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

Juárez, Sol Pía; Wagner, Phillip; Merlo, Juan

Published in:
BMJ Open

DOI:
10.1136/bmjopen-2014-005388

Published: 2014-01-01

Citation for published version (APA):
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 Merlo2

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.


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.

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 morbidities2 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 boys8 as well as Barrett’s oesophagus and oesophagitis9 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.
Two reviews, one from 1987 and the other from 2009, pointed out that SGA is associated with a broad number of genetic, obstetric, demographic and socioeconomic 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. For example, it is known that maternal hypertension during pregnancy gives a 5.5-fold increased risk of delivering an SGA baby. 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.

Although measures of discriminatory accuracy are extensively applied in other ﬁelds of epidemiology like the identiﬁcation of new biomarkers for cardiovascular diseases, these measures are still unusual in public health and epidemiology. 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. 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. We excluded 13 539 babies born with signiﬁcant congenital anomalies according to the MBR. Following previously established criteria, 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, using the last category as the reference. This variable was available at the MBR, where it is routinely calculated following standard intrauterine growth curves. 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.

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 and birth order, classified into three categories (ie, ﬁrstborn, second, and third or more). Among maternal characteristics we included birth interval between newborns, 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, categorised into yes, no, ‘only child’ and ﬁrst child; education, categorised into low (primary education or less), middle (secondary school) and high education (graduate and PhD); marital status, categorised into single, widowed, or divorced, and married or cohabiting; and maternal age at delivery, categorised into four groups (ie, <20, 20–24, 25–34 and >35 years old), as well as information on smoking habits, 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), and maternal origin, classiﬁed as being born in Sweden or not.

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)
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. A traditional measure of discriminatory accuracy is the area under the ROC curve (AU-ROC) or C statistic. 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.

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

Figure 1 Flow diagram showing the individuals excluded from the study population.
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.
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 units.

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

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

Values are ORs and (95% CI). Crude models.

SGA, small for gestational age.
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. 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*. 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
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. 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. 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, and encourage further research aiming to better capture IUGR.

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

---

**Figure 3** Area under the receiver operating characteristic curve for specific maternal and newborn characteristics.

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

The approach based on discriminatory accuracy understands average effects as an idealised mean value that does not necessarily represent the heterogeneity of individual effects. 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. 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. See elsewhere for a better explanation of these ideas. 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. 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, they

---

**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).
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 authors64 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.65 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 midwife 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.

Acknowledgements The authors thank Dr Karin Källén for her very helpful comments on a previous version of this paper.

Contributors JM had the original idea of applying measures of discriminatory accuracy for the interpretation of risk factors and discussed it with PW and SPJ. JM and SPJ initiated the study, JM, SPJ and PW contributed to the design of the study; SPJ performed the analyses under the supervision of JM and PW. SPJ wrote the first draft of the manuscript, and JM contributed to the writing of the final version. All authors made substantial contributions to the interpretation of the results and manuscript revision and approved the final version of the manuscript.

Funding  This was supported by the Swedish Research Council (VR) (Dnr #2013-2484, PI Juan Merlo), the Centre for Economic Demography and the SIMSAM early life Lund (Dnr #2013-5474).

Competing interests  None.

Ethics approval  The database was approved by the regional ethical review board in southern Sweden. Being a register-based study, the board did not require explicit informed consent from the women.

Provenance and peer review  Not commissioned; externally peer reviewed.

Data sharing statement  No additional data are available.

Open Access  This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

REFERENCES


64. Hilden 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.


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, Phillip Wagner and Juan Merlo

*BMJ Open* 2014 4:
doi: 10.1136/bmjopen-2014-005388

Updated information and services can be found at:
http://bmjopen.bmj.com/content/4/7/e005388

These include:

**References**
This article cites 65 articles, 14 of which you can access for free at:
http://bmjopen.bmj.com/content/4/7/e005388#BIBL

**Open Access**
This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

**Email alerting service**
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Notes**

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/