OPEN ACCESS

Variation in hospital morbidities in an Australian neonatal intensive care unit network

Mohamed E Abdel-Latif ,[1,2,3] Oyelola Adegboye ,[4,5] Gen Nowak ,[6] Faiz Elfaki ,[7] Barbara Bajuk ,[8] Kathryn Glass ,[9] David Harley [9,10]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/archdischild-2022-324940).

For numbered affiliations see end of article.

Correspondence to
Professor Mohamed E Abdel-Latif, Department of Neonatology, Centenary Hospital for Women and Children, Canberra Hospital, Garran, Canberra, ACT, Australia; abdel-Latif.Mohamed@act.gov.au

Received 24 September 2022
Accepted 20 December 2022
Published Online First 2 January 2023

ABSTRACT

Objective There is an expectation among the public and within the profession that the performance and outcome of neonatal intensive care units (NICUs) should be comparable between centres with a similar setting. This study aims to benchmark and audit performance variation in a regional Australian network of eight NICUs.

Design Cohort study using prospectively collected data.

Setting All eight perinatal centres in New South Wales and the Australian Capital Territory, Australia.

Patients All live-born infants born between 23^{+0} and 31^{+6} weeks gestation admitted to one of the tertiary perinatal centres from 2007 to 2020 (n=12 608).

Main outcome measures Early and late confirmed sepsis, intraventricular haemorrhage, medically and surgically treated patent ductus arteriosus, chronic lung disease (CLD), postnatal steroid for CLD, necrotising enterocolitis, retinopathy of prematurity (ROP), surgery for ROP, hospital mortality and home oxygen.

Results NICUs showed variations in maternal and neonatal characteristics and resources. The unadjusted funnel plots for neonatal outcomes showed apparent variation with multiple centres outside the 99.8% control limits of the network values. The hierarchical model-based risk-adjustment accounting for differences in patient characteristics showed that discharged home with oxygen is the only outcome above the 99.8% control limits.

Conclusions Hierarchical model-based risk-adjusted estimates of morbidity rates plotted on funnel plots provide a robust and straightforward visual graphical tool for presenting variations in outcome performance to detect aberrations in healthcare delivery and guide timely intervention. We propose using hierarchical model-based risk adjustment and funnel plots in real or near real-time to detect aberrations and start timely intervention.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The characteristics of infants admitted to neonatal intensive care units differ, so comparing unadjusted morbidity rates should be avoided.

WHAT THIS STUDY ADDS

⇒ Variations in hospital morbidities estimates plotted on funnel plots provide a powerful visual graphical tool for presenting quality performance data.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Prospective and timely hierarchical model-based risk adjustment for centre-to-centre variation in morbidities is a useful method to inform hospitals to readily appraise their practices and start timely intervention.

BACKGROUND

There is an increasing need for accurate patient quality, safety and hospital performance measures in healthcare. The public and health professionals expect the performance and outcome of neonatal services should be comparable between centres of similar settings. However, there are challenges in meeting these expectations.[1]

Centre-to-centre (CTC) variation in neonatal health outcomes may result from patient characteristic differences (intrinsic factors) rather than centre or service differences (extrinsic factors). Outcome variations related to intrinsic and extrinsic factors are called common-cause and special-cause variations, respectively.[2] Differentiating these sources of variability is critical to service improvement.

We previously published the risk-adjusted CTC variation in mortality rates for preterm infants admitted to New South Wales (NSW) and the Australian Capital Territory (ACT) Neonatal Network (NICUS), Australia.[3] This study aims to report the risk-adjusted CTC variation in major neonatal morbidities for infants born <32 weeks and admitted to the eight tertiary neonatal intensive care units (NICUs) in NSW and the ACT Neonatal Network. We also assess the benefits of adjusting CTC variation for population characteristics using hierarchical model-based risk adjustment.

METHODS

Study design

This prospective population-based cohort study uses data from all tertiary NICUs in well-defined geographic regions of NSW and the ACT.

Study centres and network

A full description of the NSW and the ACT neonatal service organisation and networking, medical and nursing staffing of the collaborating NICUs is available elsewhere.[4–6] In summary, there is a network of 10 units within NSW and the ACT. These include eight perinatal centres (referred to as A to H in this study) and two children's hospitals. Among the perinatal centres, three units (C, G and H) have

To cite: Abdel-Latif ME, Adegboye O, Nowak G, *et al*. *Arch Dis Child Fetal Neonatal Ed* 2023;**108**:F400–F407.

surgical capabilities to operate on conditions like necrotising enterocolitis (NEC) and diaphragmatic hernia.

Coordination of in-utero or ex-utero high-risk referrals among the network is assisted by an intranet bed availability bulletin board, redirecting referrals when any particular unit is full or nearly full to reduce the risk of overloading. The NSW Neonatal and Paediatric Emergency Transport Service (NETS) is an integrated centralised transport service covering NSW and the ACT which coordinates the transfer of sick infants and children from non-tertiary to tertiary centres as well as surgical cases from non-surgical (A, B, D, E, F) to surgical units (C, G, H).[7] Retrieved premature infants (outborns) are preferentially admitted to the eight perinatal centres instead of the two paediatric hospitals.[6 8]

Study participants

The study population comprised all live-born infants born between 23^{+0} and 31^{+6} weeks gestation who were admitted to one of the eight tertiary perinatal centres in NSW and the ACT from 2007 to 2020. As of December 2020, NSW and the ACT had a population of 8 599 314 and approximately 99 752 live births per year.[9]

Data source

Data for this study were obtained from *The Neonatal Intensive Care Units' Data Registry*, which is an ongoing prospective statewide audit of infants admitted to the 10 units (8 perinatal centres and 2 children's hospitals) for one of the following reasons: gestation 22^{+0} to 31^{+6} weeks, birth weight ≤1500 g, assisted ventilation (mechanical ventilation, continuous positive airway pressure, high flow humidified gas), major surgery (opening of a body cavity), insertion of a central line, exchange transfusion for hyperbilirubinemia or therapeutic hypothermia. In this region and according to the National Health and Medical Research Council recommendations, wherever possible preterm birth at <33 weeks should occur in one of the eight perinatal centres.[10] Preterm infants <33 weeks who are born in non-tertiary hospitals are transferred to tertiary centres by NETS.[7]

Data from the two children's hospitals (n=24) were excluded from this study due to the low patient load, as retrieved premature infants (outborns) were preferentially admitted to the eight perinatal centres instead of the two paediatric hospitals.[3] This means that the outcome of premature infants in these two hospitals may be affected because of this policy, as shown elsewhere.[6 8]

Definitions

NICUS data definitions and data accuracy have been described elsewhere.[11–13] Chronic lung disease (CLD) was defined as the requirement for respiratory support at 36 weeks postmenstrual age.[14] Intraventricular haemorrhage (IVH) was graded I–IV by Papile's classification;[15] NEC was staged according to Bell's classification;[16] retinopathy of prematurity (ROP) was staged I–V according to international criteria.[17 18] The details laser therapy for ROP 'Surgery for ROP' can be found elsewhere.[19] Patent ductus arteriosus (PDA) was diagnosed in infants with 'clinical evidence of left to right shunt documented by continuous murmur, hyperdynamic precordium, bounding pulses, wide pulse pressure, congestive heart failure, increased pulmonary vasculature or cardiomegaly by chest x-ray and/or increased oxygen requirement or ECHO evidence of PDA with documentation of left to right ductal shunting'.[20] PDA pharmacological (medical) and surgical management protocols differ between centres.[21] Proven sepsis is defined as a clinical picture consistent with sepsis and either a positive bacterial, viral or fungal culture of blood and/or cerebrospinal fluid occurring less than 48 hours from birth (early) or from 48 hours after birth (late).[22] Infections with coagulase-negative staphylococci, and other potential contaminants, were included only if the baby was considered clinically septic and there was supporting evidence such as raised white cell count or thrombocytopenia.

Primary outcome measures

We selected major neonatal morbidities to be benchmarked. These are shown in online supplemental tables 1–12.

Statistical analysis

Statistical analyses were performed using SAS 9.4 (SAS Institute, Cary, North Carolina, USA). Categorical variables were described by frequencies and percentages, while continuous variables were presented as medians and quartiles (25th and 75th percentile).

We used hierarchical model-based risk adjustment to estimate risk-adjusted CTC variation in major neonatal morbidities for infants born <32 weeks and admitted to the eight NICUs after adjusting for case-mix and the random effect of the centre using the steps below.

First, we used a multivariable Poisson model to control for antenatal and perinatal variables other than *intermediate* variables (eg, CLD, IVH, NEC, ROP), as these may be related, directly or indirectly, to the quality of the hospital management and might thus act as *intermediate* comorbidities through which the effect of the 'hospital' is mediated.[23] For each outcome/morbidity, the model was used to estimate the expected and predicted risk of morbidity for each patient. The level of statistical significance for model selection was based on a 5% level of significance through a stepwise variable selection approach. The calibration of the model was determined by the Hosmer-Lemeshow goodness-of-fit χ^2 test.[24] The ability of the model to discriminate between those who had versus those who did not have the morbidity was summarised using the C-statistic. A C-statistic of 0.5 indicates that the model discriminates no better than chance alone, whereas a value of 1.0 indicates perfect discrimination.[25 26]

Second, risk-adjusted standardised ratios (RAR) for each outcome/morbidity within each hospital were then computed as the ratio of predicted-to-expected hospital morbidity multiplied by the network's observed rate.

Third, a risk-adjusted standardised incidence ratio (RSIR) was produced by dividing the hospital crude rate by RAR. The 95% CI for RSIR were computed using Monte Carlo simulation as described elsewhere.[27]

Last, we used funnel plots to provide a visual indication to differentiate between common-cause and special-cause variation in risk-adjusted hospital morbidity among NICUs.[28] These plots indicate whether morbidity rates in a NICU differ significantly from the average network rate, assuming only random sampling variation influences the NICU's rate. A solid horizontal line represents the overall network morbidity rate while the 95% (2.5th percentile represents the lower control limit and 97.5th percentile represents the upper control limit) and 99.8% (0.1th percentile represents the lower control limit, and 99.9th percentile represents the upper control limit) control limits are represented by the curved dotted lines. Assuming differences arise from random sampling variation alone, the chance of the hospital being within limits is 95% for the inner funnels and 99.8% for the outer funnel.

RESULTS

Study population

A total of 12 608 live-born infants $<32^{+0}$ weeks gestation were admitted to one of the eight tertiary perinatal centres during the study period. The maternal and neonatal characteristics of the study group stratified by admitting hospital are presented in tables 1 and 2. There is variation in patient characteristics (intrinsic factors) between the admitting hospitals. The median (25th and 75th percentile) age of mothers in this study was 31.0 (26.0–35.0) years. The percentage of Indigenous Australians was 7.3% and ranged from 3% (hospital F) to 14.3% (hospital G) (table 1). Three of the eight hospitals have onsite surgical support (table 2). The median (25th and 75th percentile) length of hospital stay among neonatal centres was relatively homogenous, ranging from 51.2 (37.5–71.8) in hospital G to 58.9 (42.9–83.1) in hospital C.

Major neonatal morbidities

Table 3 presents the observed (unadjusted) and hierarchical model-based risk-adjustment estimates of twelve major neonatal morbidities across eight hospitals.

The risk-adjusted estimates presented in table 3 were plotted as funnel plots (online supplemental figures 1–12) to visualise the unadjusted and adjusted estimates. Centres above and below the limits likely indicate special-cause variation, whereas centres within limits indicate common-cause variation. The unadjusted morbidities indicate an apparent variation between the hospitals, especially for NEC, postnatal steroids for CLD, ROP grade III to V, surgery for ROP, hospital mortality and home oxygen. The estimates for these health outcomes were outside the 99.8% control limits of the network values. With 6.35%, hospital H recorded the highest prevalence rate among the network of hospitals in the study for NEC, which is above the upper 99.8% control limit (online supplemental figure 6A). While the overall prevalence of postnatal steroid administration for CLD was 6.96%, hospitals C (9.68%) and F (9.75%) recorded estimates higher than the upper 99.8% control limit of the network prevalence rate (online supplemental figure 8A). There was an overall prevalence of 3.2% for surgery for ROP, with hospital F having a prevalence rate of 4.79%, which is above the upper 99.8% control limit (online supplemental figure 9A).

To accurately estimate various morbidities, it is essential to consider each hospital's patient profile and adjust accordingly. Online supplemental tables 1–12 present the patient characteristics incorporated into the hierarchical model-based risk adjustment for models. After accounting for differences in patient characteristics, the resulting risk-adjusted estimates are presented in table 3 and online supplemental figures 1–12 (panel B). In contrast to unadjusted rates, discharge home with oxygen is the only outcome for which some hospitals lie above the 99.8% control limits after adjusting for hospital-level patient characteristics (online supplemental figure 12B). For hospitals B and G, the estimates were above the upper 99.8% control limit, while for hospitals A, E and F, the estimates were below the lower 99.8% control limit.

DISCUSSION

We have presented benchmarking for major neonatal morbidities in NSW and the ACT. We used hierarchical model-based risk adjustment rather than the traditional logistic regression for case-mix adjustment. Hierarchical model-based risk adjustments have been shown to have the additional benefit of adjusting for centre sample size and clustering issues and avoiding

Table 1 Maternal characteristics of the study group stratified by admitting hospital A to H

Characteristic	Hospital								
	A (n=1413)	B (n=1453)	C (n=1622)	D (n=1640)	E (n=1367)	F (n=1815)	G (n=2149)	H (n=1149)	All (n=12608)
Maternal age, years*	32.0 (28.0–36.0)	33.0 (29.0–36.0)	32.0 (28.0–36.0)	30.0 (25.0–34.0)	29.0 (24.0–33.0)	31.0 (27.0–35.0)	29.0 (25.0–33.0)	31.0 (26.0–34.0)	31.0 (26.0–35.0)
Indigenous Australian	86 (6.1)	47 (3.2)	86 (5.3)	73 (4.5)	167 (12.2)	54 (3.0)	308 (14.3)	96 (8.4)	917 (7.3)
Assisted conception	219 (15.5)	283 (19.5)	265 (16.3)	178 (10.9)	156 (11.4)	300 (16.5)	210 (9.8)	134 (11.7)	1745 (13.8)
Multiple pregnancies	444 (31.4)	447 (30.8)	490 (30.2)	437 (26.7)	336 (24.6)	489 (26.9)	649 (30.2)	326 (28.4)	3618 (28.7)
Chorioamnionitis	271 (19.2)	243 (16.7)	312 (19.2)	240 (14.6)	205 (15.0)	307 (16.9)	318 (14.8)	161 (14.0)	2057 (16.3)
Intrauterine growth restriction	271 (19.2)	243 (16.7)	312 (19.2)	240 (14.6)	205 (15.0)	307 (16.9)	318 (14.8)	161 (14.0)	2057 (16.3)
Any antenatal steroids	1331 (94.2)	1346 (92.6)	1511 (93.2)	1527 (93.2)	1245 (91.1)	1589 (87.5)	1989 (92.6)	1070 (93.1)	11 608 (92.1)
Vaginal breech delivery	37 (2.6)	73 (5.0)	84 (5.2)	111 (6.8)	53 (3.9)	151 (8.3)	121 (5.6)	62 (5.4)	692 (5.5)
Caesarean section	1036 (73.3)	915 (63.0)	1041 (64.2)	955 (58.3)	927 (67.8)	1123 (61.9)	1306 (60.8)	709 (61.7)	8012 (63.6)

Data are presented as *n* (%). Chorioamnionitis includes clinically suspected as well as pathologically proven cases.
*Median (25th and 75th percentile).

Table 2 Neonatal characteristics of the study group stratified by admitting hospital A to H

	Hospital								
Characteristic	A (n=1413)	B (n=1453)	C (n=1622)	D (n=1640)	E (n=1367)	F (n=1815)	G (n=2149)	H (n=1149)	All (n=12 608)
Onsite surgical support	No	No	Yes	No	No	No	Yes	Yes	–
Born in a non-tertiary hospital (outborn)	115 (8.1)	164 (11.3)	162 (10.0)	151 (9.2)	119 (8.7)	157 (8.7)	282 (13.1)	104 (9.1)	1254 (9.9)
Male sex	761 (53.9)	819 (56.4)	894 (55.1)	879 (53.6)	727 (53.2)	962 (53.0)	1196 (55.7)	638 (55.5)	6876 (54.5)
Gestational age, week*	29.0 (27.0–31.0)	29.0 (27.0–30.0)	29.0 (27.0–30.0)	29.0 (27.0–31.0)	29.0 (27.0–31.0)	29.0 (27.0–30.0)	29.0 (27.0–31.0)	29.0 (27.0–31.0)	29.0 (27.0–30.0)
Birth weight, g*	1235.0 (970.0–1527.0)	1220.0 (925.0–1515.0)	940.0 (1240.0–1530.0)	1233.0 (910.0–1549.0)	1255.0 (955.0–1576.0)	1200.0 (920.0–1494.0)	1269.0 (960.0–1540.0)	1280.0 (981.0–1550.0)	1240 (943.0–1530.0)
Birth weight <10th percentile	121 (8.6)	119 (8.2)	130 (8.0)	136 (8.3)	106 (7.8)	165 (9.1)	164 (7.6)	103 (9.0)	1044 (8.3)
Head circumference <10th percentile	65 (4.8)	57 (4.0)	86 (5.4)	78 (5.3)	51 (3.9)	73 (4.2)	101 (4.7)	67 (5.9)	578 (4.8)
Apgar score <7 at 5 min	291 (20.7)	284 (19.6)	307 (19.0)	299 (18.3)	249 (18.3)	437 (24.2)	358 (16.8)	279 (24.4)	2504 (20.0)
Surfactant	892 (63.1)	827 (56.9)	997 (61.5)	902 (55.0)	912 (66.7)	1273 (70.1)	1133 (52.7)	627 (54.6)	7563 (60.0)
Major surgery	55 (3.9)	63 (4.3)	152 (9.4)	83 (5.1)	58 (4.2)	144 (7.9)	91 (4.2)	88 (7.7)	734 (5.8)
Length of hospital stay, days*	57.8 (44.5–77.1)	56.5 (40.7–78.8)	58.9 (42.9–83.1)	53.1 (38.0–79.7)	55.0 (40.0–76.8)	58.7 (41.7–84.9)	51.2 (37.5–71.8)	57.4 (42.3–81.0)	56.0 (40.8–79.0)

Data are presented as *n* (%).
*Median (25th and 75th percentile).

overestimation of intercentre variability and consequent false outlier classification.[27 29]

Of note is that hierarchical and other statistical models do not provide direct guidance on improving quality despite flagging areas for further investigation.[30] However, combining this robust statistical modelling with funnel plots is helpful in providing a systematic structure for quality improvement, as discussed below.

From the adjusted funnel plots with the upper and lower control limits, the eight perinatal centres could be divided into three categories with guidance for appropriate action:[2 31 32]

- Category 1, positive special-cause variation: performance of these centres is below the lower control limit. Lessons could be learnt from these centres to improve the performance of other centres.
- Category 2, common-cause variation: performance is within the control limits. This is most likely the result of factors intrinsic to the centres. The reduction of common-cause variation requires a fundamental change in the underlying process.[22] This should be informed by lessons learnt from Category 1. There are no grounds for acting in individual centres in this group.
- Category 3, negative special-cause variation: performance is above the upper control limit. This is most likely the result of factors extrinsic to the centres, and its reduction requires identification of and action on the special causes.[22] These centres need to identify and eliminate the special causes of their poorer results. Again, this should be informed by lessons learnt from Category 1.

Generally, variations in outcome rates among hospitals may be caused by measurement inaccuracy in assessing the outcome, differences in case mix, sampling variability or differences in hospital clinical practices.[25 26] Our hierarchical model-based risk-adjusted approach adjusts for variation arising from differences in the case-mix. In our data, we used prospectively collected statewide data using standardised definitions to preclude any data inaccuracy and sampling variations.

A proposed strategy to examine the special-cause variation is the pyramid investigation model.[33] This model checks five variables: data accuracy, patient case mix, structure and the process of care and carers.[2 33] Identifying the exact causes of common-cause and special-cause variation between NICUs in NSW and the ACT needs further research.

Our study is not without limitations. We adjusted for selected variables, but there may be others that were not collected in the database. The transfer pattern of outborn infants may disadvantage certain hospitals with a higher proportion of outborns. However, our hierarchical model adjusted for 'outborn' and other factors.

Our analysis demonstrated the utility of adjusted funnel plots for effectively identifying NICUs with high morbidity rates that may require intervention. Similar applications have been shown to improve quality and detect aberrations elsewhere in healthcare settings.[2 34] These methods have been in use in the manufacturing industry since the mid-1900s and have greatly improved the quality of products.[22] Adjusted funnel plots have two major advantages. First, they focus on the centre that fails relative to the best centre so, enabling a systematic approach to guide improvement.[2 34 35] Second, they can be employed in real or semi-real-time to detect aberrations early and act promptly. The method is generalisable for evaluation and performance improvement for NICUs and other similar healthcare settings.

Table 3 Major neonatal morbidities of the study group stratified by admitting hospital A to H

		Hospital								
		A	B	C	D	E	F	G	H	All
Morbidity	Morbidity estimates	(n=1413)	(n=1453)	(n=1622)	(n=1640)	(n=1367)	(n=1815)	(n=2149)	(n=1149)	(n=12 608)
Proven early sepsis	Observed (%)	18 (1.27)	31 (2.13)	29 (1.79)	41 (2.5)	33 (2.41)	35 (1.93)	37 (1.72)	24 (2.09)	248 (1.97)
	Expected (%)	29.99 (2.12)	31.41 (2.16)	28.67 (1.77)	33.52 (2.04)	23.27 (1.70)	41.22 (2.27)	34.16 (1.59)	26.09 (2.27)	248.37 (1.97)
	Predicted (%)	25.15 (1.78)	31.24 (2.15)	28.39 (1.75)	37.23 (2.27)	27.20 (1.99)	38.12 (2.10)	35.67 (1.66)	25.28 (2.20)	248.38 (1.97)
	RAR	1.65	1.96	1.95	2.18	2.30	1.82	2.06	1.91	1.98
	RSIR (95% CI)	0.77 (0.34–1.11)	1.09 (0.65–1.34)	0.92 (0.66–1.39)	1.15 (0.80–1.42)	1.05 (0.85–1.59)	1.06 (0.61–1.23)	0.84 (0.72–1.37)	1.09 (0.57–1.33)	1.00 (0.88–1.13)
Late sepsis	Observed (%)	217 (15.36)	237 (16.31)	272 (16.77)	258 (15.73)	222 (16.24)	297 (16.36)	351 (16.33)	183 (15.93)	2041 (16.16)
	Expected (%)	229.13 (16.22)	265.58 (18.28)	304.47 (18.77)	298.74 (18.22)	240.60 (17.60)	342.86 (18.89)	368.23 (17.13)	196.95 (17.14)	2246.56 (17.82)
	Predicted (%)	228.62 (16.18)	244.98 (16.86)	258.71 (15.95)	255.84 (15.60)	226.65 (16.58)	290.94 (16.03)	394.00 (16.24)	182.35 (15.87)	2037.45 (16.16)
	RAR	16.13	14.90	13.73	13.84	15.22	13.71	15.32	14.96	14.69
	RSIR (95% CI)	0.95 (0.84–1.06)	1.09 (0.86–1.34)	1.22 (0.95–1.46)	1.14 (0.90–1.14)	1.07 (0.87–1.09)	1.19 (0.92–1.22)	1.07 (0.92–1.10)	1.06 (0.88–1.13)	1.11 (0.97–1.39)
IVH grade III and IV†	Observed (%)	29 (14.36)	57 (19.59)	76 (24.2)	59 (13.2)	66 (24.44)	69 (16.05)	87 (20.57)	61 (23.19)	504 (19.09)
	Expected (%)	58.31 (4.27)	59.56 (4.28)	66.29 (4.29)	68.26 (4.33)	56.93 (4.35)	75.17 (4.32)	88.96 (4.29)	47.45 (4.31)	520.90 (4.30)
	Predicted (%)	60.34 (4.27)	62.19 (4.28)	69.58 (4.29)	71.01 (4.33)	59.46 (4.35)	78.41 (4.32)	92.19 (4.29)	49.52 (4.31)	542.14 (4.30)
	RAR	4.30	4.30	4.30	4.30	4.30	4.30	4.30	4.30	4.30
	RSIR (95% CI)	0.90 (0.63–1.12)	0.92 (0.64–1.13)	1.04 (0.76–1.22)	0.83 (0.56–1.02)	1.05 (0.74–1.24)	1.15 (0.88–1.31)	1.09 (0.85–1.25)	0.97 (0.66–1.20)	1.00 (0.88–1.04)
Medically treated PDA	Observed (%)	318 (22.51)	357 (24.57)	367 (22.63)	387 (23.6)	304 (22.24)	450 (24.79)	514 (23.92)	265 (23.06)	2968 (23.49)
	Expected (%)	332.15 (23.51)	357.68 (24.62)	375.73 (23.16)	369.14 (22.51)	329.20 (24.08)	422.22 (23.26)	506.81 (23.58)	265.63 (23.12)	2958.55 (23.47)
	Predicted (%)	328.24 (23.23)	357.44 (24.60)	373.06 (23.00)	374.58 (22.84)	322.48 (23.59)	431.24 (23.76)	509.53 (23.71)	265.53 (23.11)	2961.62 (23.49)
	RAR	23.22	23.49	23.34	23.84	23.02	24.00	23.62	23.49	23.15
	RSIR (95% CI)	0.97 (0.89–1.05)	1.05 (0.92–1.08)	0.97 (0.91–1.06)	0.99 (0.96–1.11)	0.97 (0.86–1.02)	1.03 (0.97–1.11)	1.01 (0.94–1.07)	0.98 (0.91–1.09)	1.00 (0.98–1.03)
Surgically treated PDA	Observed (%)	31 (2.19)	32 (2.2)	37 (2.28)	35 (2.13)	23 (1.68)	38 (2.09)	46 (2.14)	27 (2.35)	269 (2.13)
	Expected (%)	30.10 (2.13)	32.89 (2.26)	34.11 (2.10)	33.27 (2.03)	30.38 (2.22)	38.43 (2.12)	46.14 (2.15)	23.69 (2.06)	269.11 (2.13)
	Predicted (%)	30.10 (2.13)	32.84 (2.26)	34.06 (2.10)	33.29 (2.03)	30.35 (2.22)	38.48 (2.12)	46.20 (2.15)	23.67 (2.06)	268.55 (2.13)
	RAR	2.13	2.13	2.13	2.13	2.13	2.13	2.13	2.13	2.13
	RSIR (95% CI)	1.03 (0.68–1.38)	1.03 (0.64–1.31)	1.07 (0.76–1.41)	1.00 (0.72–1.38)	0.79 (0.41–1.10)	0.98 (0.68–1.30)	1.01 (0.72–1.28)	1.10 (0.75–1.53)	1.00 (0.88–1.12)
NEC clinically or proven radiologically or at surgery	Observed (%)	52 (3.68)	42 (2.89)	56 (3.45)	72 (4.39)	39 (2.85)	93 (5.12)	52 (2.42)	73 (6.35)	479 (3.8)
	Expected (%)	47.52 (3.36)	49.77 (3.43)	78.66 (4.85)	59.72 (3.64)	45.88 (3.36)	77.48 (4.27)	72.90 (3.39)	50.36 (4.38)	482.02 (3.82)
	Predicted (%)	52.14 (3.69)	43.30 (2.98)	58.23 (3.59)	71.34 (4.35)	41.56 (3.04)	93.29 (5.14)	55.23 (2.57)	69.51 (6.05)	484.15 (3.84)
	RAR	4.17	3.30	2.82	4.54	3.44	4.58	2.88	5.25	3.82
	RSIR (95% CI)	0.88 (0.79–1.29)	0.88 (0.73–1.29)	1.23 (0.77–1.24)	0.97 (0.83–1.26)	0.83 (0.69–1.26)	1.12 (0.87–1.23)	0.84 (0.73–1.22)	1.21 (0.89–1.29)	1.00 (0.95–1.11)
Chronic lung disease	Observed (%)	283 (20.03)	359 (24.71)	407 (25.09)	409 (24.94)	299 (21.87)	410 (22.59)	495 (23.03)	254 (22.11)	2916 (23.13)
	Expected (%)	295.62 (20.92)	328.63 (24.61)	388.98 (25.98)	373.13 (24.75)	302.98 (21.16)	436.64 (22.05)	461.19 (23.46)	253.38 (22.05)	2840.26 (23.52)
	Predicted (%)	288.39 (20.41)	341.02 (23.47)	389.93 (24.04)	385.73 (23.52)	298.83 (21.86)	417.81 (23.02)	470.20 (21.88)	250.83 (21.83)	2843.10 (22.55)
	RAR	22.56	24.00	23.18	23.91	22.81	22.13	23.58	22.90	23.15
	RSIR (95% CI)	0.89 (0.74–1.13)	1.03 (1.01–1.18)	1.08 (1.02–1.18)	1.04 (1.02–1.18)	0.96 (0.95–1.13)	1.02 (0.96–1.11)	0.98 (0.89–1.18)	0.97 (0.96–1.15)	1.00 (0.85–1.10)

Continued

Table 3 Continued

Morbidity	Morbidity estimates	Hospital A (n=1413)	B (n=1453)	C (n=1622)	D (n=1640)	E (n=1367)	F (n=1815)	G (n=2149)	H (n=1149)	All (n=12 608)
Postnatal steroid for CLD	Observed (%)	71 (5.02)	105 (7.23)	157 (9.68)	98 (5.98)	98 (7.17)	177 (9.75)	84 (3.91)	87 (7.57)	877 (6.96)
	Expected (%)	79.76 (5.64)	100.83 (6.94)	119.85 (7.39)	118.09 (7.20)	91.50 (6.69)	132.71 (7.31)	133.18 (6.20)	73.45 (6.39)	849.26 (6.74)
	Predicted (%)	74.18 (5.25)	101.86 (7.01)	143.55 (8.85)	101.68 (6.20)	96.65 (7.07)	162.62 (8.96)	88.32 (4.11)	83.88 (7.30)	877.52 (6.96)
	RAR	6.48	7.03	8.33	5.99	7.34	8.52	4.62	7.95	6.96
	RSIR (95% CI)	0.78 (0.69–1.20)	1.03 (0.90–1.24)	1.16 (1.01–1.29)	1.00 (0.82–1.17)	0.98 (0.88–1.23)	1.15 (1.01–1.28)	0.85 (0.80–1.18)	0.95 (0.89–1.26)	1.00 (1.00–1.13)
ROP grade III to V‡	Observed (%)	64 (7.33)	68 (5.87)	80 (6.54)	86 (7.21)	61 (5.03)	114 (8.46)	68 (4.24)	77 (10.1)	618 (6.59)
	Expected (%)	44.60 (5.11)	81.06 (7.00)	83.49 (6.83)	84.01 (7.05)	79.16 (6.53)	91.76 (6.81)	96.84 (6.04)	43.18 (5.67)	600.52 (6.41)
	Predicted (%)	77.15 (5.46)	92.12 (6.34)	107.38 (6.62)	116.11 (7.08)	81.34 (5.95)	134.49 (7.41)	101.65 (4.73)	86.98 (7.57)	796.83 (6.32)
	RAR	7.04	5.97	6.40	6.62	6.00	7.16	5.16	8.80	6.48
	RSIR (95% CI)	1.04 (0.64–1.22)	0.98 (0.57–1.13)	1.02 (0.59–1.23)	1.09 (0.59–1.19)	0.84 (0.57–0.93)	1.18 (0.71–1.33)	0.82 (0.50–0.84)	1.15 (0.72–1.25)	1.02 (0.72–1.33)
Surgery for ROP	Observed (%)	33 (2.34)	45 (3.10)	48 (2.96)	49 (2.99)	59 (4.32)	87 (4.79)	38 (1.77)	45 (3.92)	404 (3.2)
	Expected (%)	36.19 (2.56)	49.79 (3.43)	56.08 (3.46)	57.15 (3.48)	43.38 (3.17)	62.12 (3.42)	64.63 (3.01)	33.35 (2.90)	402.68 (3.19)
	Predicted (%)	33.77 (2.39)	45.91 (3.16)	49.47 (3.05)	50.35 (3.07)	56.05 (4.10)	83.67 (4.61)	42.55 (1.98)	42.28 (3.68)	403.46 (3.20)
	RAR	2.98	2.95	2.82	2.82	4.14	4.31	2.11	4.06	3.22
	RSIR (95% CI)	0.78 (0.66–1.29)	1.05 (0.71–1.25)	1.05 (0.72–1.23)	1.06 (0.72–1.23)	1.04 (0.82–1.29)	1.11 (0.85–1.23)	0.84 (0.61–1.18)	0.97 (0.79–1.33)	1.00 (0.91–1.09)
Hospital mortality	Observed (%)	70 (4.95)	111 (7.64)	121 (7.46)	147 (8.96)	98 (7.17)	113 (6.23)	190 (8.84)	77 (6.7)	927 (7.35)
	Expected (%)	85.04 (6.02)	111.60 (7.68)	120.31 (7.42)	121.87 (7.43)	100.62 (7.36)	145.29 (8.01)	145.54 (6.77)	75.89 (6.61)	906.23 (7.19)
	Predicted (%)	73.76 (5.22)	111.30 (7.66)	119.70 (7.38)	141.04 (8.60)	98.29 (7.19)	117.98 (6.50)	179.66 (8.36)	76.87 (6.69)	919.12 (7.29)
	RAR	6.37	7.33	7.32	8.51	7.18	5.97	9.08	7.44	7.47
	RSIR (95% CI)	0.78 (0.75–1.16)	1.04 (0.84–1.16)	1.02 (0.86–1.17)	1.05 (0.91–1.19)	1.00 (0.83–1.17)	1.04 (0.80–1.12)	0.97 (0.94–1.19)	0.90 (0.81–1.20)	0.99 (0.96–1.07)
Home oxygen	Observed (%)	18 (1.27)	140 (9.64)	70 (4.32)	113 (6.89)	46 (3.37)	40 (2.2)	219 (10.19)	39 (3.39)	685 (5.43)
	Expected (%)	53.57 (3.79)	66.99 (4.61)	75.34 (4.64)	75.36 (4.59)	59.74 (4.37)	83.97 (4.63)	89.17 (4.15)	47.03 (4.09)	551.17 (4.37)
	Predicted (%)	19.78 (1.40)	138.47 (9.53)	70.07 (4.32)	112.18 (6.84)	46.48 (3.40)	41.38 (2.28)	217.26 (10.11)	39.30 (3.42)	684.61 (5.43)
	RAR	2.01	11.23	5.06	8.09	4.23	2.67	13.23	4.54	6.74
	RSIR (95% CI)	0.63 (0.48–1.34)	0.86 (0.83–1.16)	0.85 (0.78–1.22)	0.85 (0.84–1.18)	0.80 (0.72–1.26)	0.82 (0.67–1.26)	0.77 (0.69–1.13)	0.75 (0.69–1.29)	0.81 (0.73–1.07)

Data are presented as n (%). Predicted and expected hospital morbidities were calculated using hierarchical model-based risk-adjusted for intrinsic factors described in online supplemental table 1–12.
†Based on 12131 records.
‡Based on 9371 records.
CLD, chronic lung disease; IVH, intraventricular haemorrhage; NEC, necrotising enterocolitis; PDA, patent ductus arteriosus; RAR, risk-adjusted rate (%); ROP, retinopathy of prematurity; RSIR, risk-adjusted standardised incidence ratio.

Author affiliations
[1]Department of Neonatology, Centenary Hospital for Women and Children, Canberra Hospital, Garran, Canberra, ACT, Australia
[2]Department of Public Health, College of Science Health and Engineering, La Trobe University, Bundoora, Melbourne, VIC, Australia
[3]Discipline of Neonatology, School of Medicine and Psychology, College of Health and Medicine, Australian National University, Acton, Canberra, ACT, Australia
[4]Public Health and Tropical Medicine, College of Public Health, Medical and Veterinary Sciences, James Cook University, Townsville, Queensland, Australia
[5]Australian Institute of Tropical Health and Medicine, James Cook University, Townsville, QLD, Australia
[6]Research School of Finance, Actuarial Studies, and Statistics, College of Business and Economics, Australian National University, Acton, Canberra, ACT, Australia
[7]Department of Mathematics, Physics, and Statistics, Qatar University, Doha, Qatar
[8]Critical Care Program, Sydney Children's Hospitals Network, Westmead, Sydney, NSW, Australia
[9]National Centre for Epidemiology and Population Health, Australian National University, Acton, Canberra, ACT, Australia
[10]University of Queensland Centre for Clinical Research (UQCCR), University of Queensland, Brisbane, QLD, Australia

Twitter Mohamed E Abdel-Latif @Latifme1

Acknowledgements The authors thank the Directors, the NICUS members and the audit officers of all tertiary units in supporting this collaborative study: NICUS, Dr Himanshu Popat (Chairperson), Barbara Bajuk (Coordinator), Sara Sedgley (CNS); Canberra Hospital, Dr Hazel Carlisle (Director), Melanie Rosin; John Hunter Children's Hospital, Dr Larissa Korostenski (Director), Alissa Argomand, Stacey Leonard; Royal Prince Alfred Hospital, Dr Mark Greenhalgh (Director), Shelley Reid; Liverpool Hospital, Dr Jacqueline Stack (Director), Dr Ian Callander, Kathryn Medlin, Amanda Beesley; Nepean Hospital, Dr Lyn Downe, Mee Fong Chin; The Children's Hospital at Westmead, Professor Nadia Badawi (Director), Karina Rogers; Royal North Shore Hospital, Dr Mary Paradisis (Director), Professor Martin Kluckow, Lyn Barnes; Sydney Children's Hospital, Dr Andrew Numa (Director), Dr Gary Williams, Janelle Young; Westmead Hospital, Dr Melissa Luig (Director), Jane Baird, Gemma Lowe and Royal Hospital for Women, Dr Srinivas Bolisetty (Director), Professor Julee Oei, Diane Cameron. We also thank the babies and their families, the nursing and midwifery, obstetric and medical records staff of the obstetric and children's hospitals in NSW and the ACT.

Contributors MEA-L: conceptualised and designed the study, contributed to statistical analyses and interpretation, drafted the initial manuscript, led the writing and reviewed and revised the manuscript. OA and FE: designed the statistical plan, carried out the initial analyses and critically reviewed and revised the manuscript. GN: contributed to the statistical interpretation, critically reviewed the statistical methods, analysis and interpretation and critically reviewed and revised the manuscript. BB: coordinated and supervised data collection, retrieved and cleaned the data, contributed to statistical interpretation and critically reviewed and revised the manuscript. KG: contributed to the statistical interpretation critically and reviewed and revised the manuscript. DH: contributed to the study design and statistical interpretation and critically reviewed and revised the manuscript. All authors approved the final manuscript as submitted. MEA-L and OA had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval The Australian National University (No 2015/029), the ACT Health Human Research Ethics Committee (No ETH.11.09/1031) and Hunter New England Research Ethics and Governance Unit have approved this study. The Ethics Committee waived the requirement for individual patient consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. All data relevant to the study are included in the article or uploaded as supplementary information. The data for this study were extracted from The Neonatal Intensive Care Units (NICUs) Data Registry, an ongoing prospective statewide audit of infants admitted to the 10 NICUs in NSW and the ACT during the neonatal period. The data are available on reasonable request from the data custodian, NICUS Data Management Committee, Neonatal Intensive Care Units' Data Registry, Agency for Clinical Innovation, NSW, Australia. All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iDs
Mohamed E Abdel-Latif http://orcid.org/0000-0003-4306-2933
Oyelola Adegboye http://orcid.org/0000-0002-9793-8024
Gen Nowak https://orcid.org/0000-0001-5486-6318
Faiz Elfaki https://orcid.org/0000-0002-4081-9235
Barbara Bajuk https://orcid.org/0000-0002-7439-4132
Kathryn Glass https://orcid.org/0000-0001-5905-1310
David Harley https://orcid.org/0000-0001-6691-5420

REFERENCES

1 Field D, Manktelow B, Draper ES. Bench marking and performance management in neonatal care: easier said than done! *Arch Dis Child Fetal Neonatal Ed* 2002;87:163F–4.
2 Mohammed MA, Cheng KK, Rouse A, *et al*. Bristol, shipman, and clinical governance: shewhart's forgotten lessons. *Lancet* 2001;357:463–7.
3 Abdel-Latif ME, Nowak G, Bajuk B, *et al*. Variation in hospital mortality in an Australian neonatal intensive care unit network. *Arch Dis Child Fetal Neonatal Ed* 2018;103:F331–6.
4 Abdel-Latif ME, Bajuk B, Oei J, *et al*. Mortality and morbidities among very premature infants admitted after hours in an Australian neonatal intensive care unit network. *Pediatrics* 2006;117:1632–9.
5 Abdel-Latif ME, Bajuk B, Oei J, *et al*. Does rural or urban residence make a difference to neonatal outcome in premature birth? A regional study in Australia. *Arch Dis Child Fetal Neonatal Ed* 2006;91:F251–6.
6 Lui K, Abdel-Latif ME, Allgood CL, *et al*. Improved outcomes of extremely premature outborn infants: effects of strategic changes in perinatal and retrieval services. *Pediatrics* 2006;118:2076–83.
7 Abdel-Latif ME, Berry A. Analysis of the retrieval times of a centralised transport service, New South Wales, Australia. *Arch Dis Child* 2009;94:282–6.
8 Shah PS, Shah V, Qiu Z, *et al*. Improved outcomes of outborn preterm infants if admitted to perinatal centers versus freestanding pediatric hospitals. *J Pediatr* 2005;146:626–31.
9 Australian Bureau of Statistics. National, state and territory population, 2021. Available: https://www.abs.gov.au/statistics/people/population [Accessed 06 Jan 2022].
10 National Health and Medical Research Council. *Clinical practice guidelines: care around preterm birth*. Canberra, Australia: National Health and Medical Research Council, 1997.
11 Australian and New Zealand Neonatal Network. *ANZNN 2022 data dictionary*. Sydney, Australia, 2021. https://anznn.net/dataresources/datadictionaries
12 Bajuk B. Validation of the neonatal intensive care units' data collection. Proceedings of the 5th Annual Conference of the Perinatal Society of Australia and New Zealand; The Perinatal Society of Australia and New Zealand, Canberra, ACT, Australia, 2001.
13 Bajuk B. *Neonatal Intensive Care 'Units' (NICUS) Data Registry Definitions*. Randwick, New South Wales, Australia, 2020.
14 Shennan AT, Dunn MS, Ohlsson A, *et al*. Abnormal pulmonary outcomes in premature infants: prediction from oxygen requirement in the neonatal period. *Pediatrics* 1988;82:527–32.
15 Papile LA, Burstein J, Burstein R, *et al*. Incidence and evolution of subependymal and intraventricular hemorrhage: a study of infants with birth weights less than 1,500 GM. *J Pediatr* 1978;92:529–34.
16 Bell MJ, Ternberg JL, Feigin RD, *et al*. Neonatal necrotizing enterocolitis. therapeutic decisions based upon clinical staging. *Ann Surg* 1978;187:1–7.
17 Flynn JT. An international classification of retinopathy of prematurity: clinical experience. *Trans Am Ophthalmol Soc* 1984;82:218–38.
18 International Committee for the Classification of Retinopathy of Prematurity. The International classification of retinopathy of prematurity revisited. *Arch Ophthalmol* 2005;123:991–9.
19 Brown KA, Heath Jeffery RC, Bajuk B, *et al*. Sight-threatening retinopathy of prematurity: changing trends in treatment. *J Pediatr Ophthalmol Strabismus* 2016;53:90–5.
20 Evans N. Current controversies in the diagnosis and treatment of patent ductus arteriosus in preterm infants. *Adv Neonatal Care* 2003;3:168–77.

21 Hoellering AB, Cooke L. The management of patent ductus arteriosus in Australia and New Zealand. *J Paediatr Child Health* 2009;45:204–9.
22 Isaacs D, Barfield CP, Grimwood K, *et al*. Systemic bacterial and fungal infections in infants in Australian neonatal units. Australian study group for neonatal infections. *Med J Aust* 1995;162:198–201.
23 Abdel-Latif ME, Bajuk B, Oei J, *et al*. Population study of neurodevelopmental outcomes of extremely premature infants admitted after office hours. *J Paediatr Child Health* 2014;50:E45–54.
24 Lemeshow S, Hosmer DW. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982;115:92–106.
25 Kleinbaum DG. *Logistic regression: a self-learning text*. New York: Springer-Verlag, 1994.
26 Hosmer DW, Lemeshow S. *Applied logistic regression*. 2nd. New York: Wiley, 2000.
27 Austin PC, Alter DA, Tu JV, . The use of fixed- and random-effects models for classifying hospitals as mortality outliers: a monte carlo assessment. *Med Decis Making* 2003;23:526–39.
28 Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med* 2005;24:1185–202.
29 Shahian DM, Torchiana DF, Shemin RJ, *et al*. Massachusetts cardiac surgery report card: implications of statistical methodology. *Ann Thorac Surg* 2005;80:2106–13.
30 McNeil BJ, Pedersen SH, Gatsonis C. Current issues in profiling quality of care. *Inquiry* 1992;29:298–307.
31 Mohammed MA, Worthington P, Woodall WH. Plotting basic control charts: tutorial notes for healthcare practitioners. *Qual Saf Health Care* 2008;17:137–45.
32 Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. *Qual Saf Health Care* 2003;12:458–64.
33 Mohammed MA, Rathbone A, Myers P, *et al*. An investigation into general practitioners associated with high patient mortality flagged up through the shipman inquiry: retrospective analysis of routine data. *BMJ* 2004;328:1474–7.
34 Berwick DM. Controlling variation in health care: a consultation from walter shewhart. *Med Care* 1991;29:1212–25.
35 Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical issues in comparisons of institutional performance. *J R Stat Soc Ser A Stat Soc* 1996;159:385.