THEORY AND METHODS

Performance profile of an outcome measure: morbidity assessment index for newborns

Anila Verma, Angela Weir, Jane Drummond, Bryan F Mitchell

J Epidemiol Community Health 2005;**59**:420–426. doi: 10.1136/jech.2003.019109

See end of article for authors' affiliations

Correspondence to: Dr A Verma, Clinical Safety and Epidemiology, Bldg 419/1150, One Health Plaza, East Hanover, Novartis Pharmaceuticals, New Jersey, USA 07936; anila.verma@pharma.novartis.com

Accepted for publication 12 December 2004

Objective: Few tools have been optimised for use over the entire spectrum of neonatal morbidity and standardised for use in perinatal population and community health studies. The objective of this study was to determine the performance profile of the recently developed morbidity assessment index for newborns (MAIN score). This score was designed as a discriminative index of morbidity for the entire population of babies delivered >28 weeks gestation without a major congenital anomaly.

Design and setting: MAIN score items were extracted retrospectively from the health records of 2892 consecutively born babies delivered beyond 28 weeks gestation in Edmonton area hospitals between June and December of 1999.

Main results: The mean MAIN score in the general newborn population was 70.3 (95% confidence intervals 64.2 to 76.4). With the MAIN score tool, 84.6% of newborns scored from 0 to 150 (no/minimal morbidity), 11.3% from 151 to 500 (mild), 3.1% from 501 to 800 (moderate), and 1% had >800 (severe) score. The MAIN score tool was sufficiently sensitive to detect significant effects of low gestational age, low birth weight, male sex, caesarean delivery, tertiary hospital delivery, twins/triplets, non-vertex presentation, prenatal illicit drug use, and medical complications of pregnancy.

Conclusion: The MAIN score fulfills the need for a simple, universal, yet sensitive and robust tool to provide a numerical index of early neonatal outcomes of prenatal care and adverse prenatal exposures in babies delivered beyond 28 weeks gestation. The performance of the MAIN score agrees well with the current medical awareness regarding the impact of adverse prenatal exposures on newborn morbidity.

Few tools of newborn morbidity have been optimised and standardised for use over the entire spectrum, from minimal to severe, for use in perinatal population studies of early neonatal period. Traditionally, indicators such as perinatal or infant mortality, birth weight, or gestational age have been used as outcome measures of effectiveness of national health policies and interventions for pregnant women. While mortality rates still have some value in underdeveloped countries, they are of diminishing significance for countries with highly developed healthcare systems. In the absence of valid measures of newborn morbidity, low birth weight (BW), low gestational age (GA), or Apgar score are considered to be proxy measures to report morbidity at birth.[1–3] However, these proxy measures are insufficiently sensitive or precise to provide the information necessary to shape current maternal-child healthcare policy. The Apgar score was originally designed as a discriminative index to identify infants in need of immediate cardiopulmonary resuscitation and performs this function well. However, it is a poor index of a newborn's health beyond the first few hours of life.[4 5] A joint statement published by the American Association of Pediatricians with the American College of Obstetricians and Gynecologists emphasises that a low Apgar score has little meaning unless it occurs in combination with profound metabolic or mixed acidaemia (pH<7) and neurological manifestations, for example, seizure, coma, or hypotonia to diagnose neurological injury at birth.[6]

Several specific measurement tools have been developed in the past to assess morbidity in very low birthweight (<1500 g) or low GA babies.[7–9] Such babies constitute <2% of all births and morbidity in this highly selected group does not emulate the morbidity pattern of most of the newborn population. Perinatal epidemiologists and community health researchers need a standardised, valid tool for assessing general newborn health across the entire spectrum of morbidity at birth. Such measures should not only be reliable, universal, and commonly found in the study population but should also be sufficiently sensitive to warrant concern when there is a positive finding.

Recently, we developed and validated the structure of the morbidity assessment index for newborns (MAIN score). The details of the design strategy of the MAIN score have been published previously.[10–12] This score was designed specifically for babies born without major congenital anomalies at ⩾28 weeks' gestation. This population includes about 97% of all newborns. The MAIN score is based on items of routine clinical and laboratory examination of newborns. It was designed to be simple and easily completed by retrospective chart review. Its purpose is to reflect morbidity in the first week of life. The specific objective of the design of the MAIN score was to provide a tool to assist population health researchers in comparing healthcare delivery systems available for pregnant women. The MAIN score could also serve as a baseline assessment for subsequent studies of developmental outcomes in infants.

The specific aim of this study was to examine the performance profile of MAIN score in the entire newborn population in three community hospitals in the city of Edmonton, Canada. We hypothesised that, if the MAIN score is a sensitive and discriminative measure of newborn morbidity, there would be significantly higher scores in newborns from pregnancies complicated by a variety of adverse prenatal exposures and medical conditions. Therefore, after the initial analyses, the population of newborns was stratified according to factors known to influence newborn health (for example, GA, BW, singletons/multiple

Abbreviations: BW, birth weight; GA, gestational age; MAIN, morbidity assessment index for newborns

births, malpresentation, maternal medical complications, prenatal adverse exposures, etc). These analyses have provided evidence for validity of the MAIN score as a discriminative tool of newborn morbidity.

METHODS

The Health Research Ethics Board of University of Alberta approved this study for data collection by chart review at three Edmonton hospitals: Royal Alexandra (RAH), Grey

Table 1 MAIN score morbidity items with their scale values and goodness of fit parameters

Circle all items that apply between birth and discharge from the hospital or up to seven days of life, whichever is earlier. *≥2 consecutive readings				
Item	**Morbidity attribute**	**Discrimatory index†**	**χ^2‡**	**Scale**
Within 24 hours of birth				
	Cord blood pH			
1	⩽7.1	0.50	6.7	151
	Resuscitation at birth			
2	Intubation	0.88	6.5	127
	Meconium			
3	Meconium below cords	0.35	12.7	155
	Apgar score (5 min)			
4	score 4–7	0.47	13	125
5	score 1–3	0.97	0.8	162
6	score<1	1.6	0.3	193
	Apgar score (10 min)			
7	score 4–7	1.4	2	154
8	score 1–3	1.7	0.4	183
	Altered colour*			
9	Dusky/central cyanosis	0.74	2.2	145
	Respiratory rate/min*			
10	<30 or >60 at 3–24 h	0.54	5.3	131
11	>100 between 3–24 h	0.58	11.5	140
Within seven days of birth				
	Heart rate/min*			
12	160–200 beat	0.54	10.3	120
13	>200 beat	0.67	NA	183
14	<100 beat	0.5	3.1	157
	Hypotonia*			
15	Present at 1–120 h of age	1.0	1.5	129
16	Present beyond 120 h of age	1.36	9.2	156
	Flaccidity*			
17	Present at 1–120 h	1.07	7.4	154
	Apnea*			
18	Apnea corrected by oxygen	1.26	10.8	115
19	Apnea corrected by resuscitation	1.04	2.2	140
	Bleeding disorder			
20	Thrombocytopenia with or without bleeding disorder (GI/lungs/skin)	0.84	0.5	147
21	Need for transfusion due to anemia or item 20	1.1	0.2	170
	Systolic BP (mean, mm of Hg)*			
	28–32 weeks 32–42 weeks			
22	<30 <40	0.85	3.4	136
	Urine output*			
23	Low (<2 ml/kg/h)	1.25	7.1	141
	Seizures			
24	Tremors/single seizure	0.81	4.5	137
25	Multiple seizures	1.67	1.7	155
26	If >2 drugs used for seizures	1.94	1.0	183
	Level of consciousness*			
27	Drowsy/lethargic	0.92	4.2	137
28	Stupor/obtundation/coma	1.75	0.6	187
	Oral feeding difficulties*			
29	Poor sucking within 24 h	1.04	10.2	81
30	Poor sucking at 24 h–7 days	1.32	8.0	98
31	Poor sucking beyond 7 days	1.04	22.6	119
32	Persistent vomiting	0.62	13.9	136
	Respiratory status—assisted ventilation*			
33	Assisted ventilation beyond 24 h	0.66	30.8	117
	Respiratory status—mechanical ventilation*			
34	Mechanical ventilation within 24 h	1.09	5.5	130
35	Mechanical ventilation at 24 h–7 days	1.4	6.2	135
36	Mechanical ventilation beyond 7 days	1.16	2.6	162
	Birth trauma			
37	Bone fracture—long bone/clavicle/skull	0.7	NA	176
38	Nerve injury (facial/peripheral)	0.49	NA	183
39	Subdural or intracerebral haematoma	0.71	NA	179
	Hypoglycaemia (lowest level)			
40	Blood glucose <2.2 mmol/l	0.44	1.6	151
	Hyperbilirubinaemia, mmol/l (peak level)			
41	Serum bilirubin >250/phototherapy	0.37	28.9	103
42	Serum bilirubin >340/exchange transfusion	0.42	NA	179
	Bacterial culture			
43	Blood positive	0.86	1.8	162
44	CSF positive	0.93	0.1	187

Table 1 Continued

Circle all items that apply between birth and discharge from the hospital or up to seven days of life, whichever is earlier. *≥2 consecutive readings				
Item	**Morbidity attribute**	**Discrimatory index†**	**χ^2‡**	**Scale**
	Intra-ventricular haemorrhage			
45	Grade 1 or 2	0.86	0.6	152
46	Grade 3 or 4	1.84	0.8	186
	Cardiopulmonary resuscitation			
47	Any time before discharge	2.06	2.6	162

MAIN score, sum of the scale values of all checked items. †Discriminatory index, this is an indicator of discriminative capability of an item to capture various different grades of morbidity on the morbidity continuum. ‡χ^2, this statistic reflects fitness of observed on predicted test score distributions. Scale, relative scale values derived from item response analyses that provide weighting to the item(s) reflecting its contribution to the overall morbidity score. NA, not available because of small number of subjects for χ^2 analysis. All statistical parameters are derived fitting two parameter item analysis model using BILOG-W software. Please see text for further details.

Nuns (GNH), and Misericordia (MH). In addition to a community maternal-child health service, the RAH is the regional tertiary care facility for high risk deliveries. The GNH and MH are community hospitals. Data on the MAIN score morbidity items were obtained from the health records of 2892 newborns delivered at the three hospitals from June to December 1999, born at ⩾28 weeks' gestation with no major life threatening anomaly at birth. The MAIN score for each recruited newborn was calculated by retrospective review of the hospital chart. Relevant maternal data (see later) were collected retrospectively from the chart of each newborn's mother.

For designing the MAIN score, innovative item response theory (IRT)[13 14] and dimensionality testing techniques[15] were used for structural validation and item analysis, as published previously.[10–12] Structural validation techniques helped evaluate relation between the morbidity items and the underlying construct of newborn morbidity by determining the goodness of fit of items included in the MAIN score inventory. Item analysis provided scale values or weights that were assigned to each morbidity item contained in the MAIN score inventory. The scale values are relative values that provide weighting to the item(s) reflecting its contribution to the overall morbidity score. The statistical parameters of goodness of fit, including χ^2 statistics, discrimination statistics (shown in table 1), and item characteristic curves (ICC) were also obtained from the item analysis model, using BILOG-W software (BILOG, Scientific Software, Chicago, IL, 2nd edition, 1990).

The χ^2 statistic reflects fitness of observed on predicted test score distributions. The discrimination statistic is an indicator of discriminative capability of an item to capture various grades of morbidity on the morbidity continuum. The discrimination statistic in conjunction with the ICCs provided evidence regarding information function of an individual item and the goodness of fit of items in the MAIN score. In brief, after a number of iterations and re-evaluations of goodness of fit of different combinations of items, an inventory containing 47 binary items representing 24 attributes of early neonatal morbidity was specified as the best fit for the MAIN score. Table 1 contains the scale values, χ^2, and discrimination statistics of the morbidity items estimated using BILOG-W software for a two parameter IRT model. The MAIN score for a newborn is computed by aggregating the scale values of all items selected in the MAIN score inventory (table 1) from birth until seven days of life or discharge from the hospital, whichever is earlier. For example, a newborn having transient morbidity attributable to apnea corrected by oxygen (item 18) and respiratory rate >60/min between 3 to 24 hours of life (item 10) will receive a MAIN score of 246; whereas, a newborn with an Apgar score of 6 at five minutes (item 4) requiring resuscitation with intubation (item 2), who had a single seizure (item 24), in addition to hypotonia (item 15) and poor sucking for the first three days after birth (item 30) will receive a MAIN score of 616.

The review of charts of many newborns showed no indication or suspicion of morbidity during the early neonatal period and therefore, no assignment of morbidity according to items in the MAIN score. These infants were assigned a MAIN score of "zero". We have assumed that such newborns were deemed healthy and in no need for elaborate testing (for example, bacterial culture or urine output assessment) at birth before discharge from the hospital.

Sample size considerations

Sample size calculations were based on the probability of occurrence of 47 morbidity items contained in the MAIN score. Measurement theories advise that to provide stable parameter estimates for the measurement tool, at least five participants per item need to be recruited.[16] Considering that the prevalence of uncommon morbidity items may be as low as 0.2% (for example for peripheral nerve injury), we calculated that a sample of 2500 newborns was required for providing stable parameter estimates of morbidity items in the MAIN score. A total of 2892 eligible newborns were recruited to this study.

Newborn data collection

Data were collected for 2892 infants. Of these, 1405 (48.5%) newborns were recruited from the RAH, 875 (30.3%) from the GNH, and 606 (21%) from the MH. Depending on the complexity of morbidity, it takes from 2 to 15 minutes per chart to extract data for the MAIN score from hospital records. Customised electronic forms were prepared using Microsoft Access (version 5.0) for data entry and computation of the MAIN score. Three research assistants were trained using 20 charts each to extract morbidity data from the health records. Inter-rater reliability among the three data abstractors was tested on other sets of 20 charts for each pair—that is, a total of 40 charts. Inter-rater reliability between two raters in each pair was greater than 98%.

Maternal data collection

In all three hospitals there is a standardised nursing assessment of maternal risk indicators for both antepartum and intrapartum periods. The antepartum risk score assessed the number and severity of any maternal conditions that may adversely affect the pregnancy outcome for either mother or newborn. The risk items included pre-pregnancy maternal disease(s), history of obstetrical complications, and maternal complications associated with the current pregnancy. A pregnancy was considered low risk if the antepartum risk score was ⩽2, moderate risk if 3–5, high risk if 6–10, and extreme risk if >10. The most common maternal complications were defined by the following criteria:

Antepartum haemorrhage: bleeding from the genital tract after 28 weeks of gestation.

Table 2 Distribution of MAIN score in the overall population and in the subgroups stratified by specific maternal or newborn characteristics

			Mean MAIN score (% across morbidity categories)			
	N (%)	Overall mean (95% CI)	Nil/Minimal	Mild	Moderate	Severe
Overall population	2892 (100)	70.3 (64.2, 76.4)	11.7 (84.6)	282.6 (11.3)	619.7 (3.1)	977.1 (1.0)
Subgroups						
Sex						
Female	1412 (48.8)	60.9 (52.7, 69.0)	10.8 (86.8)	277.8 (9.8)	622.8 (2.6)	999.3 (0.8)
Male	1480 (51.2)	81.1 (71.8, 90.3)	12.6 (82.5)	286.2 (12.6)	617.3 (3.6)	964.3 (1.3)
Number of babies						
Singleton	2809 (97.1)	66.5 (60.4, 72.5)	11.6 (85.6)	282.1 (10.6)	618.6 (2.8)	974.6 (0.9)
Multiple	83 (2.9)	231.8 (170.7, 293.0)	22.2 (49.4)	288.6 (33.7)	627.1 (12.0)	993.7 (4.8)
Route of delivery, γ1						
Vaginal	2284 (79.1)	55.5 (49.7, 61.3)	10.6 (87.4)	284 (9.8)	613.2 (2.4)	926.0 (0.4)
Caesarean	602 (20.9)	130.8 (111.3, 150.2)	12.8 (73.8)	277 (16.9)	630 (5.8)	999.0 (3.5)
Malpresentation at delivery*, γ2						
No	2664 (92.4)	61.7 (56.0, 67.5)	11.3 (86.1)	280.1 (10.7)	615.3 (2.7)	975.53 (0.6)
Yes	220 (7.6)	185.7 (146.4, 224.9)	17.9 (66.4)	295.0 (18.6)	636.8 (8.2)	978.5 (6.8)
Hospital care type, γ3						
Community	1481 (51.3)	55.1 (48.2, 62.0)	12.3 (87.2)	257.8 (10.3)	619.9 (2.1)	883.9 (0.5)
Tertiary	1405 (48.6)	88.2 (77.8, 98.7)	11.1 (82.3)	306.5 (11.8)	619.4 (4.3)	1005.5 (1.6)
Maternal age (y)						
<35	2421 (83.7)	69.3 (62.5, 76.0)	11.3 (85.1)	279.8 (10.7)	616.7 (3.2)	998.7 (1.0)
⩾35	471 (16.3)	81.2 (65.4, 96.9)	13.5 (82.0)	293.8 (14.0)	636.6 (2.8)	891.0 (1.3)
Drug misuse, γ4						
No drug misuse†	2697 (93.7)	68.1 (61.8, 74.4)	11.6 (85.3)	280.6 (10.8)	626.7 (3.0)	979.9 (0.9)
Recreational	121 (4.2)	85.1 (52.9, 117.2)	15.3 (81.8)	282.4 (12.4)	548.6 (4.1)	899.5 (1.7)
Street	28 (1.0)	115.5 (50.5, 180.4)	23.5 (71.4)	310.6 (25.0)	589.0 (3.6)	0 (0)
Multiple	31 (1.1)	225.8 (128.4, 323.2)	26.9 (45.2)	299.3 (38.7)	556.3 (9.7)	869.5 (6.5)
Smoking, γ5						
No smoking†	2083 (72.7)	65.3 (58.2, 72.3)	11.3 (85.8)	273.3 (10.3)	616.4 (3.0)	1019.0 (0.9)
⩽10 cigarettes/day	464 (16.2)	82.4 (66.8, 98.0)	14.0 (81.0)	287.6 (14.9)	639.2 3.2)	887.7 (0.9)
>10 cigarettes/day	319 (11.1)	86.6 (65.6, 107.5)	12.1 (82.4)	309.5 (11.9)	611.4 (3.8)	888.7 (1.9)
Gestational age (weeks), four categories						
28–32	59 (2.0)	540.3 (468.7, 612.0)	73.3 (11.9)	385.0 (32.2)	644.0 (40.7)	955.3 (15.3)
>32–36	265 (9.2)	238.1 (206.8, 269.3)	37.8 (50.2)	312.8 (33.6)	610.6 (11.3)	915.2 (4.9)
>36–40†	2180 (75.4)	42.6 (37.4, 47.8)	10.1 (90.1)	266.0 (8.3)	611.8 (1.3)	1143.7 (0.3)
>40	388 (13.4)	46.7 (34.6, 58.9)	9.8 (87.9)	239.4 (9.8)	607.4 (2.1)	812.0 (0.3)
Gestational age (weeks), two categories						
<37 weeks	324 (11.2)	293.1 (261.8, 324.4)	39.6 (5.7)	325.4 (33.1)	625.4 (60.0)	1102.3 (6.8)
⩾37 weeks	2568 (88.8)	43.2 (38.4, 48.0)	10.0 (94.3)	261.4 (66.9)	610.8 (40.0)	931.6 (0.3)
Birth weight (g)						
⩾4500	55 (1.9)	84.6 (31.1, 138.2)	20.0 (83.6)	280.9 (12.7)	610.0 (1.8)	1159.0 (1.8)
4000–4499	288 (10.0)	60.7 (45.0, 76.3)	11.5 (82.3)	226.8 (15.3)	678.7 (2.4)	0 (0)
2500–3999†	2312 (79.9)	44.0 (39.0, 49.1)	10.2 (89.8)	267.2 (8.3)	587.5 (1.6)	1116.0 (0.3)
1501–2499	210 (7.3)	316.4 (277.6, 355.3)	44.8 (40.5)	346.1 (34.8)	623.6 (16.2)	898.8 (8.6)
<1500	27 (0.9)	574.5 (459.3, 689.7)	0 (3.7)	353.5 (40.7)	678.0 (40.7)	1041.3 (14.8)

*Includes face presentation, breech or transverse lie; †reference category for statistical significance where the number of categories for comparison is >2; missing cases: γ1 = 6; γ2 = 8; γ3 = 6; γ4 = 15, γ5 = 26.

Intrauterine growth restriction: ultrasound predicted BW <3rd centile using Canadian standards for gestational age.

Hypertension: chronic or pregnancy related hypertension: systolic blood pressure (BP) >140 mm Hg or diastolic BP >90 mm Hg on ⩾2 occasions, recorded at least six hours apart.

Pre-eclamptic toxaemia: pregnancy related hypertension plus proteinuria >1+ on a clean-catch urine specimen or when the total urinary protein is greater than 300 mg during a 24 hour period in a woman without a urinary tract infection.

Diabetes mellitus: insulin dependent diabetes mellitus before pregnancy.

Gestational diabetes mellitus: two or more values exceeding the recommended maternal blood glucose concentration criteria after a two hour 100 g oral glucose tolerance test.

Group B streptococcus: a positive culture growth of *Streptococcus agalactiae* upon routine screening at 35–36 weeks' gestation by lower vaginal/perianal swab.

Malpresentation: any obstetrical presentation other than cephalic.

Smoking was recorded as no smoking, <10 cigarettes/day, or >10 cigarettes/day.

The assessment form also showed the use of alcohol and marijuana (termed recreational drug use) or street drug use (including misuse of cocaine, narcotics, amphetamines, inhalation of fumes, etc) There was no distinction as to the quantity or frequency of drug ingestion in drug misusers.

The intrapartum risk score assessment included any of the above complications that onset during labour or present as labour related complications. This score also included assessment of gestational age, prolonged rupture of fetal membranes (>24 hours), maternal fever, fetal heart rate abnormalities, etc. An intrapartum score of 0 was considered no risk and a score of ⩾5 extreme risk.

Missing data

Data on the hospital of delivery and the type of delivery were missing from six maternal records; maternal smoking status was absent from 26 records; data regarding drug misuse, diabetes mellitus, and pre-eclamptic toxaemia were missing from 15 records and on presentation at delivery from eight

Table 3 Mean MAIN score (95% CI) in selected high risk maternal populations

Characteristics	N (%)	Mean MAIN score (95% CI)
Overall total*	2892 (100)	70.3 (64.2, 76.4)
APH	163 (5.7)	155.8 (113.5, 198.0)
IUGR	79 (2.7)	235.1 (168.6, 301.7)
Group B streptococcus	485 (18.5)	61.5 (47.8, 75.1)
Hypertension	223 (7.8)	127.2 (98.1, 156.3)
PET	118 (4.1)	157.2 (103.8, 210.7)
GDM	112 (3.9)	79.8 (51.5, 108.1)
Diabetes mellitus	30 (1.0)	135.6 (47.6, 223.6)

*Reference category. APH, antepartum haemorrhage; IUGR, intrauterine growth restriction; PET, pre-eclemptic toxemia; GDM, gestational diabetes mellitus.

records. The number of newborns thus available for various different analyses is shown in the respective tables. The hospital charts of all infants were available. As noted earlier, we have assumed that all abnormal examinations and tests were recorded on the chart. In the absence of checked items on the MAIN score inventory, we have considered that infant completely healthy with a MAIN score of zero. Thus, there were no missing data for calculation of the MAIN score.

Statistical analyses

Analyses were performed using the SPSS 10.0 version (IL, Chicago). To facilitate analysis and to improve the clinical application of MAIN score tool, it was elected to divide the MAIN scores into four subgroups according to their relative severity of morbidity. After consultation with the obstetrical and paediatric consultant groups from the three hospitals, morbidity scores were subgrouped empirically into: (1) none/minimal, score 0–150; (2) mild, 151–500; (3) moderate, 501–800; or (4) severe, >800.

Bivariate analytical techniques were used to test differences in the MAIN score among various specific newborn and maternal populations shown in tables 2 and 3. Linear trend analysis was performed to show the influence of increasing degree of adverse prenatal exposure and obstetrical complications on newborn morbidity.

RESULTS

Data from 2892 newborns (1480 males and 1412 females) delivered at or beyond 28 weeks' gestation, with no congenital anomaly were used to compute and analyse the MAIN score profile. The average MAIN score in our population was 70.3 with 95% CI 64.2 to 76.4 (table 2). On classifying 2892 newborns into clinically relevant qualitative categories, 84.6% had none/minimal, 11.3% mild, 3.1% moderate, and 1% severe morbidity. Table 2 also shows the distribution of MAIN scores within specific subgroups of newborn and maternal populations. Both the mean morbidity score and the proportion of morbidity were significantly higher in all of the following: males compared with females; twin/triplets compared with singletons; infants delivered by caesarean section compared with vaginal delivery; malpresentation compared with cephalic presentation; tertiary compared with community hospital delivery, infant of smoker compared with non-smoker; or infant of multiple drug misusers compared with recreational/no drug misusers. As shown in table 2, with each increment in GA (except ⩾40 weeks) and BW (except ⩾4500 g), there was significant reduction in newborn morbidity. Among twins and triplets, about half (49%) had no/minimal morbidity at birth but the remaining had significantly higher morbidity over the entire spectrum of morbidity compared with singletons.

Table 3 provides the MAIN scores for selected high risk maternal populations. Except for group B streptococcus and gestational diabetes, the MAIN scores for infants from mothers with other maternal conditions were statistically significantly higher than the overall population. With each

Key points

- The MAIN score fulfils the need for a simple, universal, yet sensitive and robust tool to provide a numerical index of the influence of prenatal care and adverse prenatal exposures on early newborn morbidity. Data presented show that the MAIN score provides fine discrimination across the spectrum of newborn morbidity in the general newborn population.
- In the public health domain, this tool could assist in community health programming by early identification of infants who may require increased use of community healthcare resources.

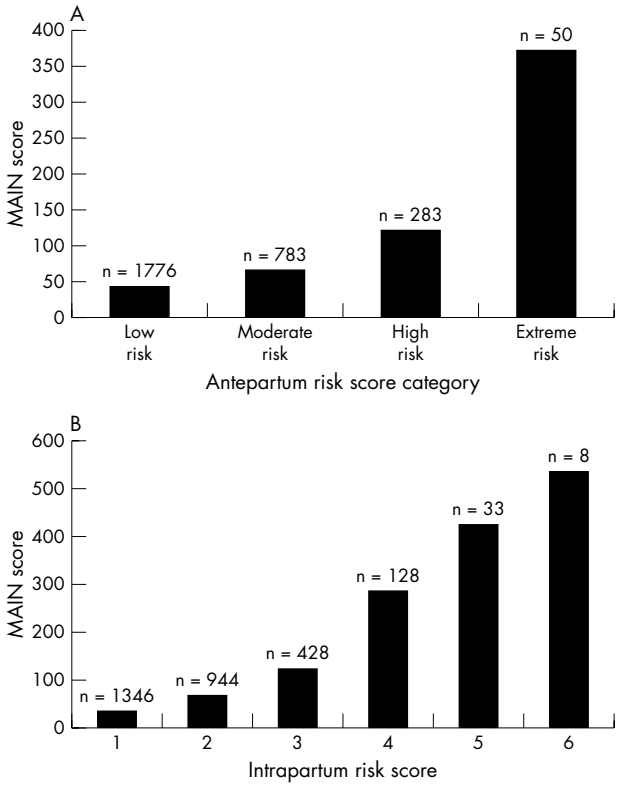

Figure 1 Relation between the MAIN scores and antepartum risk score category (A) and intrapartum risk score (B). There was a significant correlation between the MAIN scores and both of these obstetrical risk indicators. The numbers above the histograms show the number of infants in that group.

successive increase in the degree of adverse prenatal exposure, there was a significant trend towards increase in the MAIN score, for example, from no drug to recreational, street and multiple drug misuse ($p = 0.001$); from no smoking to smoking $\leqslant 10$ to >10 cigarettes/day ($p = 0.009$); from hypertension to PET ($p = 0.001$); and from gestational diabetes to pregestational diabetes ($p = 0.05$), as shown in tables 2 and 3.

In Figure 1 we have plotted the antepartum and intrapartum risk scores for all pregnant women against the mean MAIN scores of their infants. The MAIN scores had a significant linear positive correlation with both antepartum (Pearson's correlation coefficient, $r = 0.35$, $p = 0.01$) and intrapartum risk scores ($r = 0.38$, $p = 0.01$).

DISCUSSION

Sound decision making in perinatal epidemiology and policy development is dependent upon the use of readily available, generalisable, and valid outcome measures of morbidity. The primary aim of our study was to describe the performance of and establish normative values for the recently developed MAIN score. This score was developed specifically as a discriminative index of morbidity for newborns born beyond 28 weeks' gestation with no congenital or metabolic abnormality. This population constitutes about 97% of the total newborn population. The MAIN score captures the full range of morbidity, from minimal to severe. Our data show that the score is sensitive enough to capture subtle differences in morbidity according to GA, BW, sex, and maternal influences. We are not aware of any other measures of newborn morbidity that have illustrated such fine discrimination in evaluating severity of morbidity. This study also confirms that the MAIN score agrees well with the current medical awareness about the impact of adverse prenatal exposures on newborn morbidity.

In contrast with the parameters developed for very low BW (<1500 g) babies to predict developmental outcome, mortality, or severe morbidity,[7–9] the MAIN score was designed to function as a simple and universal measure of early neonatal morbidity over the entire spectrum of morbidity. Optimal performance of the MAIN score can be attributed to two major factors in the construction of this tool: (1) a comprehensive inventory of standard assessment items that reflect pathophysiology in early newborn period; and (2) the use of IRT to produce robust yet sensitive scaling units for each individual morbidity item.

We had excluded babies born with major structural anomaly or life threatening birth defect diagnosed at birth; however, we recognise that a small proportion of babies whose defect was not detected at birth (for example, metabolic anomaly or heart defect) may have been included in our dataset. Also, the four morbidity subgroups used to describe the severity of morbidity, from very mild to severe, are merely empirical, which should be substantiated in future studies by involving a larger number of perinatal consultants and with the help of large datasets.

We propose several potential uses for the validated MAIN score tool for policy makers and health administrators. Firstly, the score has utility in comparing the health of newborn populations across regions/countries or within regions after a change in the healthcare delivery system. Secondly, the MAIN score could be used as an outcome measure in research projects investigating the effects of novel interventions in the maternal-child healthcare delivery system. Furthermore, the score could be used to support hypotheses regarding the effects of adverse prenatal exposures on neonatal outcome. An example of this could be the data presented in this manuscript regarding the effects of maternal smoking or drug misuse on newborn's health.

Policy implications

We propose several potential uses of the validated MAIN score tool for policy makers and health administrators:

- The score has utility in comparing the health of newborn populations across regions/countries or within regions after a change in the healthcare delivery system. (Who could initiate this change? Public health decision makers and advocacy groups)
- The MAIN score could also be used as an outcome measure in research projects investigating the effects of novel interventions in the maternal-child healthcare delivery system. (Who could initiate this change? Public health, community health, and clinical researchers, and epidemiologists)
- The score could be used to support hypotheses regarding the effects of adverse prenatal exposures on neonatal outcome. An example of this could be the data presented in this manuscript regarding the effects of maternal smoking or drug misuse on newborn's health. (Who could initiate this change? Public health, community health, and clinical researchers, and clinical professionals)

The MAIN score fulfills the need for a simple, universal, yet sensitive and robust tool to provide numerical index of the influence of prenatal care and adverse prenatal exposures on early newborn morbidity. In summary, data presented herein show that the MAIN score provides fine discrimination across the spectrum of newborn morbidity in the general newborn population. In the public health domain, this tool could assist in community health programming by early identification of infants who may require increased use of community healthcare resources. Currently analyses of studies are underway to determine predictive value of MAIN score in health services use by one year of infant's age.

Authors' affiliations

A Verma, B F Mitchell, Department of Obstetrics and Gynaecology, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada
A Weir, J Drummond, Faculty of Nursing, University of Alberta

Funding: this research was supported by funds from three sources: The Alberta Heritage Foundation for Medical Research, record number 199800308; (2) University of Alberta Industry Liaison Office; (3) The University of Alberta Perinatal Research Center, Canada.

Conflicts of interest: none declared.

REFERENCES

1 **Johnson TS**. Hypoglycemia and the full-term newborn: how well does birth weight for gestational age predict risk? *J Obstet Gynecol Neonatal Nurs* 2003;**32**:48–57.
2 **Hansen JP**. Older maternal age and pregnancy outcome: a review of the literature. *Obstet Gynecol Surv* 1986;**41**:726–42.
3 **Brosson-Caroll G**, Fraser W, Beart G, *et al*. The effect of routine early amniotomy on spontaneous labor: a meta-analysis. *Obstet Gynecol* 1996;**87**:891–6.
4 **Schmidt B**, Kirpalani H, Rosenbaum P, *et al*. Strengths and limitations of the Apgar score: a critical appraisal. *J Clin Epidemiol* 1988;**41**:843–50.
5 **Jepson HA**, Talashek ML, Tichy AM. The Apgar score: evolution, limitations, and scoring guidelines. *Birth* 1991;**18**:83–92.
6 **Committee on Fetus and Newborn, American Academy of Pediatrics, and Committee on Obstetric Practice, American College of Obstetricians and Gynecologists**. Use and abuse of the Apgar score. *Pediatrics* 1996;**98**:141–2.
7 **de Courcy-Wheeler RH**, Wolfe CD, Fitzgerald A, *et al*. Use of the CRIB (clinical risk index for babies) score in prediction of neonatal mortality and morbidity. *Arch Dis Child Fetal Neonatal Ed* 1995;**73**:F32–6.

8 **Richardson DK**, Gray JE, McCormick MC, *et al*. Score for neonatal acute physiology: a physiology severity index for neonatal intensive care. *Pediatrics* 1993;**91**:617–23.
9 **Palta M**, Gabbert D, Fryback D. Development and validation of an index for scoring baseline respiratory disease in the very low birth weight neonate. *Pediatrics* 1990;**86**:714–21.
10 **Verma A**, Okun NB, Maguire TO, *et al*. Morbidity assessment index for newborns: a composite tool for measuring newborn health as an outcome following obstetrical interventions. *Am J Obstet Gynecol* 1999;**181**:701–8.
11 **Verma A**, Maguire TO, Mitchell BF. Criterion-related validation of an obstetrical outcome measure. (Abstract). *Am J Epidemiol* 1999;**149**:S25.
12 **Verma A**. *Development of an obstetrical outcome measure of newborn morbidity: the MAIN score*. (PhD Thesis). Edmonton: University of Alberta, 1997.
13 **Hambleton RK**, Swaminathan H, Rogers HJ, eds. *Fundamentals of item response theory*. Newbury Park, CA: Sage, 1991.
14 **Embretson SE**, Hershberger SL, eds. *The new rules of measurement: what every psychologist and educator should know*. Mahwah, NJ: Lawrence Erlbaum Associates, 1999.
15 **Stout WF**. A new response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika* 1990;**52**:589–617.
16 **Crocker L**, Algina J, eds. *Introduction to classical and modern test theory*. New York: Holt, Rinehart, and Winston, 1986:137–59.