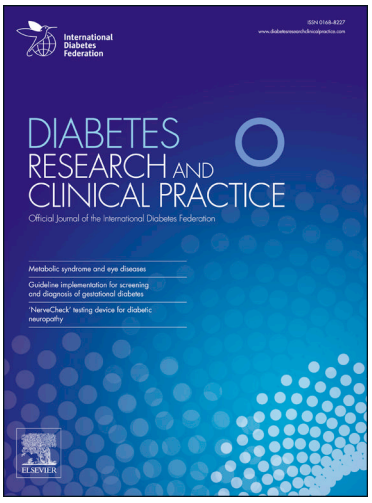
Accepted Manuscript



Review

Development of a risk engine relating maternal glycemia and body mass index to pregnancy outcomes

H. David McIntyre, Kristen S Gibbons, Julia Lowe, Jeremy JN Oats

PII: S0168-8227(18)30272-9
DOI: https://doi.org/10.1016/j.diabres.2018.02.036
Reference: DIAB 7249

To appear in: *Diabetes Research and Clinical Practice*

Received Date: 19 February 2018
Accepted Date: 27 February 2018

Please cite this article as: H. David McIntyre, K.S. Gibbons, J. Lowe, J. JN Oats, Development of a risk engine relating maternal glycemia and body mass index to pregnancy outcomes, *Diabetes Research and Clinical Practice* (2018), doi: https://doi.org/10.1016/j.diabres.2018.02.036

Title: Development of a risk engine relating maternal glycemia and body mass index to pregnancy outcomes

H David McIntyre[a,b], Kristen S Gibbons[b], Julia Lowe[c], Jeremy JN Oats[d]

[a] Mater Clinical Unit and [b] Mater Research, Faculty of Medicine, The University of Queensland, Raymond Terrace, South Brisbane, Queensland, 4101

[b]Mater Research, Level 3, Aubigny Place, Raymond Terrace, Brisbane, Queensland, Australia, 4101

[c]University of Toronto, 27 King's College Circle, Toronto, Ontario, Canada, M5S 1A1

[d]University of Melbourne, Parkville, Victoria, Australia, 3010

Corresponding author: H David McIntyre; david.mcintyre@mater.org.au

Word count: Abstract 168; Text 2696; Tables 3; Figures 3

Abstract

Aims: To develop a risk "engine" or calculator, integrating the risks of hyperglycemia, maternal BMI and other basic demographic data commonly available at the time of the pregnancy oral glucose tolerance test (OGTT), to predict an individual's absolute risk of specific adverse pregnancy outcomes.

Methods: Data from the Brisbane HAPO cohort was analysed using logistic regression to determine the relationship between four clinical outcomes (primary CS, birth injury, large-for-gestational age, excess neonatal adiposity) with different combinations of OGTT results and maternal demographics (age, height, BMI, parity). Existing sets of international GDM diagnostic criteria were also applied to the cohort.

Results: 191 (15.3%) women were diagnosed as GDM by one or more existing criteria. All international criteria performed poorly compared to risk models utilising OGTT results only, or OGTT results in combination with maternal demographics.

Conclusions: The risk engine's empirical performance on receiver – operator curve analysis was superior to the existing GDM diagnostic criteria tested. This concept shows promise for use in clinical practice, but further development is required.

Keywords: gestational diabetes, obesity, ROC analysis, risk

1.0 Introduction

In 2010, the International Association of Diabetes in Pregnancy Study Groups (IADPSG) published consensus based recommendations on the diagnosis and classification of hyperglycemia in pregnancy (HIP) [1], based primarily on data from the Hyperglycemia and Adverse Pregnancy Outcome Study (HAPO) [2]. A guiding principle of these recommendations was that classification should be based on the risk of adverse pregnancy outcomes, specifically in the case of IADPSG - risk of delivery of a large for gestational age (LGA) infant, risk of excess infant adiposity and risk of neonatal hyperinsulinemia. Other reports from the HAPO study have outlined the association of increasing maternal body mass index (BMI) with pregnancy complications [3] and described the combined associations of hyperglycemia and BMI with these complications [4]. In particular, the latter study by Catalano et al clearly demonstrated that the risks of obesity and IADPSG-defined gestational diabetes (GDM), considered as categorical variables, were additive and that the presence of both identified a group of women at particularly high risk of a broad range of adverse pregnancy outcomes.

However, these reports still considered both GDM and maternal obesity as dichotomous (YES / NO) variables, despite the fact that the underlying associations between both glucose [2] and maternal BMI were clearly continuous [3]. Thus, for example, a woman with one glucose value on an oral glucose tolerance test (OGTT) marginally above the diagnostic threshold, but with a BMI well in the population normal range would be considered as having GDM and thus requiring treatment. By contrast, a woman with obesity and all OGTT glucose values just below diagnostic thresholds would be classified as "normal" from the glycemic viewpoint. The latter woman might receive some dietary or lifestyle advice during pregnancy, but, depending on local clinical policy, might also continue in routine antenatal care.

The aim of the current work was to develop a risk "engine" or "calculator" connecting HIP to the frequency of adverse pregnancy outcomes, similar to the approach used for predicting cardiovascular risk in diabetes outside pregnancy [5]. We particularly wished to integrate consideration of the risks of and maternal glycemia and BMI, along with other basic demographic data commonly available at the time of OGTT, to determine an individual woman's absolute risk of specific adverse pregnancy outcomes.

2.0 Subjects, Materials and Methods

This study is a *post hoc* analysis of data collected at Mater Health Services, South Brisbane under the auspices of the HAPO study. Approval from Mater Health Services Human Research Ethics Committee was current throughout the course of the study.

Exclusion criteria for this study were: unblinding from the HAPO study; multiple pregnancy (not included in HAPO), stillbirth, congenital anomaly, non-Caucasian mother, birth at less than 33 completed weeks of gestation and missing data for key independent variables.

Four key clinical outcomes were extracted from the Mater HAPO dataset for use in this study; 1) primary caesarean section as noted in the HAPO data file, 2) birth injury (including shoulder dystocia) defined as in the primary HAPO cohort [2], 3) large-for-gestational age (LGA) (corrected for gender and gestational age at delivery) defined as birthweight greater than the 90th centile in comparison to the Australian national standard values published by Roberts and Lancaster [6], 4) neonatal adiposity defined as % body fat greater than the 90th centile derived from within the Mater HAPO dataset, 5) neonatal hyperinsulinemia defined as elevated cord c-peptide (>1.7µg/L, 90th centile in the HAPO dataset) and 6) neonatal hypoglycaemia defined as a blood glucose level of <2.2mmol/L (10th centile in the HAPO dataset.

Since a "risk engine" would potentially offer an alternative to dichotomous GDM diagnosis, it was considered important to define the predictive ability of the currently used diagnostic criteria for GDM to identify women at risk of these outcomes. Thus, four sets of recognised GDM diagnostic criteria were firstly applied to the data:

(1) the 1991Australasian Diabetes in Pregnancy Society (ADIPS) criteria [7] (fasting ≥5.5mmol/L and/or 2-hour ≥8.0mmol/L),

and current international criteria:

(2) IADPSG criteria [1] (now also adopted by ADIPS and the World Health Organization [8]; any of fasting ≥5.1mmol/L, 1-hour ≥10.0mmol/L, 2-hour ≥8.5mmol/L),
(3) National Institute for Clinical Excellence (NICE criteria) [9] (fasting ≥5.6mmol/L and/or 2-hour ≥7.8mmol/L), and
(4) American College of Obstetricians and Gynecologists (ACOG) / Carpenter and Coustan criteria [10, 11] (two or more of fasting ≥5.3mmol/L, 1-hour ≥10.0mmol/L, 2-hour ≥8.6mmol/L).

It must be noted that although we applied differing sets of GDM diagnostic thresholds, we were unable to mimic the full diagnostic protocols of NICE, which involves risk factor screening and selective biochemical testing, or ACOG, which involves two step testing with an initial non-fasting glucose "challenge" test and subsequent 100-gram OGTT.

The sensitivity, specificity, receiver operator characteristic (ROC) area under the curve (AUC) (calculated as [sensitivity + specificity] / 2) and Youden Index [12](calculated as [sensitivity +

specificity - 1]), along with associated 95% confidence intervals (CIs), were calculated for each combination of outcome and criteria. The optimum set of criteria was determined using comparison of the ROC AUCs and 95% CIs.

We then sought to develop and assess multivariable regression models incorporating potential predictors of pregnancy outcomes available within the HAPO dataset and likely to be available in routine clinical practice. The choice of independent variables was limited to those available for all pregnant women and thus previous birth outcomes were excluded. Exploratory analyses suggested that maternal age, height and body mass index (BMI) were relevant predictors and these variables were included in the models. Their values were mean centred based on the data from the HAPO Brisbane sample and the association of glucose values with outcomes was assessed using the individual values for fasting/one hour/two hour glucose on the 75g HAPO OGTT as well as using a standardised average measure of glycemia during the OGTT, calculated using the following formula: average OGTT = ([standardised fasting glucose] + [standardised one hour glucose] + [standardised two hour glucose]) / 3. Parity was collapsed to nulliparous (referent) vs parous.

The following models were developed: A) fasting OGTT result only; B) fasting and one hour OGTT results; C) fasting and two hour OGTT results; D) fasting, one hour and two hour OGTT results; E) average OGTT results; F) Hemoglobin A1c (HbA1c); G) fasting, one hour and two hour OGTT results and maternal demographics (age, height, BMI, parity); H) average OGTT components and maternal demographics (age, height, BMI, parity).

Logistic regression analyses were undertaken for the four clinical outcomes using the independent variables outlined above (Models A – H). Given the exploratory nature of this study, potential predictor variables were included in the regression analyses for all outcome variables, even if they were not individually significant on bivariate analysis for some outcomes (see Table 2). Statistical assumptions for models were tested and met. For each model the predicted risk of each outcome was calculated and compared using the Mann-Whitney U test between actual cases and non-cases. To enable comparison to existing criteria, the Youden Index "optimal cut-off" and the "diagnostic cut-off" (where the specificity is approximately three times the sensitivity)[13] for each model was determined. Sensitivity, specificity and ROC AUC was calculated for existing criteria and for Models A - H. Each model was also applied to the whole dataset to determine how many participants would be identified as 'at-risk' of each outcome using both the Youden Index "optimal" cut-off and the diagnostic cut-off thresholds. These dichotomous predictions derived from Models A – H were then compared to existing dichotomous criteria. Analyses were conducted in Stata/SE 13.1 (StataCorp LLC, College Station, Texas, USA) and deemed statistically significant at the 0.05 level.

3.0 Results

A total of 1248 women from the Brisbane HAPO cohort met the inclusion criteria. Baseline characteristics are noted in Table 1.

Comparing the existing sets of international criteria, 113 (9.1%) participants were classified as GDM by the previous ADIPS criteria, 131 (10.5%) using IADPSG, 134 (10.7%) using NICE and 33 (2.6%) using the ACOG / Carpenter and Coustan criteria. In total, 191 (15.3%) participants were diagnosed by one or more sets of GDM diagnostic criteria, with only 30 participants diagnosed as GDM by all four criteria (Figure 1).

All sets of existing criteria yielded low sensitivities, specificities, ROC AUC and Youden Indices when related to the clinical adverse outcomes. Sensitivities for birth injury ranged from 7.7% (95% CI 0.2%-36.0%, ACOG) to 30.8% (95% CI 9.1%-61.4%, IADPSG), specificity 89.3% (95% CI 87.5%-91.0%, NICE) to 97.4% (95% CI 96.4%-98.2%, ACOG), ROC AUC 0.523 (95% CI 0.421-0.626, NICE) to 0.602 (95% CI 0.472-0.733, IADPSG) and Youden Index 4.7% (NICE) to 20.5% (IADPSG). Results were similar for primary CS, LGA and adiposity (data not shown). There was no statistically significant difference between the predictive ability of the four pre-existing sets of GDM diagnostic criteria, however visual inspection indicated that the IADPSG criteria were consistently slightly better (Figure 2a), and therefore IADPSG criteria were used in subsequent comparisons to the new models developed.

Similar analyses were carried out for the newly developed regression models A to F, using the Youden Index to determine a potential cut-off value to categorise participants as "at-risk" of each clinical outcome. If applied in clinical practice, this designation would be similar to a GDM "diagnosis" using a conventional dichotomous approach.

When comparing the five models, there was no statistically significant difference between the models. However, in general, the point estimates for model C (fasting and 2 hour OGTT results) were higher than for the other models. There was a statistically significant benefit over IADPSG for prediction of primary CS (Figure 2b). Using model C, there were 39.0%, 36.6%, 36.2%, 53.5%, 31.7% and 60.1% of participants identified as at-risk of primary CS, birth injury, LGA, adiposity, hyperinsulinemia and hypoglycaemia, respectively, compared with actual rates of 19.3%, 1.6%, 14.0%, 8.2%, 7.7% and 10.5%.

In order to develop a model that provided less over-classification of "at-risk" patients, models A to F (containing glycemic variables only) were re-evaluated using the diagnostic cut-off (Figure 2c). There were no statistically significant differences between the five models. Using model C (for consistency with the model chosen based on the Youden Index), there were 18.7%, 14.3%, 16.7%, 16.8%, 17.7% and 23.1% of participants identified as at-risk of primary CS, birth injury, LGA, adiposity, hyperinsulinemia and hypoglycaemia respectively, compared with actual rates of 19.3%, 0.8%, 14.0%, 8.2%, 7.7% and 10.5%.

Model G (individual GTT components) and model H (averaged GTT results) included maternal characteristics, in addition to the OGTT results, as independent variables. While both models demonstrated similar strength of relationships between the independent variables and the clinical outcomes, model G had slightly better predictive ability (data not shown), and as such results from model H will not be presented.

Differing combinations of maternal age, height, BMI, parity, fasting glucose and two hour OGTT glucose were significantly related to the clinical outcomes of interest (Table 2). Parity, BMI and fasting glucose were most frequently associated with a clinical outcome. When using the logistic regression models to calculate the predicted risk for each case and outcome, there was a significant difference in the median predicted risk between actual cases and non-cases.

The potential utility of these models arguably lies in their potential ability to calculate a predicted absolute risk for outcome(s) on a continuous scale, rather than to dichotomise outcome risks into "at risk" and "not at risk" categories. However, to allow comparison with the dichotomous results of current diagnostic criteria, the sensitivity, specificity and ROC AUC were calculated for model G using both the Youden Index and the diagnostic cut-off (Table 3). Figure 5 presents the ROC curves for the dataset for the four key clinical outcomes, along with the point estimates of sensitivity and specificity for each of the four existing international criteria, model C (Youden Index and diagnostic cut-off) and model G (Youden Index and diagnostic cut-off).

4.0 Discussion

Despite spirited defence of their supposed virtues by proponents aligned with various national and international professional and scientific organizations, all the existing criteria for dichotomous classification or "diagnosis" of HIP performed extremely poorly in terms of objective ROC analysis in the current study, with ROC AUCs generally < 55%. Thus, they are barely superior to "tossing a coin" in terms of specifically identifying individual women at risk of adverse pregnancy outcomes. Such a result is perhaps not unexpected given the known continuous relationship between glucose measures and pregnancy outcomes [2]. When the commonly used Youden index is applied and "optimal" cut-offs used to delineate women "at risk" of particular adverse outcomes, the ROC AUCs improve to approximately 60%, but at the cost of over classification, with 30 – 60% of the cohort identified as "at risk". Higher potential rates of GDM diagnosis have already been identified as a barrier to implementation of the IADPSG / WHO 2013 criteria into clinical practice [14, 15] and a further increase in the frequency of women identified as "at risk" and potentially requiring intervention appears untenable.

Given this pragmatic obstacle, we also considered the potential use of "diagnostic" thresholds, also derived from ROC analysis, as a way of better identifying women "at risk". Due to the continuous risk gradient and the inevitable "trade – off" between sensitivity and sensitivity involved, this

approach gave lower frequencies (between 14 – 23%) of women identified as "at risk", but nonetheless performed better than existing dichotomous diagnostic criteria (see Figure 5a).

Overall, the ROC curve analyses demonstrate both the poor performance of existing strategies for classification of HIP and the difficulty of developing dichotomous approaches to separating "normal" from "abnormal" or "at risk" from "not at risk" across a spectrum of risk. The risk gradient is acknowledged as continuous, but tradition and "clinical thinking" seek to dichotomise it at all costs, even in defiance of the observed data.

We consider that a more rational future approach, in line with the findings described in this study, would be to define a point estimate of the absolute risk (and appropriate confidence intervals) of specific adverse outcomes for each woman on the basis of her individual demographic, clinical and laboratory findings. The variables used in the current study provide the initial basis for a logical approach to this issue. The data used in our regression models have significant strengths, being derived from carefully standardized clinical observational data with strict quality control and data supervision. Furthermore, glucose measurements in HAPO were blinded and women received no glucose lowering treatment, eliminating treatment confounding on the basis of perceived risk. However, some important limitations must also be noted. The data used in development of these models were confined to a single HAPO center and to women of Caucasian ethnicity (89% of HAPO participants at the Mater site) and require validation across geographically and ethnically diverse cohorts. Women with marked hyperglycemia were unblinded and treated under the HAPO study protocol [2] and thus the models do not encompass the full range of glucose values which might be encountered in routine clinical practice, especially in areas with a high prevalence of diabetes. Further, previous randomized controlled trials (RCTs) have been conducted using inclusion criteria based on existing dichotomous diagnostic frameworks [16, 17] and have demonstrated meaningful clinical benefits in terms of immediate pregnancy outcomes. Thus, an attempt at risk classification using an alternative approach such as a "risk engine" would require validation before being used routinely in clinical care.

Risk models have been applied to other aspects of GDM. Barnes et al [18] have reported a risk model designed to predict the likelihood of a woman with GDM requiring insulin therapy, but this model employed a conventional dichotomous GDM diagnosis strategy [7]. Kalter – Lebovici have reported a risk model designed to determine which women should proceed from initial fasting glucose testing to a full OGTT, taking into account other clinical features [19]. However, this model again viewed GDM in a dichotomous fashion and the primary outcome considered was GDM diagnosis rather than specific pregnancy outcomes.

Despite these caveats we believe that the "risk engine" approach as described in this paper offers a viable alternative to current dichotomous approaches to diagnosis and management of maternal hyperglycaemia in pregnancy and merits further development and evaluation.

5.0 Acknowledgements

We would like to thank all HAPO participants at the Mater site for their voluntary participation in this study.

6.0 References

[1] Metzger BE, Gabbe SG, Persson B, Buchanan TA, Catalano PA, Damm P, et al. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care. 2010;33:676-82.
[2] Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, Coustan DR, et al. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med. 2008;358:1991-2002.
[3] HAPO Study Cooperative Research Group. Hyperglycaemia and Adverse Pregnancy Outcome (HAPO) Study: associations with maternal body mass index. BJOG. 2010;117:575-84.
[4] Catalano PM, McIntyre HD, Cruickshank JK, McCance DR, Dyer AR, Metzger BE, et al. The Hyperglycemia and Adverse Pregnancy Outcome Study: Associations of GDM and obesity with pregnancy outcomes. Diabetes Care. 2012;35:780-6.
[5] Almeda-Valdes P, Cuevas-Ramos D, Mehta R, Gomez-Perez FJ, Aguilar-Salinas CA. UKPDS Risk Engine, decode and diabetes PHD models for the estimation of cardiovascular risk in patients with diabetes. Curr Diabetes Rev. 2010;6:1-8.
[6] Roberts CL, Lancaster PA. Australian national birthweight percentiles by gestational age. Med J Aust. 1999;170:114-8.
[7] Martin FI. The diagnosis of gestational diabetes. Ad Hoc Working Party. Med J Aust. 1991;155:112.
[8] Colagiuri S, Falavigna M, Agarwal MM, Boulvain M, Coetzee E, Hod M, et al. Strategies for implementing the WHO diagnostic criteria and classification of hyperglycaemia first detected in pregnancy. Diabetes Res Clin Pract. 2014;103:364-72.
[9] NICE guidelines. Diabetes in pregnancy: management from preconception to the postnatal period. 2015.
[10] ACOG Committee on Practice Bulletins. Practice Bulletin No. 137: Gestational diabetes mellitus. Obstet Gynecol. 2013;122:406-16.
[11] Carpenter MW, Coustan DR. Criteria for screening tests for gestational diabetes. Am J Obstet Gynecol. 1982;144:768-73.
[12] Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3:32-5.
[13] Tripepi G, Jager KJ, Dekker FW, Zoccali C. Diagnostic methods 2: receiver operating characteristic (ROC) curves. Kidney Int. 2009;76:252-6.
[14] McIntyre HD, Oats JJ, Zeck W, Seshiah V, Hod M. Matching diagnosis and management of diabetes in pregnancy to local priorities and resources: an international approach. Int J Gynaecol Obstet. 2011;115 Suppl 1:S26-9.
[15] NIH consensus panel. Final Statement on "Diagnosing Gestational Diabetes". Bethesda MD: National Institutes of Health, USA; 2013.
[16] Crowther CA, Hiller JE, Moss JR, McPhee AJ, Jeffries WS, Robinson JS. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. N Engl J Med. 2005;352:2477-86.
[17] Landon MB, Spong CY, Thom E, Carpenter MW, Ramin SM, Casey B, et al. A multicenter, randomized trial of treatment for mild gestational diabetes. N Engl J Med. 2009;361:1339-48.
[18] Barnes RA, Wong T, Ross GP, Jalaludin BB, Wong VW, Smart CE, et al. A novel validated model for the prediction of insulin therapy initiation and adverse perinatal outcomes in women with gestational diabetes mellitus. Diabetologia. 2016;59:2331-8.
[19] Kalter-Leibovici O, Freedman LS, Olmer L, Liebermann N, Heymann A, Tal O, et al. Screening and Diagnosis of Gestational Diabetes Mellitus: Critical appraisal of the new International Association of Diabetes in Pregnancy Study Group recommendations on a national level. Diabetes Care. 2012;11:11.

Table 1: Descriptive statistics for the cohort (N=1248)

Variable	**Mean**	**Standard Deviation**
Gestation at delivery (weeks)	39.8	1.5
Birthweight (g)	3522	502
Maternal age (years)	29.4	5.3
Maternal height (cm)	166	6.1
Maternal BMI	29.0	5.7
Basal glucose	4.4	0.4
One hour glucose	7.4	1.5
Two hour glucose	6.2	1.2
	N	%
Parity – 0	680	54.5
Parity – 1	348	27.9
Parity – 2	140	11.2
Parity – 3	80	6.4
Parity – parous	568	45.5
Primary Caesarean section	241	19.3
Birth injury including shoulder dystocia	29	1.6
Large-for-gestational age	175	14.0
Adiposity *(N=1,000)*	100	10.0
Hyperinsulinemia *(N=992)*	76	7.7
Hypoglycaemia *(N=695)*	73	10.5

Table 2: Results of logistic regression analysis including maternal age, maternal height, parity, maternal BMI and individual GTT components for model G

Outcome	Standardised coefficients							Median (IQR) Predicted Risk		
	Maternal Age	Maternal Height	Parous	BMI	Fasting GTT	One-hour GTT	Two-hour GTT	Non-cases	Cases	p-value[#]
Primary CS	0.267*	-0.165*	-0.632*	0.243*	0.074	-0.023	0.088	13.6% (16.6%)	27.5% (21.2%)	<0.001
Birth Injury	-0.101	-0.239	0.042	-0.408	0.541*	0.358	-0.074	0.5% (0.8%)	1.8% (1.9%)	<0.001
LGA	-0.123	0.340*	0.191*	0.247*	0.168*	0.075	0.177*	11.0% (8.8%)	16.9% (13.5%)	<0.001
Adiposity	-0.278*	0.143	0.131	0.206*	0.219*	0.148	0.044	8.3% (6.0%)	10.4% (9.0%)	<0.001
Hyperinsulinemia	0.140	-0.061	0.009	0.063	0.156	0.255	0.045	6.6% (4.4%)	8.4% (5.7%)	<0.001
Hypoglycaemia	-0.045	-0.055	-0.110	-0.119	0.186	0.091	0.036	10.1% (3.2%)	10.9% (4.3%)	0.032

* indicates a statistically significant relationship at the 0.05 level with the outcome

Mann-Whitney U test

CS Caesarean section; LGA large-for-gestational age; IQR interquartile range; BMI body mass index; GTT glucose tolerance test

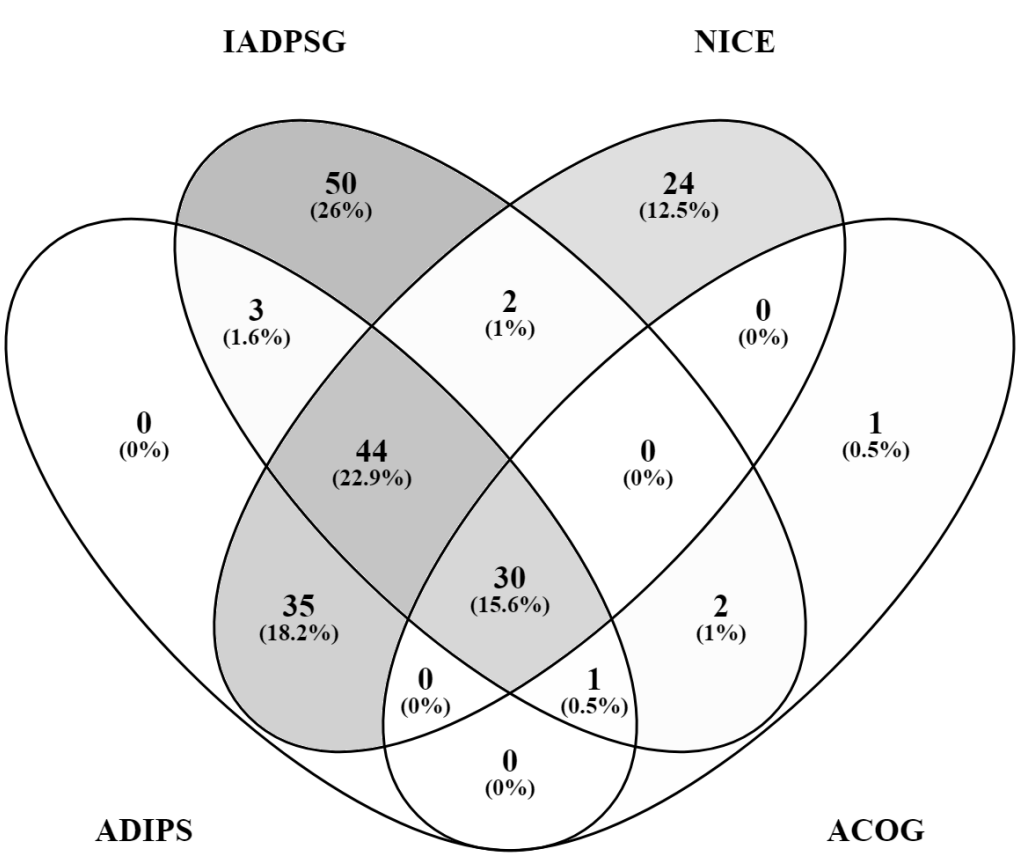
Table 3: Statistics and 95% confidence intervals for the primary five models evaluated

	Primary CS (n=241, 19.3%)				Birth Injury (n=29, 1.6%)				LGA (n=175, 14.0%)			
Model	Sens	Spec	ROC AUC	n (%) classified	Sens	Spec	ROC AUC	n (%) classified	Sens	Spec	ROC AUC	n (%) classified
IADPSG	13.3% (9.3%-18.2%)	90.2% (88.2%-91.9%)	0.517 (0.494-0.541)	-	30.8% (9.1%-61.4%)	89.7% (87.9%-91.4%)	0.602 (0.472-0.733)	-	16.0% (10.9%-22.3%)	90.4% (88.5%-92.1%)	0.532 (0.503-0.561)	-
Model C (Youden)	52.3% (45.8%-58.7%)	64.2% (61.1%-67.1%)	0.582 (0.547-0.617)	487 (39.0%)	84.6% (54.6%-98.1%)	63.9% (61.1%-66.6%)	0.743 (0.640-0.845)	457 (36.6%)	56.0% (48.3%-63.5%)	67.0% (64.1%-69.8%)	0.615 (0.576-0.655)	452 (36.2%)
Model C (Diagnostic)	27.8% (22.2%-33.9%)	83.5% (81.1%-85.8%)	0.557 (0.526-0.587)	233 (18.7%)	23.1% (5.0%-53.8%)	85.7% (83.7%-87.7%)	0.544 (0.425-0.664)	179 (14.3%)	28.6% (22.0%-35.9%)	85.3% (83.0%-87.3%)	0.569 (0.534-0.604)	208 (16.7%)
Model G (Youden)	67.6% (61.3%-73.5%)	71.2% (68.3%-74.0%)	0.694 (0.661-0.727)	453 (36.3%)	69.2% (38.6%-90.9%)	70.5% (67.9%-73.1%)	0.699 (0.568-0.830)	373 (29.9%)	85.1% (79.0%-90.1%)	45.7% (42.7%-48.7%)	0.654 (0.624-0.684)	732 (58.7%)
Model G (Diagnostic)	30.3% (24.6%-36.5%)	90.8% (88.8%-92.5%)	0.605 (0.575-0.636)	166 (13.3%)	30.8% (9.1%-61.4%)	92.7% (91.1%-94.1%)	0.617 (0.487-0.748)	94 (7.5%)	29.1% (22.5%-36.5%)	88.5% (86.5%-90.4%)	0.588 (0.553-0.623)	174 (13.9%)

Sens sensitivity; Spec specificity

	Adiposity (n=100/1000, 10.0%)				Hyperinsulinemia (n=76/992, 7.7%)				Hypoglycaemia (n=73/695, 10.5%)			
Model	Sens	Spec	ROC AUC	n (%) classified	Sens	Spec	ROC AUC	n (%) classified	Sens	Spec	ROC AUC	n (%) classified
IADPSG	17.0% (10.2%-25.8%)	90.1% (88.0%-92.0%)	0.536 (0.497-0.574)	-	15.8%	90.5%	0.531	-	9.6%	89.1%	0.493	-
Model C (Youden)	70.0% (60.0%-78.8%)	46.3% (43.0%-49.7%)	0.582 (0.534-0.630)	668 (53.5%)	51.3% (39.6%-63.0%)	70.5% (67.5%-73.5%)	0.609 (0.551-0.668)	396 (31.7%)	69.9% (58.0%-80.1%)	42.8% (38.8%-46.8%)	0.563 (0.507-0.620)	750 (60.1%)
Model C (Diagnostic)	28.0% (19.5%-37.9%)	84.0% (81.4%-86.3%)	0.560 (0.514-0.606)	209 (16.8%)	28.9% (19.1%-40.5%)	83.7% (81.2%-86.1%)	0.563 (0.511-0.616)	221 (17.7%)	27.4% (17.6%-39.1%)	78.3% (74.8%-81.5%)	0.528 (0.474-0.582)	288 (23.1%)
Model G (Youden)	86.0% (77.6%-92.1%)	35.6% (32.4%-38.8%)	0.608 (0.570-0.645)	824 (66.0%)	36.8% (26.1%-48.7%)	83.4% (80.8%-85.8%)	0.601 (0.545-0.657)	227 (18.2%)	34.2% (23.5%-46.3%)	80.5% (77.2%-83.6%)	0.574 (0.517-0.632)	271 (21.7%)
Model G (Diagnostic)	29.0% (20.4%03 8.9%)	87.7% (85.3%-89.7%)	0.583 (0.537-0.629)	172 (13.8%)	28.9% (19.1%-40.5%)	86.4% (84.0%-88.5%)	0.577 (0.524-0.629)	187 (15.0%)	27.4% (17.6%-39.1%)	82.6% (79.4%-85.5%)	0.550 (0.497-0.604)	242 (19.4%)

Figure 1



Legend Figure 1

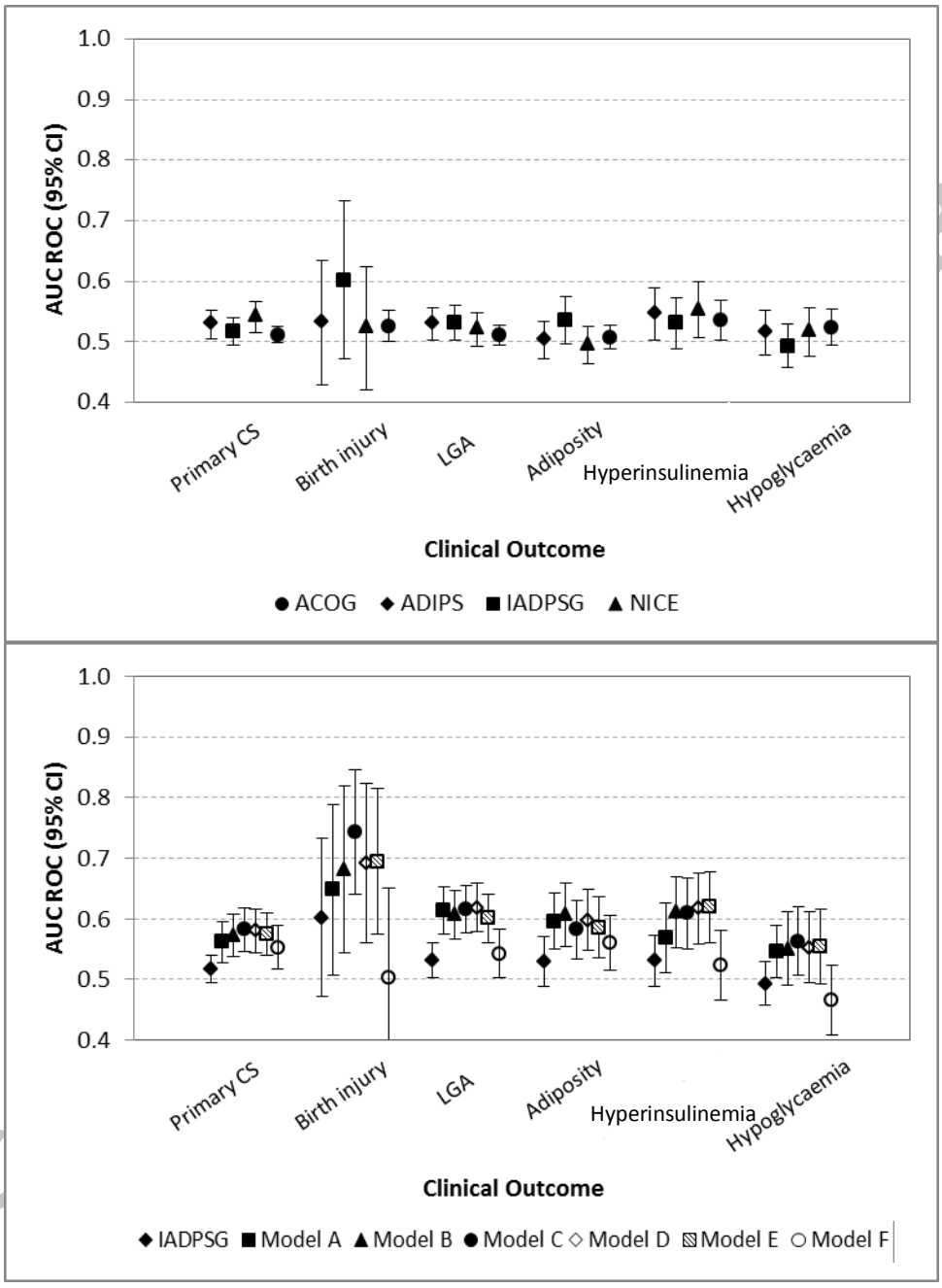
Venn diagram depicting the participants who received gestational diabetes diagnoses by the four sets of international criteria

Abbreviations: ADIPS Australasian Diabetes in Pregnancy Society (ADIPS), IADPSG International Association of Diabetes in Pregnancy Study Groups, NICE National Institute for Clinical Excellence, ACOG American College of Obstetricians and Gynecologists

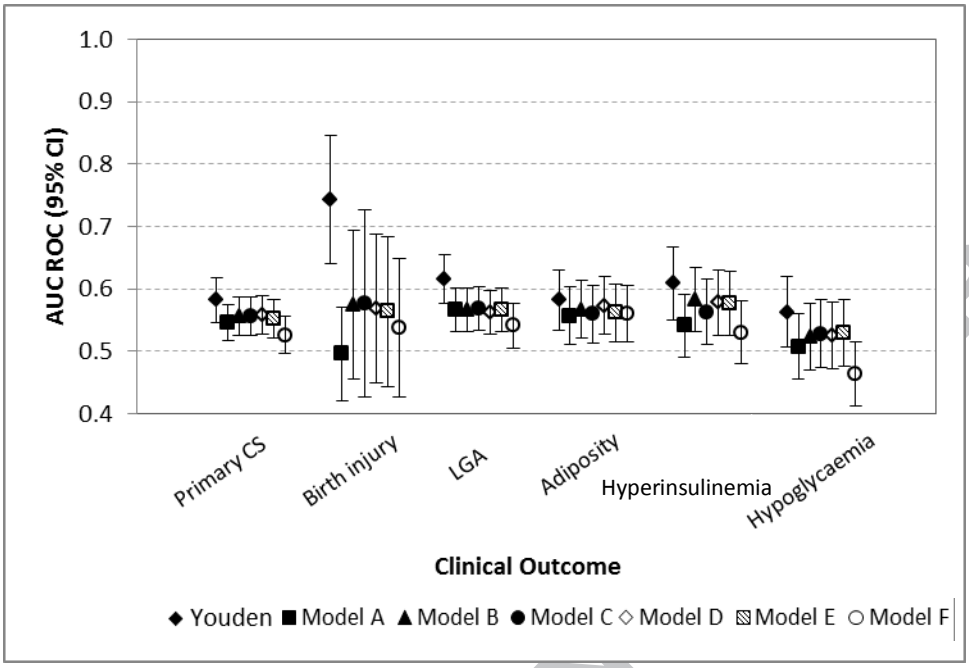
Figure 2:

a)

b)



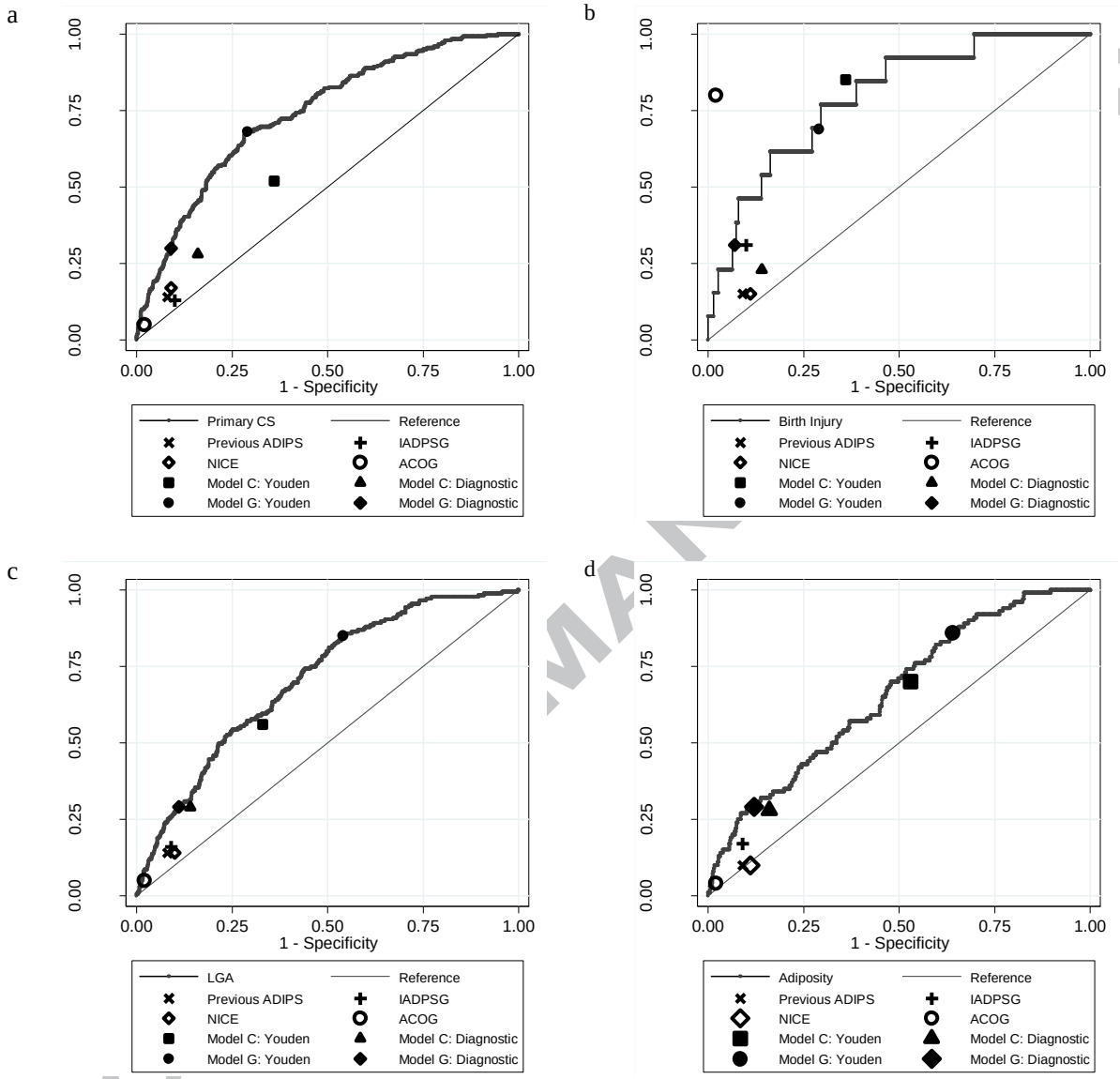
c)



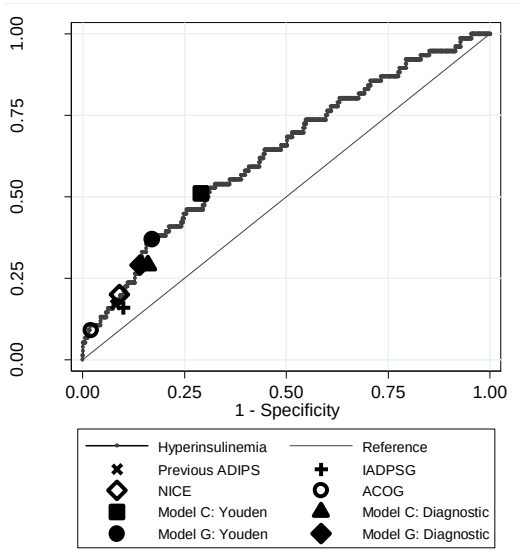
Legend Figure 2:

Comparison of AUC ROC (95% CI) for six key clinical outcomes. Models compared a) existing international criteria, b) IADPSG and the five new models using Youden index, c) optimal model using Youden index cut-off and the five new models using the diagnostic cut-off

Figure 3:



e



f

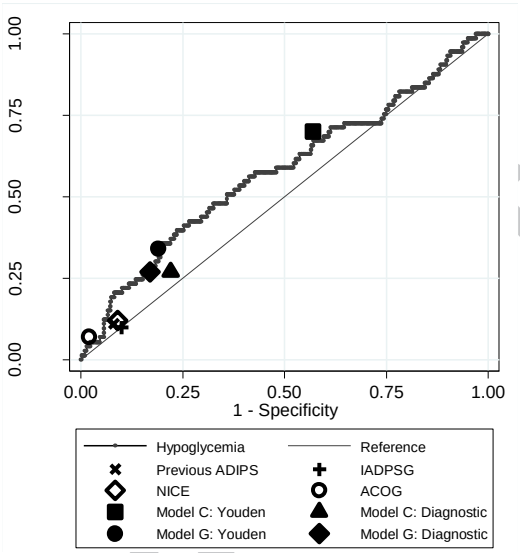


Figure 3 Legend:

Model G ROC curve with point estimates of sensitivity and specificity for each of the four existing international GDM diagnosis criteria, as well as models C and G, for a) primary Caesarean section, b) birth injury, c) large-for-gestational age, d) neonatal adiposity, e) hyperinsulinemia, f) hypoglycaemia.