BMJ Open

Applying measures of discriminatory accuracy to revisit traditional risk factors for being small for gestational age in Sweden: a national cross-sectional study

Sol Pía Juárez,[1,2] Phillip Wagner,[2] Juan Merlo[2]

CrossMark

[1]Center for Economic Demography, Lund University, Sweden
[2]Department of Clinical Sciences, Unit of Social Epidemiology, Lund University, Malmö, Skåne University Hospital (SUS Malmö), Malmö, Sweden

Correspondence to
Dr Sol Pía Juárez;
Sol.juarez@ekh.lu.se

ABSTRACT

Objectives: Small for gestational age (SGA) is considered as an indicator of intrauterine growth restriction, and multiple maternal and newborn characteristics have been identified as risk factors for SGA. This knowledge is mainly based on measures of average association (ie, OR) that quantify differences in average risk between exposed and unexposed groups. Nevertheless, average associations do not assess the discriminatory accuracy of the risk factors (ie, its ability to discriminate the babies who will develop SGA from those that will not). Therefore, applying measures of discriminatory accuracy rather than measures of association only, our study revisits known risk factors of SGA and discusses their role from a public health perspective.

Design: Cross-sectional study. We measured maternal (ie, smoking, hypertension, age, marital status, education) and delivery (ie, sex, gestational age, birth order) characteristics and performed logistic regression models to estimate both ORs and measures of discriminatory accuracy, like the area under the receiver operating characteristic curve (AU-ROC) and the net reclassification improvement.

Setting: Data were obtained from the Swedish Medical Birth Registry.

Participants: Our sample included 731 989 babies born during 1987–1993.

Results: We replicated the expected associations. For instance, smoking (OR=2.57), having had a previous SGA baby (OR=5.48) and hypertension (OR=4.02) were strongly associated with SGA. However, they show a very small discriminatory accuracy (AU-ROC≈0.5). The discriminatory accuracy increased, but remained unsatisfactorily low (AU-ROC=0.6), when including all variables studied in the same model.

Conclusions: Traditional risk factors for SGA alone or in combination have a low accuracy for discriminating babies with SGA from those without SGA. A proper understanding of these findings is of fundamental relevance to address future research and to design policymaking recommendations in a more informed way.

Strengths and limitations of this study

- Our study emphasises the use and interpretation of measures of discriminatory accuracy (ie, capacity to distinguish between small for gestational age (SGA) and non-SGA babies) when evaluating risk factors.
- We confirm statistical associations between maternal and newborn characteristics and risk for SGA, but we underline that the discriminatory capacity of all the risk factors studied was very low.
- This low discriminatory capacity suggests that we know very little about the determinants of SGA in the population and that more efforts should be devoted to understand individual heterogeneity of effects.
- Our finding is of fundamental relevance to address future research and to design policy-making recommendations in a more informed way.

INTRODUCTION

Small for gestational age (SGA) is commonly identified as a proxy for intrauterine growth restriction (IUGR).[1] This disorder has been associated with neonatal mortality and morbidities[2] as well as with major medical problems across the life course, such as a higher risk of neurodevelopmental impairments,[3 4] autism,[5] schizophrenia,[6] impaired cognitive function,[7] coeliac disease in boys[8] and reduced bone mass during early infancy,[9] as well as Barrett's oesophagus and oesophagitis[10 11] and others.[12 13] Therefore, the identification of maternal and newborn characteristics (denominated as 'risk factors' in the rest of this work) associated with an increased risk for SGA is of obvious relevance in public health and preventive medicine.

BMJ

Two reviews, one from 1987[14] and the other from 2009,[15] pointed out that SGA is associated with a broad number of genetic, obstetric, demographic and socio-economic factors as well as maternal morbidities and toxic exposures before and during pregnancy. However, the identification of these risk factors has been exclusively based on measures of average association (eg, OR) but without considering their accuracy for discriminating babies with, from those without, SGA. Indeed, it is common practice to use measures of average association to gauge the ability of a factor to discriminate future cases of disease.[16] For example, it is known that maternal hypertension during pregnancy gives a 5.5-fold increased risk of delivering an SGA baby.[17] Therefore, this variable is implicitly used as a predictive test to classify who will and who will not deliver an SGA baby. However, in spite of this popular belief, measures of association alone are inappropriate for this discriminatory purpose insofar as there are different scenarios of sensitivity/specificity for a given OR.[16–23]

Although measures of discriminatory accuracy are extensively applied in other fields of epidemiology like the identification of new biomarkers for cardiovascular diseases,[18–21] these measures are still unusual in public health and epidemiology.[22] In fact, as far as we know, they have never been explicitly used to formally revisit established maternal and newborn risk factors for SGA.

With this background our study aims to revisit the role of current risk factors for SGA in public health. We do it in two steps. First, using measures of average association, we aim to replicate previous findings and identify maternal and newborn risk factors for SGA. Second, we apply measures of discriminatory accuracy to assess the ability of those risk factors (alone or in combination) to discriminate babies with, from those without, SGA in the whole population and in different subgroups according to gestational age.

DATA AND METHODS

Study design, setting and participants

This is a cross-sectional study based on a population-based register. We identified all the 811 599 babies born alive and recorded at the Swedish Medical Birth Registry (MBR) between 1 January 1987 and 31 December 1993. The MBR collects detailed and standardised information on nearly all pregnancies in Sweden culminating in delivery.[23 24] Using a unique personal identification number, the Swedish authorities (National Board of Health and Welfare and Statistics Sweden) linked the MBR to the Register of the Total Population and the Swedish 1990 population census and created a research database. This database was delivered to us without the personal identification numbers to protect the anonymity of the participants.

For the purpose of our study, we selected singletons, because it is known that multiple births (n=19 167) have a different intrauterine growth pattern from gestational

weeks 28–30.[25] We excluded 13 539 babies born with significant congenital anomalies according to the MBR. Following previously established criteria,[26] we also excluded babies with inconsistent information on birth weight according to gestational age (n=9195) and babies weighing less than 500 g (n=51) as well as 15 observations with missing information on maternal age and birth order. The final sample contained 768 059 babies. Thereafter, we stratified the population by gestational age into preterm (<37 gestational weeks), term (≥37 and <42 gestational weeks) and post-term babies (≥42 gestational weeks; figure 1).

Variables

The outcome variable combined birth weight and gestational age to dichotomise as being SGA or not, and using the last category as the reference. This variable was available at the MBR, where it is routinely calculated following standard intrauterine growth curves.[27] Infants were defined as SGA if they weighed less than 2 SDs below the expected birth weight for gestational age and gender, according to a Swedish intrauterine growth curve.[28]

In our analyses we included child and maternal characteristics that are known to be associated with low birthweight and SGA.

As child characteristics we used *sex*[14 29] and *birth order*[30 31] classified into three categories (ie, firstborn, second, and third or more). Among maternal characteristics we included *birth interval between newborns*,[14 28] categorised into <1, 1–2, >2 years, 'only child' (ie, when we know that the newborns have previous siblings but we do not have their information in our setting) and first child (ie, when we know the newborn has no previous siblings); whether the mother has a *previous child with SGA*[32] categorised into yes, no, 'only child' and first child; *education*,[33 34] categorised into low (primary education or less), middle (secondary school) and high education (graduate and PhD); *marital status*,[35 36] categorised into single, widowed, or divorced, and married or cohabiting; and *maternal age at delivery*,[37–39] categorised into four groups (ie, <20, 20–24, 25–34 and >35 years old), as well as information on *smoking habits*,[40–43] categorised into non-smoking, light smoking (fewer than 9 cigarettes per day), heavy smoking (more than 9 cigarettes per day), and missing information. Finally, we included information about the presence of *hypertension during pregnancy* (yes vs no),[15 17] and *maternal origin*, classified as being born in Sweden or not.[44]

STATISTICAL METHODS

To examine the average association between, on the one hand, the categorical variables mentioned above, and on the other, being SGA, we simply calculated ORs and 95% CIs obtained from logistic regression analyses.

The discriminatory accuracy of a risk factor is better appraised by measuring *the true positive fraction* (TPF)

Figure 1 Flow diagram showing the individuals excluded from the study population.
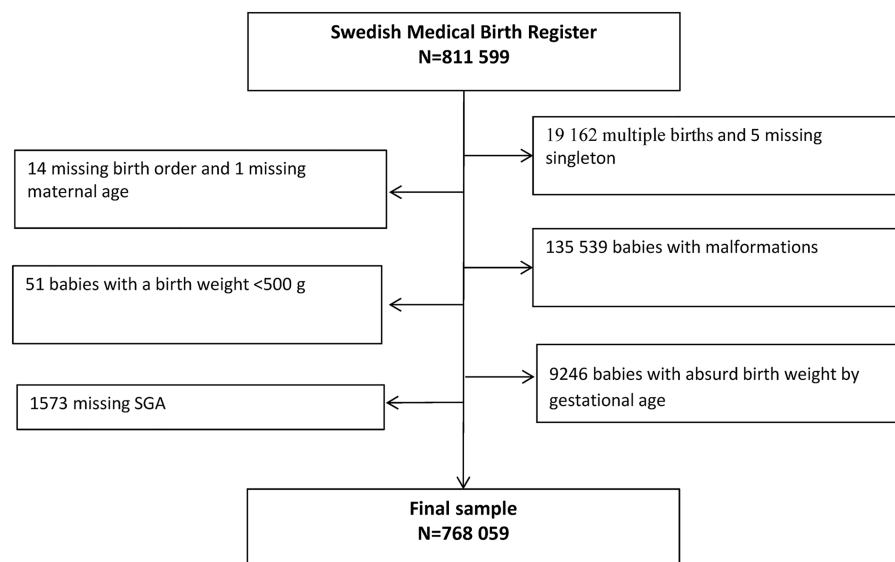
and *the false positive fraction* (FPF). For a dichotomous risk factor, the TPF expresses the probability of being exposed to the risk factor when the SGA occurs (ie, cases that are exposed to the risk factor), and the FPF indicates the probability of being exposed to the risk factor when the SGA does not occur (ie, controls exposed to the risk factor). In the ideal scenario the TPF should be 1 and the FPF should be 0, even if a lower TPF or a higher FPF. For instance, if the identification of the risk factors conveys pharmacological treatment, we should try to keep the FPF as low as possible.

For the evaluation of the discriminatory accuracy of the combination of risk factors within a risk score (ie, predicted probability) we obtained *the receiver operating characteristic* (ROC) *curve*. The ROC curve is constructed by plotting the TPF against the FPF for different risk score thresholds.[16 45 46] A traditional measure of discriminatory accuracy is the *area under the ROC curve* (AU-ROC) or C statistic.[16 45 47–49] The AU-ROC extends from 0.5 to 1.0. An AU-ROC=0.5 means that the discriminatory accuracy of the candidate risk factor is similar to that obtained by flipping a coin. That is, a risk factor with an AU-ROC=0.5 is useless. An AU-ROC=1.0 means complete accuracy.

In a series of simple logistic regression models we identified the single variables with the highest discriminatory accuracy. Using this information, thereafter, we performed two models. Model A only with the two variables with the higher discriminatory accuracy (ie, smoking and birth order) and model B which adds the rest of covariates to the initial model A. We ran this second model in order to assess the change of discriminatory accuracy when adding the rest of information to a simpler model. We appraised the incremental value of a model by the difference between AU-ROCs. Owing to a problem of collinearity, Stata automatically deleted the two categories in common (ie, 'only child' and first child) shared by the variables of birth interval and previous child with SGA,

keeping them only in the former. All models were stratified by gestational age (ie, preterm, term and post-term) because it has been suggested that SGA at term and at preterm may have been driven by a different aetiology.[50] We included post-term to complete the classification.

We performed the analyses in the whole population, stratifying by gestational age (ie, preterm, term and post-term). We performed the statistical analyses using STATA V.12.0 (College Station, StataCorp LP, Texas, USA) and SPSS V.21.0 (Armonk, IBM Corp, New York, USA).

RESULTS
Table 1 shows the maternal and individual characteristics of the population of newborns by the SGA status. We see that SGA is much more prevalent among preterm babies (10.14%) than among term (1.87%) and post-term (3.03%) babies. Females show higher prevalence of SGA than males among preterm, and slightly lower prevalence among those born post-term. Regardless of gestational age, firstborns had a higher risk of SGA than their siblings. SGA is more prevalent among children who had a previous sibling during the same year, except among those babies born at preterm, but this may be due to the larger amount of missing information about the previous siblings (11%). Mothers who had a previous child with SGA are more likely to have a current SGA baby regardless of gestational age.

SGA was more frequent in mothers younger than 20 years of age, among divorced, widowed and single women, and among those who were born outside Sweden and those with low educational achievement. In babies born with SGA, hypertension was more frequent among preterm than among post-term babies.

Table 2 indicates that the risk for being SGA was similar in boys and girls. However, as expected, not being a firstborn reduced the risk of being SGA. With respect to maternal characteristics, mothers younger

Table 1 Prevalence of SGA in the whole population of babies and in strata of gestational age in Sweden 1987–1993

	SGA		SGA stratified by gestational age					
			Term		Preterm		Post-term	
	768 059 (2.3%)		676 961 (1.9%)		36 080 (10.1%)		55 018 (3.0%)	
	N	Per cent	N	Per cent	N	Per cent	N	Per cent
Newborn babies								
Male	393 538	2.33	343 549	1.86	19 635	9.20	30 354	3.24
Female	374 521	2.34	333 412	1.87	16 445	11.26	24 664	2.78
Birth order								
First	325 326	3.30	279 426	2.66	18 796	11.67	27 104	4.10
Second	268 998	1.63	242 276	1.31	9 774	8.79	16 948	2.01
≥Third	173 735	1.63	155 259	1.30	7,510	8.07	10 866	1.97
Birth intervals								
<1 year	4557	3.55	3841	2.76	415	6.99	301	8.97
1–2 years	77 922	2.00	69 232	1.61	3585	8.23	5105	2.86
>2 years	147 354	1.97	130 979	1.57	5970	9.45	10 405	2.72
'Only child'	341 906	2.77	298 264	2.20	17 789	11.63	25 853	3.29
First child	196 320	1.97	174 645	1.60	8321	8.45	13 354	2.70
Previous child SGA								
Yes	3371	9.58	2913	7.69	259	30.12	196	9.18
No	226 462	1.90	201 139	1.52	9711	8.34	15 615	2.80
'Only child'	341 906	2.77	298 264	2.20	17 789	11.63	25 853	3.29
First child	196 320	1.97	174 645	1.60	8321	8.45	13 354	2.70
Maternal age								
<20	40 735	3.06	35 367	2.49	2485	9.93	2883	4.30
20–24	156 736	2.39	138 232	1.91	7842	9.82	10 662	3.24
25–34	481 689	2.19	425 980	1.76	20 796	9.59	34 913	2.89
>35	88 899	2.69	77 382	2.08	4957	12.04	6 560	2.90
Hypertension	3166	8.69	2648	5.44	383	32.11	135	5.93
Smoking								
Non-smoker	541 962	1.77	479 851	1.37	22 603	9.14	39 508	2.41
Light smoker	109 799	3.62	76 327	3.05	5884	11.86	7588	4.40
Heavy smoker	67 454	4.41	58 790	3.79	4234	11.76	4430	5.64
Missing	48 844	2.90	41 993	2.12	3359	11.79	3492	3.78
Marital status								
Married or cohabiting	385 173	2.02	342 330	1.61	16,50	9.36	26 343	2.69
Divorced or widowed	27 151	3.03	23 475	2.43	1706	10.67	1970	3.55
Single	355 735	2.63	311 156	2.10	17 874	10.81	26 705	3.33
Maternal origin								
Born in Sweden	663 043	2.25	584 683	1.79	30 745	10.19	47 615	2.79
Not born in Sweden	95 393	2.83	84 022	2.30	4749	10.07	6 622	4.27
Missing	9623	2.67	8256	3.02	586	8.02	781	7.30
Maternal education								
University	181 083	1.95	160 102	1.52	7367	10.26	13 614	2.57
Secondary	400 363	2.22	353 813	1.77	18 457	9.96	28 093	2.74
Primary	140 186	3.21	122 919	2.31	7631	10.78	9636	3.84
Missing	46 427	2.34	40 127	2.67	2625	9.22	3675	4.82

SGA, small for gestational age.

than 20 years and those 35 years and older had a higher risk of delivering an SGA baby than 20–24-year-old mothers. Mothers who had a previous child during the same year have a higher risk of having a SGA baby as well as those who had a previous child with SGA. Mothers who experienced hypertension during pregnancy had a higher risk of delivering SGA babies. Compared with non-smoker mothers, light and heavy smoker mothers had a higher risk of delivering an SGA baby. Divorced and widowed mothers as well as single mothers were more likely to deliver an SGA baby than married and cohabiting mothers. Mothers with primary and secondary education had a higher risk of delivering SGA babies than mothers with a university degree. Similarly, mothers who were not born in Sweden were at higher risk of delivering an SGA baby.

Figure 2 shows the values for the AU-ROC of the variables included in table 2. Overall, their discriminatory accuracy was rather low. Newborn babies had the lower discriminatory accuracy. Having a SGA child and

Table 2 Measures of association between offspring and maternal characteristics, and being small for gestational age (SGA), in the whole population of babies and in strata of gestational age in Sweden 1987–1993

|	SGA Unadjusted model			SGA term Unadjusted model			SGA preterm Unadjusted model			SGA post-term Unadjusted model		
	OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI	
Newborn babies												
Male (ref)												
Female	1.01	(0.98	1.04)	1.01	(0.97	1.04)	0.86	(0.84	0.88)	0.85	(0.77	0.94)
Birth order												
First (ref)												
Second	0.48	(0.47	0.50)	0.49	(0.47	0.51)	0.60	(0.59	0.62)	0.48	(0.42	0.54)
≥Third	0.49	(0.47	0.51)	0.48	(0.46	0.50)	0.72	(0.70	0.74)	0.47	(0.41	0.54)
Gestational age	0.80	(0.79	0.81)	0.87	(0.85	0.88)	0.86	(0.85	0.87)	1.77	(1.61	1.95)
Birth intervals												
1–2 years (ref)												
<1 year	1.81	(1.53	2.13)	1.74	(1.42	2.12)	0.84	(0.56	1.24)	3.35	(2.18	5.13)
>2 years	0.98	(0.92	1.05)	0.97	(0.90	1.05)	1.16	(1.00	1.35)	0.95	(0.77	1.16)
'Only child'	1.40	(1.32	1.05)	1.37	(1.29	1.46)	1.47	(1.29	1.67)	1.16	(0.97	1.38)
First child	0.99	(0.93	1.05)	0.99	(0.93	1.07)	1.03	(0.89	1.19)	0.94	(0.78	1.15)
Previous child SGA												
No (ref)												
Yes	5.48	(4.87	6.17)	5.45	(4.74	6.28)	4.75	(3.61	6.24)	3.95	(2.49	6.42)
'Only child'	1.47	(1.42	1.53)	1.46	(1.40	1.52)	1.45	(1.33	1.58)	1.18	(1.05	1.33)
First child	1.04	(0.99	1.09)	1.06	(1.01	1.12)	1.01	(0.91	1.13)	0.97	(0.84	1.11)
Maternal age												
25–34 (ref)												
<20	1.41	(1.33	1.49)	1.42	(1.33	1.53)	1.44	(1.38	1.50)	1.51	(1.25	1.83)
20–24	1.09	(1.05	1.13)	1.09	(1.04	1.14)	1.16	(1.13	1.19)	1.12	(0.99	1.27)
>35	1.23	(1.18	1.29)	1.18	(1.12	1.25)	1.31	(1.27	1.36)	1.00	(0.86	1.17)
Hypertension												
No (ref) vs yes	4.02	(3.55	4.55)	3.05	(2.58	3.61)	2.73	(2.45	3.04)	2.02	(0.99	4.13)
Smoking												
Non-smoker (ref)												
Light smoker	2.08	(2.01	2.16)	2.27	(2.17	2.37)	1.30	(1.26	1.33)	1.86	(1.64	2.12)
Heavy smoker	2.56	(2.46	2.67)	2.84	(2.70	2.98)	1.53	(1.48	1.58)	2.42	(2.10	2.79)
Missing	1.66	(1.57	1.76)	1.56	(1.45	1.67)	1.70	(1.64	1.76)	1.59	(1.32	1.92)
Marital status												
Married or cohabiting (ref)												
Divorced or widowed	1.52	(1.41	1.63)	1.52	(1.40	1.66)	1.51	(1.43	1.59)	1.33	(1.04	1.71)
Single	1.31	(1.27	1.35)	1.31	(1.27	1.36)	1.19	(1.17	1.22)	1.25	(1.13	1.38)
Maternal origin												
Born in Sweden (ref)												
Not born in Sweden	1.27	(1.21	1.32)	1.30	(1.23	1.36)	1.07	(1.04	1.11)	1.56	(1.37	1.77)
Missing	1.66	(1.49	1.84)	1.71	(1.51	1.94)	1.35	(1.24	1.47)	2.74	(2.08	3.61)
Maternal education												
University (ref)												
Secondary	1.14	(1.10	1.19)	1.17	(1.12	1.23)	1.13	(1.10	1.17)	1.07	(0.94	1.22)
Primary	1.49	(1.42	1.56)	1.53	(1.45	1.62)	1.35	(1.31	1.39)	1.51	(1.30	1.76)
Missing	1.67	(1.57	1.77)	1.78	(1.66	1.92)	1.42	(1.36	1.49)	1.92	(1.59	2.31)

Values are ORs and (95% CI). Crude models.
SGA, small for gestational age.

hypertension, despite the risk factors being most strongly associated with SGA (OR 5.48 and 4.02, respectively), led to a very low discriminatory accuracy (AU-ROC 0.54 and 0.51). Birth order and smoking were the variables with the highest accuracy (AU-ROC 0.59).

Figure 3 shows the AU-ROC for SGA of different risk factors after stratification by preterm, term and post-term. As in the non-stratified analysis, the discriminatory accuracy of the variables was low. Smoking at term showed the highest discrimination (AU-ROC 0.60).

Figure 4 shows that the discriminatory accuracy of the general model A, including only birth order and smoking, was slightly improved (just 0.05 proportion units), when all variables were included in the full model B. Among preterm babies model B improved the discriminatory accuracy of the model by 0.1 proportion
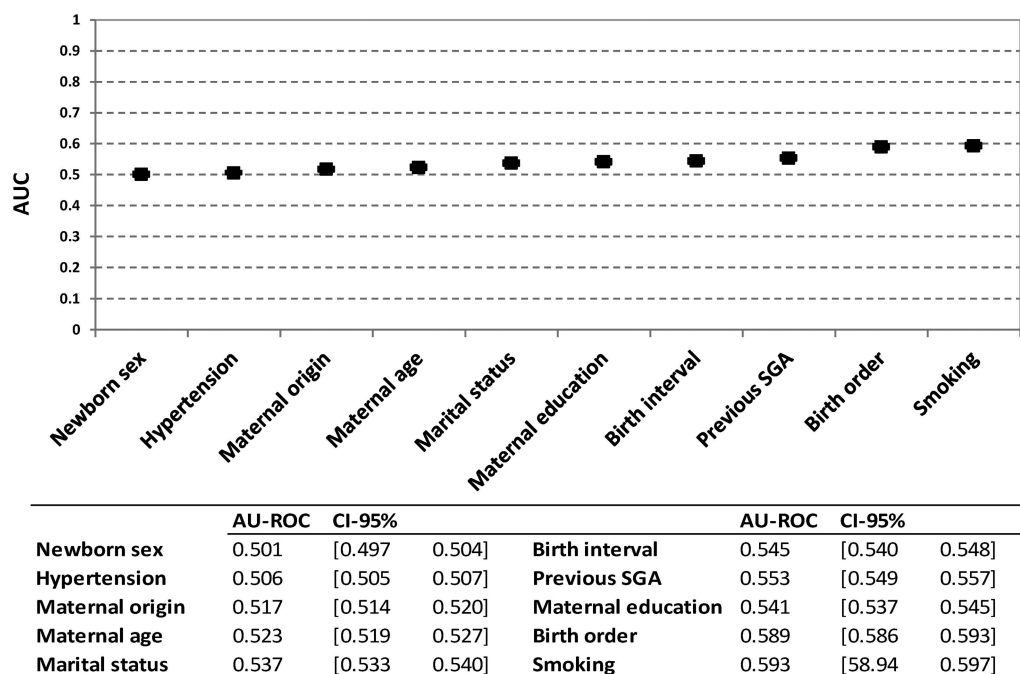
Figure 2   Area under the receiver operating characteristic curve to compare the discriminatory accuracy of different models to distinguish between small for gestational age (SGA) and non-SGA babies.

	AU-ROC	CI-95%			AU-ROC	CI-95%	
Newborn sex	0.501	[0.497	0.504]	**Birth interval**	0.545	[0.540	0.548]
Hypertension	0.506	[0.505	0.507]	**Previous SGA**	0.553	[0.549	0.557]
Maternal origin	0.517	[0.514	0.520]	**Maternal education**	0.541	[0.537	0.545]
Maternal age	0.523	[0.519	0.527]	**Birth order**	0.589	[0.586	0.593]
Marital status	0.537	[0.533	0.540]	**Smoking**	0.593	[58.94	0.597]

units, while this improvement was much lower among SGA term and SGA post-term babies.

DISCUSSIONS

We were able to verify a number of recognised maternal and newborn risk factors for SGA. For instance, we found that smoking (OR 2.56) and especially having had a previous SGA baby (OR 5.48) and maternal hypertension (OR 4.02) were 'strongly' associated with being SGA. However, even if the magnitude of the ORs was of a size normally considered as undoubtedly relevant in epidemiology, none of those traditional risk factors for SGA provided enough accuracy to discriminate babies with SGA from other babies. In fact, the AU-ROC for having had a previous SGA child and maternal hypertension was slightly higher than 0.5, which means that the accuracy of this variable for discriminating babies with SGA from those without SGA was rather similar to that obtained by flipping a coin. That is, we need to recognise that, although on average, mothers with hypertension were four times more likely to have an SGA baby, many mothers with hypertension delivered babies without SGA, and many SGA babies were born to mothers without hypertension. Our findings, therefore, seriously question the utility of maternal hypertension during pregnancy for planning strategies of prevention against SGA. This statement, however, does not mean that hypertension during pregnancy is irrelevant to understanding the origin of SGA, but rather that we need to determine who among hypertensive mothers is actually prone to deliver an SGA baby.

There is a tacit but fallacious belief that the discriminatory accuracy of a risk factor is high when it is supported by a 'strong' association (eg, an OR of 4, as in the case of maternal hypertension). However, for an association to be an accurate instrument for discrimination, it must be of a magnitude rarely identified in epidemiological studies.[16 51–53] Following our example, a low discriminatory accuracy only indicates that any attempt of intervention based on the existence of the risk factor will be inefficient and even inappropriate, because health professionals will unnecessarily treat many mothers. The decision to start an intervention should seriously take into account the existence of important (physical or emotional) side effects in the false-positive women. That is, it is always important to consider the principle of *primum non nocere*.[22]

Compared with the other variables studied, birth order and smoking presented a higher discriminatory capacity. However, their discriminatory accuracy was still very low in absolute terms (AU-ROC≈0.59). Also, combining all the variables in the same model did not substantially increase the discriminatory accuracy (AU-ROC=0.69). In other words, our results indicate that we actually do not know so much about what determines being SGA.

The existence of a low discriminatory accuracy suggests that around the population average risk there is considerable individual heterogeneity. Therefore, a logical consequence should be to identify which women are most susceptible to the risk factors. Hence, we explored the discriminatory accuracy of the chosen risk factors in different strata of gestational age at birth. We
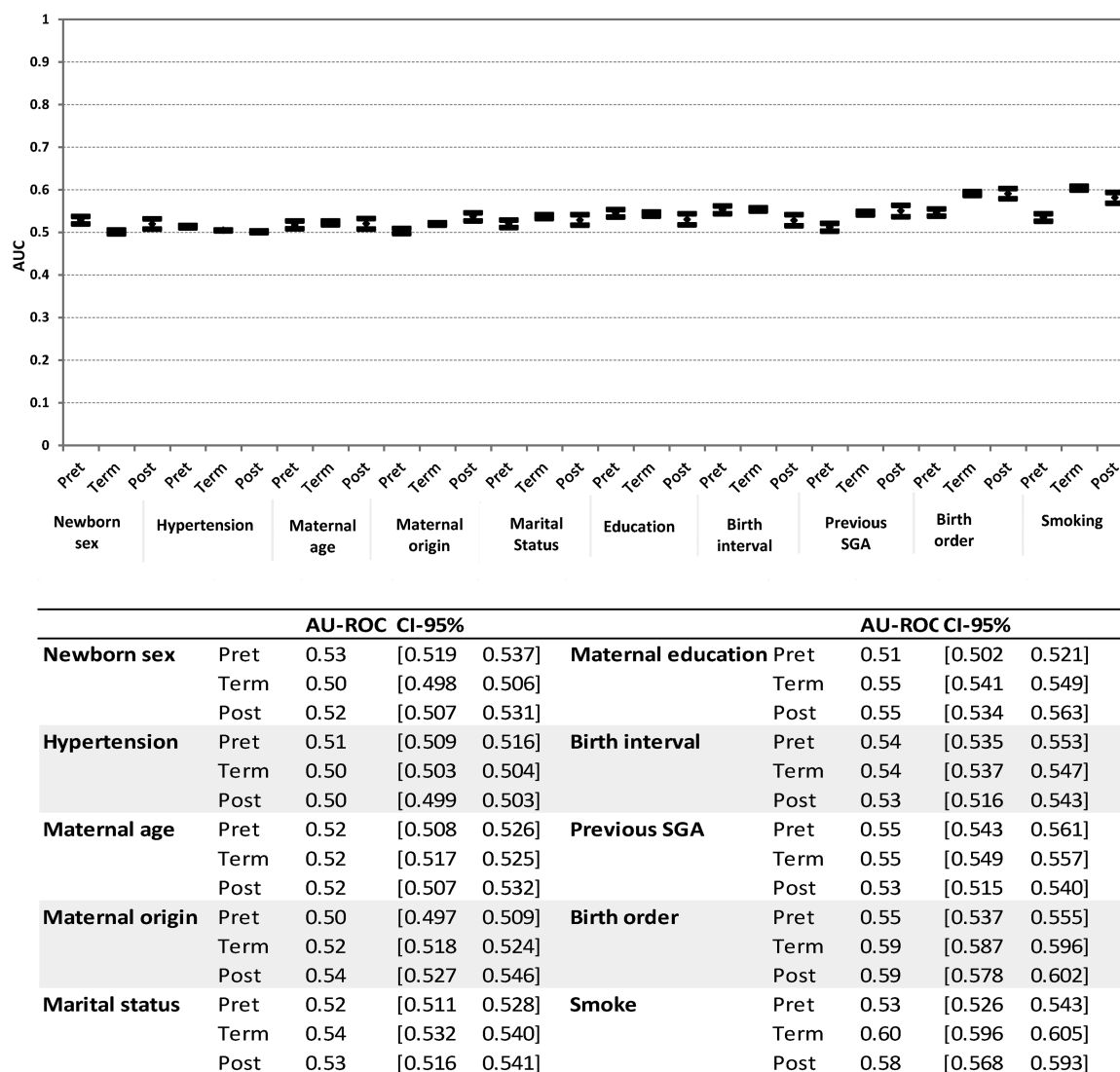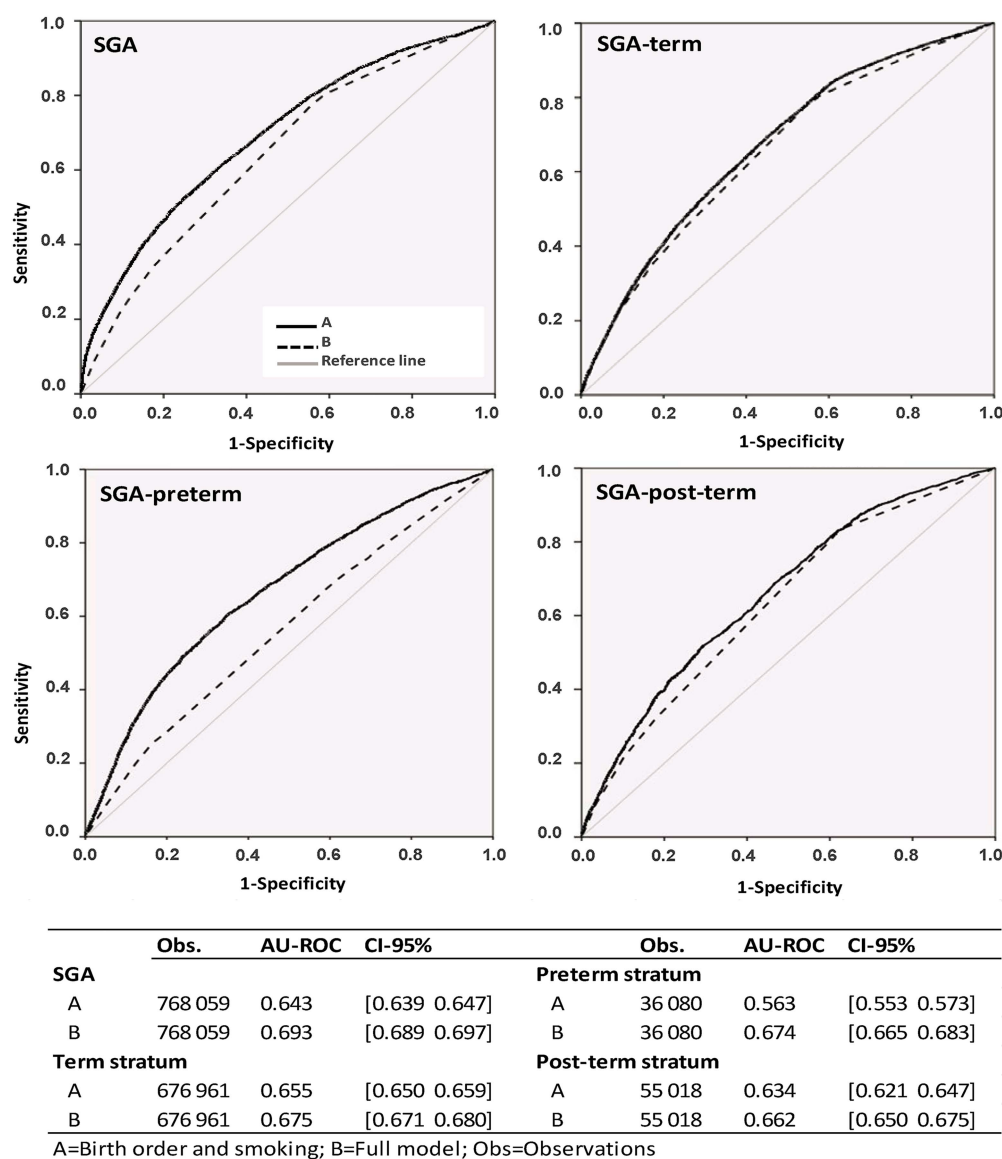
Figure 3 Area under the receiver operating characteristic curve for specific maternal and newborn characteristics.

		AU-ROC	CI-95%				AU-ROC	CI-95%	
Newborn sex	Pret	0.53	[0.519	0.537]	**Maternal education**	Pret	0.51	[0.502	0.521]
	Term	0.50	[0.498	0.506]		Term	0.55	[0.541	0.549]
	Post	0.52	[0.507	0.531]		Post	0.55	[0.534	0.563]
Hypertension	Pret	0.51	[0.509	0.516]	**Birth interval**	Pret	0.54	[0.535	0.553]
	Term	0.50	[0.503	0.504]		Term	0.54	[0.537	0.547]
	Post	0.50	[0.499	0.503]		Post	0.53	[0.516	0.543]
Maternal age	Pret	0.52	[0.508	0.526]	**Previous SGA**	Pret	0.55	[0.543	0.561]
	Term	0.52	[0.517	0.525]		Term	0.55	[0.549	0.557]
	Post	0.52	[0.507	0.532]		Post	0.53	[0.515	0.540]
Maternal origin	Pret	0.50	[0.497	0.509]	**Birth order**	Pret	0.55	[0.537	0.555]
	Term	0.52	[0.518	0.524]		Term	0.59	[0.587	0.596]
	Post	0.54	[0.527	0.546]		Post	0.59	[0.578	0.602]
Marital status	Pret	0.52	[0.511	0.528]	**Smoke**	Pret	0.53	[0.526	0.543]
	Term	0.54	[0.532	0.540]		Term	0.60	[0.596	0.605]
	Post	0.53	[0.516	0.541]		Post	0.58	[0.568	0.593]

found that the combination of all variables in the same model had a minor improvement for discriminatory accuracy among those born at term or preterm as well as post-term.

Our finding suggests the existence of individual heterogeneity of responses to some specific variables, so the discriminatory accuracy depended, for instance, on whether the baby was preterm, term or post-term. In fact, smoking, birth order, maternal origin and marital status had a lower discriminatory capacity among preterm than among babies at term. On the contrary, newborn babies and hypertension had a higher discriminatory accuracy among preterm babies than among term babies. In the same way, newborn babies and maternal origin had a lower discriminatory capacity among term than among post-term babies, but for smoking and having had a previous SGA baby we found the opposite relationship. The variation of the magnitude of the discriminatory accuracy by gestational age at birth expresses the existence of individual heterogeneity.

In addition, the definition of SGA may also actively contribute to reducing the discriminatory accuracy of the traditional risk factors since discrimination depends on the outcome and exposure. Thus, low discrimination can result from the fact that SGA fails to distinguish between pathological and constitutionally small babies, that is, to properly capture the health dimension that it is supposed to be a proxy for IUGR.[1] In order to address this shortcoming, we stratified SGA by gestational age as this has been identified as a good strategy to distinguish between these two.[50] However, we do not find support for this approach since we found a lower discriminatory accuracy among preterm SGA babies (presumably pathologically small) than among term SGA babies. In this regard, our findings show awareness of the caveats pointed out by previous studies on the use of SGA as a proxy for IUGR,[1] and encourage further research aiming to better capture IUGR.

Our findings have important research and policy-making implications. A possible reason for the low

	Obs.	AU-ROC	CI-95%		Obs.	AU-ROC	CI-95%
SGA				**Preterm stratum**			
A	768 059	0.643	[0.639 0.647]	A	36 080	0.563	[0.553 0.573]
B	768 059	0.693	[0.689 0.697]	B	36 080	0.674	[0.665 0.683]
Term stratum				**Post-term stratum**			
A	676 961	0.655	[0.650 0.659]	A	55 018	0.634	[0.621 0.647]
B	676 961	0.675	[0.671 0.680]	B	55 018	0.662	[0.650 0.675]

A=Birth order and smoking; B=Full model; Obs=Observations

Figure 4 Area under the receiver operating characteristic curve for specific maternal and newborn characteristics after stratifying by gestational age (preterm, term and post-term).

discriminatory accuracy of many average associations is that average effects are a mixture of individual level effects and therefore mix inter-individual heterogeneity (ie, some individuals respond intensively to the exposure, while others are resilient or might even respond in the opposite direction). The approach based on discriminatory accuracy understands average effects as an idealised mean value that does not necessarily represent the heterogeneity of individual effects.[22] Some scholars prefer to conceive individual outcome as the expression of a stochastic phenomenon that is best estimated by the average risk using a probabilistic approach.[54] Our understanding instead is that individual outcome reflects the interindividual heterogeneity of responses that can be potentially determined; lack of knowledge could be amended by a better understanding of individual responses.[55] See elsewhere for a better explanation of

these ideas.[22 56 57] From this perspective, reducing exposure to a risk factor would only be effective when acting on the susceptible, but not on the resilient, individuals. For instance, we need to better capture babies who suffer from IUGR, since, so far, we have been incapable of distinguishing between babies who are constitutionally small from those who are pathologically growth restricted.[1] By stratifying between preterm, term, and post-term, we might be able to better approach the underlying heterogeneity.

From the policymaking perspective, our findings suggest that hitherto there has not been enough knowledge to identify any specific risk factor or combination of them that could discriminate with accuracy children with and without the SGA status. Our findings support policymaking oriented to lifestyle modification, as according to the principle of *primum non nocere*,[22] they

have mostly positive consequences, even for 'false-positive' mothers. For instance, persuading women to quit smoking reduces the risk of SGA in some babies, but it improves general well-being in everyone. However, other risk factors with low discriminatory accuracy that lead to pharmacological treatment or screening might result in unnecessary side effects and cost. In the long run, an uncritical use of variables with low discriminatory accuracy may hinder the identification of pertinent risk factors and susceptible individuals and damage the scientific credibility of modern epidemiology.[22 56 57]

Our conclusions are based on classical measures of discriminatory accuracy such as the AU-ROC curve. These measures have been criticised as insensitive to small changes in predicted individual risk.[58] Some authors propose more specific measures of reclassification, like the net reclassification improvement (NRI), and the integrated discrimination improvement (IDI).[59–62] We applied NRI and IDI in a sensitivity analysis (results not shown in tables). For example, using NRI, we observed a reclassification of 30%. However, this figure does not add substantial information to our results, since NRI (as well as IDI) refers to the misclassification occurring all along the risk scale, instead of capturing the misclassification which takes place around the fixed threshold. Furthermore, the new NRI and IDI measures have also been criticised,[63] and some authors[64] have explicitly advised against their use in common epidemiological practice because, unlike IDI and NRI, traditional measures of discrimination like the AU-ROC curve have the advantage that prognostic performance cannot be manipulated.[64] Therefore, we preferred to quantify discriminatory accuracy by analysing ROC curves and AU-ROCs.

Our analyses are based on a national medical registry covering almost the entire population of residents in Sweden. Nearly all births are registered in the MBR, because giving birth at home is very unusual in Sweden. In addition, estimation of SGA is routinely calculated at the MBR following standard intrauterine growth curves.[27] However, our study also has a number of limitations. Because of lack of data, we could not assess many other variables identified in the literature as 'risk factors', such as genetic or nutritional factors.[15 65] In spite of the quality of the MBR, the information regarding smoking is based on a self-reported questionnaire (anamnesis) administered by the mid-wife at the first antenatal visit (ie, between 10 and 12 gestational weeks), which to some extent might bias the result by including misclassification of exposure.[66] However, a study conducted in Sweden comparing self-reported nicotine exposure and plasma levels of cotinine in early and late pregnancy concluded that self-reported smoking information had acceptable validity.[67]

Unfortunately, we could not identify those mothers who suffered from preeclampsia, for which the discriminatory accuracy concerning SGA may be higher than for hypertension in our model. Further analysis on this aspect is required. Another limitation of our study is that we calculated the discriminatory accuracy in the same sample used for constructing the predicted model. This procedure, however, might overestimate the discriminatory accuracy of the models, so the low discriminatory accuracy found may be an underestimation.

Since our study has been carried out with data from 1987 to 1993, we performed a sensitivity analysis to check possible differences in current years (2000–2010) but the results remain the same. Given the consistence of the results, we preferred to maintain the results for years 1987–1993 to use a more accurate outcome since the definition of SGA is based on standard curves estimated for Sweden with data from 1985 to 1989.[29] Moreover, the period we cover is of relevance to our study since most of the risk factors which are discussed in our paper were mainly identified in that period and, in the case of Sweden, with the data we used.

In conclusion, applying measures of discriminatory accuracy rather than measures of association only, our study revisits known risk factors of SGA and discusses their role from a public health perspective. We found that neither models including simple variables nor models including several variables at the same time have a good discriminatory accuracy to discriminate babies with SGA from those without SGA. This finding is of fundamental relevance to address future research and to design policy-making recommendations in a more informed way.

As noted elsewhere,[56 57] there is need of a new epidemiological approach that systematically provides information on the discriminatory accuracy and interindividual heterogeneity of effects and does not rely only on average measures of association.[68] In this line, new statistical methods like logic regression seem promising.[69 70] A fundamental change is needed in the way traditional risk factors are currently interpreted in public health epidemiology. If the *discriminatory accuracy* of most classical risk factors is very low, what happens with the vast majority of recommendations given so far in epidemiology and public health? Are health professionals misleading the community by raising the alarm about risks that may be harmless for most individuals? What are the ethical repercussions of using risk factors with low discriminatory accuracy? Are there problems of inefficiency, medicalisation and stigmatisation? We believe that these questions have a high significance for both the community and the future of public health research.

REFERENCES

1. Urquia ML, Ray JG. Seven caveats on the use of low birthweight and related indicators in health research. *J Epidemiol Community Health* 2012;66:971–5.
2. Grisaru-Granovsky S, Reichman B, Lerner-Geva L, *et al*. Mortality and morbidity in preterm small-for-gestational-age infants: a population-based study. *Am J Obstet Gynecol* 2012;206:150e151–6.
3. Streimish IG, Ehrenkranz RA, Allred EN, *et al*. Birth weight- and fetal weight-growth restriction: impact on neurodevelopment. *Early Hum Dev* 2012;88:765–71.
4. Eikenes L, Martinussen MP, Lund LK, *et al*. Being born small for gestational age reduces white matter integrity in adulthood: a prospective cohort study. *Pediatr Res* 2012;72:649–54.
5. Moore GS, Kneitel AW, Walker CK, *et al*. Autism risk in small- and large-for-gestational-age infants. *Am J Obstet Gynecol* 2012;206:314.
6. Nielsen PR, Mortensen PB, Dalman C, *et al*. Fetal growth and schizophrenia: a nested case-control and case-sibling study. *Schizophr Bull* 2013;39:1337–42.
7. Løhaugen GC, Østgård HF, Andreassen S, *et al*. Small for gestational age and intrauterine growth restriction decreases cognitive function in young adults. *J Pediatr* 2013;163:447–53.
8. Wingren CJ, Agardh D, Merlo J. Revisiting the risk of celiac disease in children born small for gestational age: a sibling design perspective. *Scand J Gastroenterol* 2012;47:632–9.
9. van de Lagemaat M, Rotteveel J, van Weissenbruch MM, *et al*. Small-for-gestational-age preterm-born infants already have lower bone mass during early infancy. *Bone* 2012;51:441–6.
10. Forssell L, Cnattingius S, Bottai M, *et al*. Increased risk of Barrett's esophagus among individuals born preterm or small for gestational age. *Clin Gastroenterol Hepatol* 2013;11:790–4.
11. Forssell L, Cnattingius S, Bottai M, *et al*. Risk of esophagitis among individuals born preterm or small for gestational age. *Clin Gastroenterol Hepatol* 2012;10:1369–75.
12. Dalla Pozza RD, Bechtold S, Putzker S, *et al*. Young adults born small for gestational age: is reduced baroreceptor sensitivity a risk factor for hypertension? *Clin Cardiol* 2006;29:215–18.
13. Crispi F, Figueras F, Cruz-Lemini M, *et al*. Cardiovascular programming in children born small for gestational age and relationship with prenatal signs of severity. *Am J Obstet Gynecol* 2012;207:121.e121–9.
14. Kramer MS. Determinants of low birth weight: methodological assessment and meta-analysis. *Bull World Health Organ* 1987;65:663–737.
15. McCowan L, Horgan RP. Risk factors for small for gestational age infants. *Best Pract Res Clin Obstet Gynaecol* 2009;23:779–93.
16. Pepe MS, Janes H, Longton G, *et al*. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol* 2004;159:882–90.
17. Catov JM, Nohr EA, Olsen J, *et al*. Chronic hypertension related to risk for preterm and term small for gestational age births. *Obstet Gynecol* 2008;112:290–6.
18. Zethelius B, Berglund L, Sundstrom J, *et al*. Use of multiple biomarkers to improve the prediction of death from cardiovascular causes. *N Engl J Med* 2008;358:2107–16.
19. Cooney MT, Dudina AL, Graham IM. Value and limitations of existing scores for the assessment of cardiovascular risk: a review for clinicians. *J Am Coll Cardiol* 2009;54:1209–27.
20. Law MR, Wald NJ, Morris JK. The performance of blood pressure and other cardiovascular risk factors as screening tests for ischaemic heart disease and stroke. *J Med Screen* 2004;11:3–7.
21. Melander O, Newton-Cheh C, Almgren P, *et al*. Novel and conventional biomarkers for prediction of incident cardiovascular events in the community. *JAMA* 2009;302:49–57.
22. Merlo J. Multilevel analysis of individual heterogeneity: a fundamental critique of the current probabilistic risk factor epidemiology (invited commentary). *Am J Epidemiol* 2014;180:208–12.
23. National Board of Health and Welfare, Centre for Epidemiology. Fakta om mammor, förlossningar och nyfödda barn (Facts about Mothers, Births and Newborn Children): Medicinska Födelseregistret 1973–2000. Medical Birth Register 1973–2000 (in Swedish) (No. 2002-125-12). National Board of Health and Welfare, Stockholm, 2002.
24. Cnattingius S, Ericson A, Gunnarskog J, *et al*. A quality study of a medical birth registry. *Scand J Soc Med* 1990;18:143–8.
25. Cunningham FG. *Williams obstetrics*. 3rd edn. McGraw-Hill, 2005.
26. Källén B. A birth weight for gestational age standard based on data in the Swedish Medical Birth Registry, 1985–1989. *Eur J Epidemiol* 1995;11:601–6.
27. Marsál K, Persson PH, Larsen T, *et al*. Intrauterine growth curves based on ultrasonically estimated foetal weights. *Acta Paediatr* 1996;85:843–8.
28. Grisaru Granovsky S, Gordon ES, Haklai Z, *et al*. Effect of interpregnancy interval on adverse perinatal outcomes—a national study. *Contraception* 2009;80:512–18.
29. Alexander GR, Kogan MD, Himes JH. 1994–1996 U.S. singleton birth weight percentiles for gestational age by race, Hispanic origin, and gender. *Matern Child Health J* 1999;3:225–32.
30. Swamy GK, Edwards MJ, Gelfand A, *et al*. Maternal age, birth order, and race: differential effects on birthweight. *J Epidemiol Community Health* 2012;66:136–42.
31. Elshibly EM, Schmalisch G. The effect of maternal anthropometric characteristics and social factors on gestational age and birth weight in Sudanese newborn infants. *BMC Public Health* 2008;8:244.
32. Voskamp B, Kazemier BM, Ravelli AC, *et al*. Recurrence of small-for-gestational-age pregnancy: analysis of first and subsequent singleton pregnancies in The Netherlands. *Am J Obstet Gynecol* 2013;208:374.e371–6.
33. Kramer MS, Séguin L, Lydos J, *et al*. Socio-economic disparities in pregnancy outcome: why do the poor fare so poorly? *Paediatr Perinat Epidemiol* 2000;14:194–210.
34. Auger N, Park AL, Harper S, *et al*. Educational inequalities in preterm and term small-for-gestational-age birth over time. *Ann Epidemiol* 2012;22:160–7.
35. Shah PS, Zao J, Ali S, Knowledge Synthesis Group of Determinants of preterm/LBW births. Maternal marital status and birth outcomes: a systematic review and meta-analyses. *Matern Child Health J* 2011;15:1097–109.
36. Castro T. Single motherhood and low birthweight in Spain: narrowing social inequalities in health? *Demogr Res* 2010;22:863–90.
37. Fraser AM, Brockert JE, Ward RH. Association of young maternal age with adverse reproductive outcomes. *N Engl J Med* 1995;332:1113–18.
38. Odibo A, Nelson D, Stamilo DM, *et al*. Advanced maternal age is an independent risk factor for intrauterine growth restriction. *Am J Perinatol* 2006;23:325–8.
39. Campbell MK, Cartier S, Xie B, *et al*. Determinants of small for gestational age birth at term. *Paediatr Perinat Epidemiol* 2012;26:525–33.
40. Rogers JM. Tobacco and pregnancy: overview of exposures and effects. *Birth Defects Res* 2008;84:1–15.
41. Kabir Z, Daly S, Clarke V, *et al*. Smoking ban and small-for-gestational age births in Ireland. *PLoS ONE* 2013;8:e57441.
42. Mitchell EA, Thompson JM, Robinson E, *et al*. Smoking, nicotine and tar and risk of small for gestational age babies. *Acta Paediatr* 2002;91:323–8.
43. Juárez SP, Merlo J. Revisiting the effect of maternal smoking during pregnancy on offspring birthweight: a quasi-experimental sibling analysis in Sweden. *PLoS ONE* 2013;8:e61734.
44. Li X, Sundquist K, Sundquist J. Risks of small-for-gestational-age births in immigrants: a nationwide epidemiological study in Sweden. *Scand J Public Health* 2012;40:634–40.
45. Pepe MS. *The statistical evaluation of medical tests for classification and prediction*. Oxford, New York: Oxford University Press, 2003.
46. Zweig MH, Broste SK, Reinhart RA. ROC curve analysis: an example showing the relationships among serum lipid and apolipoprotein concentrations in identifying patients with coronary artery disease. *Clin Chem Aug* 1992;38(8 Pt 1):1425–8.
47. Pepe MS, Janes H, Gu JW. Letter by Pepe *et al* regarding article, "Use and misuse of the receiver operating characteristic curve in risk prediction". *Circulation* 2007;116:e132; author reply e134.

48. Royston P, Altman DG. Visualizing and assessing discrimination in the logistic regression model. *Stat Med* 2010;29:2508–20.
49. Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. *Biom J* 2008;50:457–79.
50. Ananth CV, Vintzileos AM. Distinguishing pathological from constitutional small for gestational age births in population-based studies. *Early Hum Dev* 2009;85:653–8.
51. Boyko EJ, Alderman BW. The use of risk factors in medical diagnosis: opportunities and cautions. *J Clin Epidemiol* 1990;43:851–8.
52. Wald NJ, Hackshaw AK, Frost CD. When can a risk factor be used as a worthwhile screening test? *BMJ* 1999;319:1562–5.
53. Khoury MJ, Newill CA, Chase GA. Epidemiologic evaluation of screening for risk factors: application to genetic screening. *Am J Public Health* 1985;75:1204–8.
54. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;115:928–35.
55. Zernicka-Goetz M, Huang S. Stochasticity versus determinism in development: a false dichotomy? *Nat Rev Genet* 2010;11:743–4.
56. Merlo J, Wagner P. The tyranny of the averages and the indiscriminate use of risk factors in public health: a call for revolution. *Eur J Epidemiol* 2013;28(Suppl 1):148.
57. Merlo J, Wagner P, Juarez S, *et al.* The tyranny of the averages and the indiscriminate use of risk factors and population attributable fractions in Public Health: the case of coronary heart disease. Working paper version 2013–09–26. *Unit for Social Epidemiology, Department of Clinical Sciences, Faculty of Medicine, Lund University*. 2013; http://www.med.lu.se/english/klinvetmalmo/unit_for_social_epidemiology/abstracts_c_working_papers_c.
58. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;115:928–35.
59. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, *et al.* Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157–72.
60. Pencina MJ, D'Agostino RB, Pencina KM, *et al.* Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol* 2012;176:473–81.
61. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 2011;30:11–21.
62. Hilden J. On NRI, IDI, and "good-looking" statistics with nothing underneath. *Epidemiology* 2014;25:265–7.
63. Pepe MS. Problems with risk reclassification methods for evaluating prediction models. *Am J Epidemiol* 2011;173:1327–35.
64. Hidden J, Gerds T. *Evaluating the impact of novel biomarkers: do not rely on IDI and NRI.* Denmark: Department of Biostatistics, University of Copenhagen, 2012; Research Report 12/08.
65. Kramer MS. Determinants of low birth weight: methodological assessment and meta-analysis. *Bull World Health Organ* 1984;65:663–737.
66. Frisell T, Öberg S, Kuja-Halkola R, *et al.* Sibling comparison designs: bias from non-shared confounders and measurement error. *Epidemiology* 2012;23:713–20.
67. Cnattingius S, Ericson A, Gunnarskog J, *et al.* A quality study of a medical birth registry. *Scand J Soc Med* 1990;18:143–8.
68. Grove A. Rethinking clinical trials. *Science* 2011;333:1679.
69. Kooperberg C, Ruczinski I, LeBlanc ML, *et al.* Sequence analysis using logic regression. *Genet Epidemiol* 2001;21:s626–31.
70. Janes H, Pepe M, Kooperberg C, *et al.* Identifying target populations for screening or not screening using logic regression. *Stat Med* 2005;15:1321–38.