	100.0	100.0	100.0

Records with missing data were excluded.

PDC: Perinatal Data Collection

PPV: positive predictive value

Table 3. Agreement, ascertainment and accuracy of dichotomised pregnancy risk factors reported in the PDC and grouped by pregnancy outcomes compared with validated data

Risk	Outcome <i>n</i> = number in outcome identified by PDC sample	No. with risk identified by PDC sample	No. with risk identified by validation data	Agreement %	Kappa	Sensitivity %	PPV %
Smoking	<37 weeks gestation (<i>n</i> = 100)*	27	33	94.0	0.86	81.8	100.0
	≥37 weeks gestation (<i>n</i> = 1542)	304	322	97.1	0.91	90.4	95.7
	<2500 g (<i>n</i> = 87)*	29	35	93.1	0.85	82.9	100.0
	≥2500 g (<i>n</i> = 1555)	302	320	97.2	0.91	90.3	95.7
Hypertension	<37 weeks gestation (<i>n</i> = 102)*	20	26	94.1	0.83	76.9	100.0
	≥37 weeks gestation (<i>n</i> = 1576)	99	132	96.5	0.74	66.7	88.9
Diabetes	≥4000 g (<i>n</i> = 201)*	11	10	98.5	0.85	90.0	81.8
	<4000 g (<i>n</i> = 1473)	56	59	99.1	0.88	86.4	91.1

Instrumental birth	Perineal trauma (<i>n</i> = 575)*	144	146	98.6	0.96	96.6	97.9
	No perineal trauma (<i>n</i> = 1011)	28	32	99.6	0.93	87.5	100.0
Induction	Regional analgesia (<i>n</i> = 553)*	170	173	97.3	0.94	94.8	96.5
	No regional analgesia (<i>n</i> = 1123)	275	267	98.6	0.96	98.5	95.6
Infant resuscitation	Apgar5 <7 (<i>n</i> = 37)*	20	21	86.5	0.73	85.7	90.0
	Apgar5 = 7–10 (<i>n</i> = 1628)	106	139	96.3	0.73	66.2	86.8

Records with missing data were excluded.

PPV: positive predictive value

*Outcome known to be associated with risk factor (adverse outcome)