BMJ Open

A COST EFFECTIVENESS COMPARISON OF THE NICE 2015 AND WHO 2013 DIAGNOSTIC CRITERIA FOR GESTATIONAL DIABETES

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-016621
Article Type:	Research
Date Submitted by the Author:	03-Mar-2017
Complete List of Authors:	Jacklin, Paul; Royal College of Obstetricians and Gynaecologists, National Guideline Alliance Maresh, Michael; Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre Patterson, Chris; Queen's University Belfast, Centre for Public Health Stanley, Katharine; Norfolk and Norwich University Hospitals NHS Foundation Trust, Department of Obstetrics and Gynaecology Dornhorst, Anne; Hammersmith Hospital, Department of Investigative Medicine Burman-ROY, Shona; Royal College of Obstetricians and Gynaecologists, National Guideline Alliance Bilous, Rudy; Newcastle University Medicine Malaysia
Primary Subject Heading:	Health economics
Secondary Subject Heading:	Diabetes and endocrinology, Obstetrics and gynaecology, Diagnostics
Keywords:	HEALTH ECONOMICS, DIABETES & ENDOCRINOLOGY, OBSTETRICS

Title:

A COST EFFECTIVENESS COMPARISON OF THE NICE 2015 AND WHO 2013 DIAGNOSTIC CRITERIA FOR GESTATIONAL DIABETES

Authors: PB Jacklin[1], MJA Maresh[2], CC Patterson[3], KP Stanley[4], A Dornhorst[5], S Burman-Roy[1], RW Bilous[6]

Institutions:

1. Royal College of Obstetricians and Gynaecologists, London NW1 4RG, UK
2. St. Mary's Hospital, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester M13 9WL, UK
3. Centre for Public Health, Queen's University Belfast, Room 3.014, ICS Block B, Grosvenor Road, Belfast BT12 6BJ, UK
4. Department of Obstetrics and Gynaecology, Norfolk and Norwich University Hospitals NHS Foundation Trust, Colney Ln, Norwich NR4 7UY, UK
5. Faculty of Medicine, Department of Investigative Medicine, Hammersmith Hospital, Imperial College London, London, UK
6. Newcastle University Medicine Malaysia, Johor, Malaysia

Corresponding author:

Paul Jacklin
e-mail: pjacklin@rcog.org.uk

Abstract: 278 words

Main text: 4618 words

Abstract

Objectives To compare the cost effectiveness of The National Institute for Health and Care Excellence (NICE) 2015 and the World Health Organisation (WHO) 2013 diagnostic thresholds for gestational diabetes (GDM).

Setting: The analysis was from the perspective of the National Health Service (NHS) in England and Wales.

Participants: 6,221patients from 4 of the Hyperglycaemia and Adverse Pregnancy Outcomes (HAPO) study centres (2 UK, 2 Australian), 6,308 patients from the Atlantic Diabetes in Pregnancy (DiP) study and 12,755 patients from UK clinical practice

Primary and secondary outcome measures planned: The incremental cost per quality adjusted life year (QALY), net monetary benefit (NMB) and the probability of being cost-effective at cost-effectiveness thresholds of £20,000 and £30,000 per QALY

Results. In a population of pregnant women from the 4 HAPO study centres, and utilising NICE defined risk factors for GDM, diagnosing GDM using NICE 2015 criteria had an incremental cost effectiveness ratio (ICER) of £23,073 per QALY gained compared to £37,669 per QALY gained using WHO 2013 diagnostic criteria. At a cost-effectiveness threshold of £30,000 per QALY the NICE 2015 criteria had a 43.4% probability of being cost-effective compared to the WHO 2013 diagnostic criteria which had a 34.7% probability of being cost-effective (no treatment had a 21.9% probability of being cost-effective). The ICERs for women without NICE risk factors in this population were £43,845 and £220,638 per QALY for NICE and WHO diagnostic criteria, respectively.

Conclusion The NICE 2015 diagnostic criteria for GDM can be considered cost-effective relative to the WHO 2013 alternative at a cost-effectiveness (CE) threshold of £30,000 per QALY. Universal screening for GDM was not found to be cost-effective relative to screening based on NICE risk factors.

Keywords: Cost Effectiveness, Gestational Diabetes, Screening, Risk Factors, Diagnosis

Strengths and limitations of this study

- This economic evaluation address an important clinical and policy issue. The existing economic evidence is limited and WHO have stated that studies of this type are needed to inform a future update of their guideline
- Our paper has used patient-level data from the influential HAPO study for an economic analysis which has not been previously been published in a peer reviewed journal.
- This analysis provides strong evidence that universal screening is not cost-effective in the UK
- This analysis suggests that the NICE diagnostic criteria for GDM are more cost-effective than the WHO criteria in the UK context
- Model conclusions are sensitive to uncertainties with respect to valuation of health outcomes and the possible long term metabolic consequences for offspring for which the evidence is debated and which are hard to quantify

Introduction

The diagnostic glycaemic thresholds for GDM remain the subject of considerable debate. The original definition was based upon maternal risk for developing post partum diabetes, but subsequent thresholds have concentrated on complications during pregnancy and the health of the offspring. Since the publication of the HAPO study[1] showing that there was a linear association between increasing levels of maternal hyperglycaemia and adverse perinatal outcomes with no obvious threshold, the discussion around the diagnostic criteria that should define GDM has intensified. New diagnostic thresholds were proposed by the International Association of Diabetes in Pregnancy Study Group (IADPSG)[2] based upon the HAPO study levels of plasma glucose when fasting, and at 1 and 2 hours after an oral 75g glucose load that were associated with covariate adjusted odds ratio of 1.75 relative to the mean glucose value in the whole HAPO cohort on three offspring outcomes: exceeding the 90th centile for birth weight, for cord serum C-peptide concentration and for percent fetal body fat. These diagnostic criteria have been subsequently adopted by the WHO.[3] However, they remain controversial, and have not been supported by bodies such as the National Institutes for Health and the American College of Obstetricians.[4] Furthermore, WHO has acknowledged that they will have to be revisited in the near future in the light of new studies reporting their cost-effectiveness.[3]

In 2015 the NICE published updated guidance on Diabetes in Pregnancy[5] which included recommendations on diagnostic thresholds for GDM which differ from those adopted by WHO. These NICE thresholds were informed by an economic evaluation of the type that WHO considered important to inform future recommendations, but have attracted criticism in the UK[6] and elsewhere. Data from a recently published Spanish study[7] have been widely cited[6,8] in support of the cost effectiveness of the WHO criteria.

In this paper we compared the cost-effectiveness of NICE 2015 and WHO 2013 diagnostic thresholds for GDM. The analysis was undertaken using a revised version of the health economic model developed for the NICE guideline and was based upon data from the UK and Australian HAPO Study centres.

Methods

Model description

A decision analytic framework was used to evaluate the cost effectiveness of two recently proposed diagnostic thresholds for GDM, together with a no diagnosis/no treatment option (See Table 1). A schematic of the model is shown in **Error! Reference source not found.**. Cost-effectiveness was evaluated using both deterministic and probabilistic sensitivity analysis.

***Figure 1**: Model Schematic*

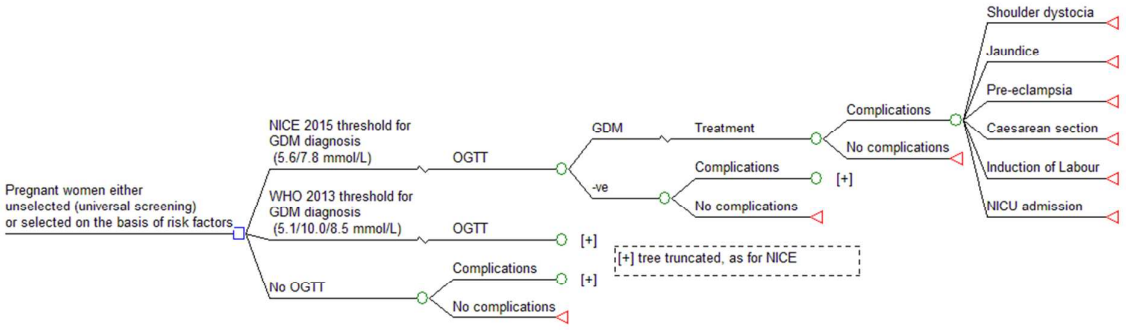


***Table 1**:* Diagnostic thresholds for plasma glucose evaluated in the economic model

Threshold name	Fasting (mmol/L)	1-hour (mmol/L)	2-hour (mmol/L)
No diagnosis/no treatment	-	-	-
NICE 2015	≥5.6	-	≥7.8
WHO 2013	≥5.1	≥10.0	≥8.5

Population

The model population comprised women of gestational age 24-28 weeks without pre-existing diabetes. The analysis utilised individual patient data from three datasets which, although not restricted to the UK, provide a representative cross section of the demographic and patient characteristics that would be found in the UK. The analyses were run separately for each dataset and, where possible, for subgroups with and without risk factors for GDM within a dataset.

i. HAPO – a dataset from the two UK (Manchester and Belfast) and two Australian (Brisbane and Newcastle) centres of the HAPO Study, referred to as HAPO (4)

ii. Norwich – these data were routinely collected between 2008 and February 2014 on women who had an oral glucose tolerance test (OGTT) on the basis of the presence of one or more risk factors for GDM. The results were obtained from laboratory records with no identifiers. Risk factors in addition to those recommended by NICE were used e.g. women with polycystic ovary syndrome, previous stillbirth or recurrent glycosuria.

iii. Atlantic Diabetes in Pregnancy (Atlantic DiP) – these data were collected between 2007 and 2013 as part of a research initiative in the Republic of Ireland intended to improve pregnancy outcomes for women with diabetes before, during and after pregnancy.

For the HAPO (4) and Atlantic DiP datasets the populations were stratified according to whether or not they had NICE risk factors for GDM (body mass index (BMI) above 30 kg/m^2, previous baby with birthweight ≥4.5 kg, previous GDM, first-degree relative with diabetes and minority ethnic family origin with a high prevalence of diabetes). This facilitated a comparison of the cost-effectiveness of universal screening for GDM when compared with a risk factor approach.

The NICE risk factor approach could not be replicated exactly because the patient data used in the model do not include information on previous offspring birth weight, and the HAPO (4) dataset does not provide information on previous GDM. Therefore, the comparison in the model was between universal screening and a subset of NICE risk factors.

Clinical outcomes

The agreed outcomes for the economic model were selected prior to model development by the NICE Guideline Development Group. They were:

i. Shoulder dystocia (SD) – this was used to estimate serious perinatal complications (SPC), a broader composite outcome (death, shoulder dystocia and birth trauma) used as a primary outcome in clinical trials. The estimation of SPC from shoulder dystocia has been described elsewhere.[5]
ii. Caesarean section (CS)
iii. Neo-natal intensive care unit (NICU) admission
iv. Jaundice requiring phototherapy (Jaund)
v. Pre-eclampsia (PE)
vi. Induction of labour (IOL)

Outcomes were prioritised for inclusion in the model if they had a direct impact on health related quality of life and/or cost. Birth weight was not included because there were few long-term outcome data for modelling any risk benefit of a reduction in birth weight for future diabetes and other health outcomes in the offspring.

In addition, outcomes were only included if the relationship with plasma glucose levels had been established in the HAPO study, and also that they had been assessed in intervention studies used to derive treatment effect size estimates. Possible double counting of certain outcomes was taken into account (e.g. preterm birth and NICU admission). The final list of outcomes included in the model was therefore a pragmatic one.

Baseline risk

Logistic regression analyses of patient data from HAPO (4) were used to predict a baseline risk for all six outcomes for each woman, based on their characteristics including their OGTT results. In the HAPO study the OGTT was blinded to the carers, unless there was overt diabetes, thus allowing direct comparison of the OGTT with perinatal outcomes without intermediate treatment effects for those meeting the new diagnostic criteria for GDM.

For each of the six outcomes, 2 logistic analyses to predict risk were assessed:

i. Prediction based on OGTT plasma glucose results and including the same covariates as used for Model 2 in the original analysis of the HAPO data[1] – this could not be applied to the Norwich and Atlantic DiP datasets as information on all HAPO covariates was not available.

ii. Prediction based only on OGTT plasma glucose results

Backward elimination of plasma glucose variables with non-significant coefficients was undertaken to arrive at a 'final' logistic regression analysis to predict baseline risk for each

outcome for the base case analysis, although a sensitivity analysis is also presented where the model was run with plasma glucose variables with non-significant coefficients retained. The logistic regression analyses used to predict the baseline risk for each outcome are shown in the Supplementary Report, Tables x1 to x6.

Clinical effectiveness

For each evaluated diagnostic threshold in

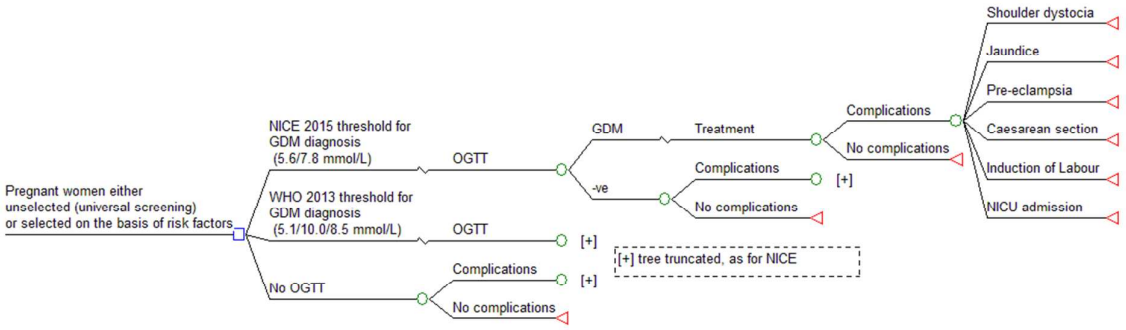


Table **1** the model determined whether a woman would be identified as having GDM based on her OGTT. If the woman was not identified as having GDM then outcome probabilities were based on the predicted baseline risk, but for women identified as having GDM the predicted baseline risk was modified to take account of the effects of treatment. Treatment effectiveness for most outcomes was estimated from a random-effects meta-analysis of two studies, Australian Carbohydrate Intolerance Study (ACHOIS) and the Landon et al. trial.[9, 10] Other published studies of treatment for GDM were adjudged to lack adequate randomisation.[11] For the NICU outcome only the Landon et al. trial data were used as it was considered to more closely represent UK practice as they utilised all neonatal nursery admissions. Similarly, the incidence of pre-eclampsia seemed high in ACHOIS in both arms, and again only Landon et al. trial data were utilised. The treatment effects for each of the model's clinical outcomes are shown in Table 2 along with parameters for probabilistic sampling. The model assumes that the relative treatment effect will be the same irrespective of the absolute baseline risk. For deterministic analyses the point estimate of relative risk was used but in order to account for uncertainty in these point estimates, these relative risks were sampled from a log-normal distribution in the simulations undertaken for probabilistic sensitivity analysis (PSA).

Table 2: Relative treatment effects for model outcomes

Outcome	Relative risk (RR)	Standard error (log RR)	Source
Shoulder dystocia	0.41	0.316	ACHOIS (2005), Landon (2009)
Caesarean section	0.88	0.095	ACHOIS (2005), Landon (2009)

NICU	0.77	0.194	Landon (2009)
Jaundice requiring phototherapy	0.83	0.136	ACHOIS (2005), Landon (2009)
Pre-eclampsia	0.46	0.345	Landon (2009)
Induction of Labour	1.16	0.126	ACHOIS (2005), Landon (2009)

207

208

Costs

Costing was undertaken from the perspective of the NHS and was calculated for each woman in the dataset being analysed and made up of three components;

- the costs of the diagnostic test – not applied in the *no test/no treat* strategy
- the costs of treatment- applied to every woman diagnosed with GDM at a particular threshold
- the costs associated with the various outcomes – with the cost for each woman being the expected (or average) cost of the outcome based on her estimated risk

The costs calculated for each woman were then summed across the entire patient dataset to give a total cost for a particular diagnostic threshold.

Costs were taken from published UK sources where possible (cost year 2015) and have not been discounted as they are all assumed to occur within 12 months of diagnosis. Model unit costs are reported in the Supplementary Report, Table x7.

Other event probabilities

Probabilities in decision analysis were used to calculate the expected costs and benefits of the various comparators. Many of these probabilities stemmed from relative treatment effects but a few additional event probabilities were included in the model in order to estimate certain costs. These probabilities are shown in Table 3 and their source is described elsewhere.[5]

Table 3: *Model event probability not derived from patient level regression*

Event	Probability
Not requiring hypoglycaemic therapy when treated for GDM	36%
Risk of hypoglycaemia if taking hypoglycaemic therapy	20%
Risk of hypoglycaemia being severe (requiring hospitalisation)	5%

Quality Adjusted Life Years (QALYs)

Following previous studies[5, 16] a QALY decrement of 2.2 was assigned to serious perinatal complications (SPC), defined as per the ACHOIS study as a composite outcome of shoulder dystocia, death and birth trauma.[9] More detail on the derivation of this QALY loss is provided in the Supplementary Report. The cost-effectiveness of a healthcare intervention is determined by the opportunity cost of the health foregone on the basis that with a fixed health budget any newly funded intervention would displace the least cost-effective treatment currently provided. In the UK, NICE typically uses a threshold of £20,000 to £30,000 per QALY as a benchmark[17] for the opportunity cost of health foregone and this paper assesses cost-effectiveness accordingly.

Sensitivity analysis

Probabilistic sensitivity analysis, using Monte Carlo simulation (with 2,000 iterations for each analysis), was undertaken in order to assess the impact of sampling uncertainty on model inputs. Parameters and distributions for the probabilistic sensitivity analysis are given in Table 2 and Table x7 in the supplementary report**Error! Reference source not found.**. For the logistic regression coefficients used to predict baseline risk, the Cholesky decomposition method[18] was

used to sample from a multivariate normal distribution in order to reflect correlations between the coefficients.

Results

Table **4** shows the percentage of women diagnosed with GDM in the three populations using both of the evaluated diagnostic thresholds. In addition, for the HAPO (4) and Atlantic DiP datasets this is additionally broken down in the subgroups with and without NICE risk factors (RF).

Table 4:. Percentage of women identified with GDM by threshold and population

Threshold name	Norwich (n=12,754)	HAPO all (n=6,163)	HAPO RF (n=3,549)	HAPO No RF (n=2,614)	DiP All (n=5,290)	DiP RF (n=1,988)	DiP No RF (n=3,302)
NICE 2015	7.0%	13.6%	17.7%	8.0%	13.1%	25.0%	5.9%
WHO 2013	13.9%	18.9%	25.7%	9.7%	21.2%	37.7%	11.2%

Detailed deterministic and probabilistic results for HAPO (4) with risk factors are shown in Table 5, Table 6, Table 7 and

Figure 2.

Table 5: *Clinical outcomes for HAPO (4) population with NICE risk factors (n=3,549)*

Diagnostic threshold	Diagnosed	SD	SPC	CS	NICU	Jaund	PE	IOL
No Treatment	0	49	67	759	345	219	146	974
NICE 2015	629	41	56	739	326	210	123	1,004
WHO 2013	912	39	54	731	321	207	117	1,016

Table 6: *Deterministic analysis for the HAPO (4 centres) population with NICE risk factors (n=3,549)*

Diagnostic threshold	Cost	QALY	Incremental cost	Incremental QALY	ICER
No Treatment	£0	0.00	n/a	n/a	n/a
NICE 2015	£546,349	23.68	£546,349	23.68	£23,073
WHO 2013	£778,993	29.86	£232,644	6.18	£37,669

Table 5 indicates that there was a relatively small difference in clinical outcomes contrasting NICE and WHO diagnostic criteria, despite there being a 45% increase in women diagnosed with GDM. Using the WHO 2013 criteria, instead of the NICE 2015 criteria, an additional 142 women would have had to be diagnosed with GDM, and treated in order to prevent 1 case of shoulder dystocia.

In the deterministic analysis the NICE 2015 diagnostic criteria would be considered cost-effective at a cost-effectiveness threshold of £30,000 per QALY (Table 6).

The probabilistic sensitivity analysis reached a similar conclusion, with the NICE 2015 diagnostic threshold having the highest probability of being the most cost-effective treatment and the highest NMB using a cost-effectiveness threshold of £30,000 per QALY (Table 7 and

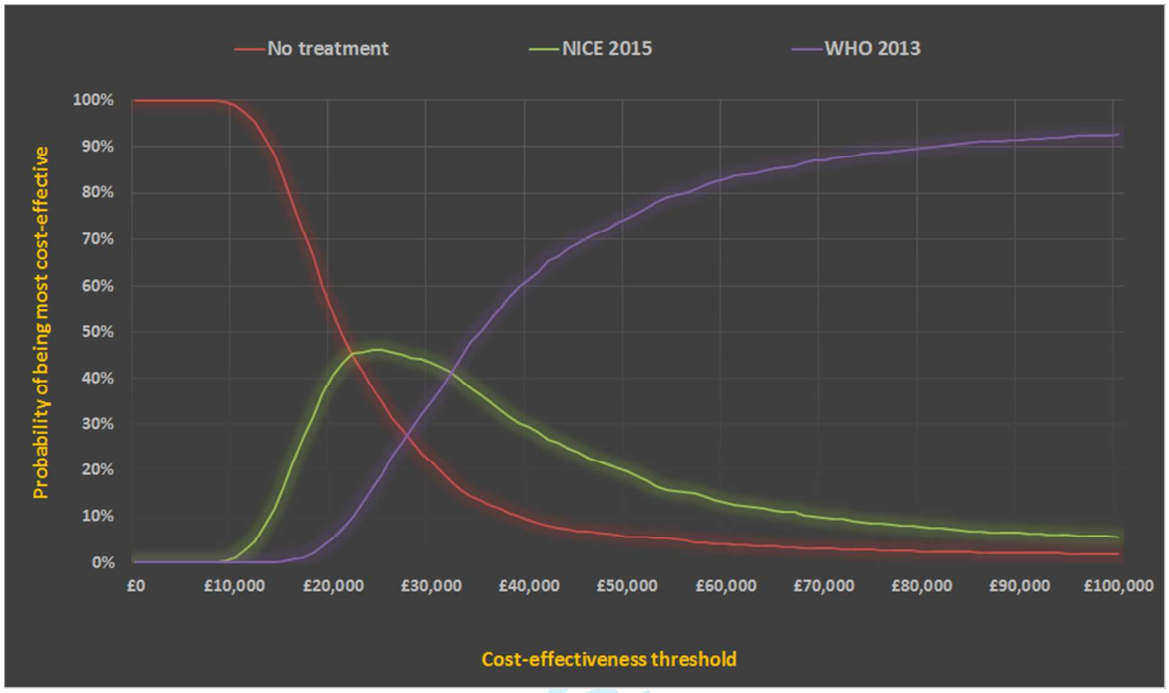
Figure 2). The analysis also suggested that no diagnosis/no treatment might be considered the most likely to be cost-effective when using a lower cost-effectiveness threshold of £20,000 per QALY. The probability of no diagnosis/no treatment being cost-effective falls sharply in the cost-effectiveness threshold range of £20,000 - £30,000 *per* QALY. As shown in the cost-effectiveness acceptability curve of

Figure 2, the WHO 2013 diagnostic threshold becomes more cost-effective as the cost-effectiveness threshold increases. Nevertheless, this would have to exceed £30,000 per QALY before becoming cost-effective, indicating that the further reduction in adverse outcomes, are achieved at an unacceptably high opportunity cost. The Supplementary Report plots the incremental cost and QALY outcomes of 2,000 simulations from the probabilistic analysis on the cost-effectiveness plane (see Figure x1). Whilst most points fall in the south-western quadrant, suggesting that WHO 2013 diagnostic criteria are likely to lead to additional QALYs when compared with NICE 2015 criteria, all points show that NICE 2015 criteria were associated with markedly lower costs.

Table 7: *Probabilistic sensitivity analysis for HAPO (4) in a population with NICE risk factors*

Diagnostic threshold	**NMB CE threshold £30,000 per QALY**	**Probability cost-effective CE threshold £20,000 per QALY**	**Probability cost-effective CE threshold WTP = £30,000 per QALY**
No Treatment	£391	54.2%	21.9%
NICE 2015	£233,192	40.7%	43.4%
WHO 2013	£200,384	5.2%	34.7%

Figure 2– Cost-effectiveness acceptability curve indicating the probability of a threshold or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for the HAPO 4 centres population with risk factors



Summaries of results for all of the model populations and more detailed results are provided in the Supplementary Report.

Table x9 and Table x10 in the Supplementary Report show that in both the HAPO (4) and Atlantic DiP populations with NICE risk factors, the NICE diagnostic threshold is the most cost-effective strategy at a cost-effectiveness threshold of £30,000 per QALY. The NICE 2015 diagnostic threshold has ICERs of less than £30,000 per QALY, and in the probabilistic sensitivity analysis it has the highest net monetary benefit and the highest probability of being the most cost-effective. For HAPO (4) the results are similar if baseline risks are estimated using logistic regression based on all covariates or a logistic regression just using plasma glucose levels

The results also suggested that universal screening would not be cost-effective as, when compared to risk factor screening (as recommended in NICE guidelines), the additional women included in such an approach would be those without risk factors and the model demonstrates that the ICERs for diagnosis and treatment are all well in excess of £30,000 per QALY; markedly so when using WHO 2013 diagnostic thresholds. These conclusions were supported by an analysis of the Norwich dataset (see Supplementary Report).

It was not possible to stratify the Norwich dataset according to risk factors, and therefore the ICERs presented relate to a comparison between no screening/treatment and universal screening and treatment. However, the results were consistent with those for HAPO (4) and Atlantic DiP. First, they showed that universal screening was not cost-effective even when compared to an alternative of no screening/no treatment. Second, the ICERs for the whole population were a weighted average of the populations with and without risk factors. The ICER for the population with risk factors would be lower than the ICER for the entire population, which was not substantially above the £30,000 per QALY threshold.

One-way sensitivity analysis

As part of a sensitivity analysis the deterministic models were re-run using the logistic regression models without backward elimination of glucose variables with non-significant coefficients, and these analyses are summarised in the Supplementary Report.

Discussion

In the NICE guideline analysis, 14 alternative diagnostic thresholds were compared and there was no single optimal diagnostic threshold which clearly emerged[5]. This is not surprising given

the small differences in patient outcomes between them. In that analysis the previous WHO 1999 criteria emerged as a relatively cost-effective strategy. However, the Guideline Committee rejected a fasting threshold of 7.0 mmol/L as there was a wide clinical consensus that this was too high, as 6.1-7.0 mmol/L is diagnostic of impaired fasting glycaemia in the non-pregnant population. Intervention studies had used a lower fasting threshold than 7.0 mmol/L as a basis for inclusion, and therefore made a case for intervention at lower levels. Based upon detailed cost effectiveness analysis of all the options, the Guideline Committee ultimately decided on recommending a fasting plasma glucose of 5.6 mmol/L and a 2 hour plasma glucose of 7.8 mmol/L. In this paper, we have restricted our analysis of cost effectiveness to the WHO 2013 and NICE 2015 criteria (with a no screening/treatment baseline also included) as these two recommendations have the most clinical currency at present.

All of the analyses presented in this paper suggest that, in a population with NICE risk factors, the NICE 2015 diagnostic criteria for GDM could be considered cost-effective relative to no screening/no treatment and to WHO 2013 diagnostic thresholds when using a cost-effectiveness threshold of £30,000 per QALY. The analyses also show that no screening/no treatment is cost-effective in populations without NICE risk factors, suggesting that universal screening does not represent value for money, at least in a UK setting.

One of the limitations of our analysis was that the 2-hour threshold was restricted to the historical WHO 1999 2-hour definition of 7.8mmol/l, or the new WHO 2013 criteria of 8.5 mmol/l. It is conceivable that a 2-hour threshold lying between these values might outperform both. Our greater focus, though was on the optimal fasting level as this is where the greatest controversy lies with respect to potentially missed treatment opportunities.

As noted by the proponents of WHO 2013 diagnostic criteria for GDM, using a lower fasting plasma glucose threshold would by definition detect more cases. Furthermore, because we assumed in the model that the relative treatment effect would be the same in additionally diagnosed cases, it follows that such a threshold could potentially yield the lowest number of adverse outcomes and the greatest QALY gain. However, our analysis suggests that the relatively small additional gains are not justified by the substantially higher costs that such lower thresholds would require.

A key driver of our results were the logistic regression models which were used to predict baseline risk. For the outcomes included in this study these regression models suggested that the 2-hour plasma glucose was a much more important predictor of adverse outcomes than the fasting plasma glucose, something we were unaware of when selecting the model's clinical outcomes.

We consider that our analysis which builds on previous modelling[5, 16] is the most comprehensive assessment of the cost-effectiveness of diagnostic thresholds for GDM yet undertaken, and will hopefully contribute to the WHO's expectation "that a substantial body of new data will emerge in the near future, providing currently scarce health and economic evaluation of the recommended criteria applied to various populations and with different approaches (universal screening, screening only women at high risk, diagnostic testing only)".

A number of commentators [19, 20] have recently advocated universal screening for GDM. The essence of the argument is based upon the number of cases of GDM that would be missed with selective screening, and the subsequent reduced opportunity to prevent a serious perinatal outcome. Of course, it is true that universal screening will detect more cases, although the

absolute numbers will depend upon the thresholds used to define GDM. Table 5 shows that many more women would need to be diagnosed in order to prevent a single adverse outcome. However, in the context of finite health care resources, it must be accepted that it may be cost-effective to miss some cases. Epidemiological measures such as number needed to treat (or number needed to screen in this case) implicitly recognise that a goal of health care systems cannot be to maximize health gain without any consideration of cost. Identifying missed cases carries an opportunity cost and it may be that those resources would achieve greater benefit if employed elsewhere in the health care system. If a population is divided into those with risk factors and those without risk factors, then the prevalence of GDM must be lower in the group without risk factors (and the number needed to screen higher) with concomitantly lower cost-effectiveness. However, the comparative cost-effectiveness of screening in those with and without risk factors is not only affected by the respective prevalence in the two groups, but also differences in severity. In those diagnosed with GDM and who had risk factors there were, as anticipated, greater levels of hyperglycaemia than in those without risk factors. As shown in Table x24 in the Supplementary Report,**Error! Reference source not found.Error! Reference source not found.** 'true positives' or identified cases (risk factor present and GDM) had higher plasma glucose values than 'false negatives' or missed cases (risk factors absent and GDM) when defining GDM positives according to WHO 2013 diagnostic thresholds.

We would therefore expect the women with risk factors and GDM to be at greater risk of adverse outcomes than the women with GDM without risk factors as a result of their higher plasma glucose levels. So the "cases" missed with selective screening would have, on average, fewer adverse outcomes than in "cases" in a population with risk factors. So the ICER would be greater in the population without risk factors because prevalence is lower and cases have fewer adverse outcomes.

Our analysis, by splitting the HAPO (4) and Atlantic DiP datasets into those with and without risk factors, was able to evaluate the cost-effectiveness of moving from risk factor screening to universal screening. Whilst diagnosis in populations with risk factors was shown to be cost-effective at a threshold of £30,000 per QALY, it was never cost-effective to diagnose and treat in those without risk factors. Table 4 indicates the large differences that exist in prevalence between the populations with and without risk factors. Our analysis suggests that the cost-effectiveness threshold would have to substantially exceed currently accepted UK norms for universal screening to be considered cost-effective. Although the NICE risk factor approach could not be replicated exactly, we felt that the approximation used was acceptable, as the only women who would be omitted from the model risk factor population were multiparous and would have had a large baby previously and/or a past history of GDM. This approximation would over-estimate slightly the benefits of universal screening, as the baseline risk in a group designated as being without NICE risk factors present would be over-stated.

A previous study[7] from Spain using WHO 2013 diagnostic criteria suggested cost effectiveness compared with a two-step protocol using the Carpenter – Coustan thresholds. However, this was largely based upon estimates of reduction of caesarean section rates of 50% which we find implausible based upon changes in diagnostic criteria alone, noting that ACHOIS and Landon et al. found only a 4% and 21% reduction in caesarean section respectively as a result of treating gestational diabetes. The Spanish study did not consider other alternative thresholds, and was a retrospective, before and after analysis which has been criticized by the Cochrane Collaboration as it does not control for possible changes in important variables, such as clinical management, over time.[21]

Our model has a number of limitations particularly with respect to the valuation of health outcomes. We did not include large for gestational age as an outcome because it was felt that shoulder dystocia was the relevant immediate complication of interest, and that possible long term metabolic consequences for the offspring were hard to quantify and therefore difficult to incorporate within the model. As previously noted, the QALY loss from a serious perinatal complication used in this analysis is likely to be overstated because of the relatively large weight given to death based on the intervention studies.[16] HAPO failed to show an association between perinatal mortality and plasma glucose levels, which may mean that perinatal mortality reduction is less amenable to reduction by treatment than other serious perinatal complications. In this respect the cost-effectiveness of diagnosing and treating GDM may be over-stated. On the other hand, the model does not take account of any potential long term effects on the offspring (e.g. adiposity and the likelihood of subsequent pathology) as these effects are difficult to quantify but may under-estimate the QALY gain from diagnosis and treatment. A US study[22] considered the potential long-term benefits to the mother whereby a diagnosis of GDM averts or delays onset of Type 2 diabetes mellitus, but this was not incorporated into our model as we did not consider that the relationship was sufficiently well established at this time. However, to the extent that such a relationship does exist our model would also underestimate the QALY gain from a diagnosis of GDM. A recent review has, however, questioned the association between maternal glycaemia and subsequent cardio-metabolic outcomes in offspring in humans.[23]

Despite these caveats, we feel our analysis represents a robust analysis of the cost-effectiveness of WHO 2013 diagnostic thresholds for GDM based upon current understanding of the impact of intervention in women with GDM, at least in the UK population using NICE criteria for assessing cost-effectiveness. We acknowledge completely that this analysis cannot be the final

word on the subject, and that further health economic evaluation is required to either corroborate our findings or to challenge them. Nevertheless, we feel that our analysis represents a constructive and evidence based contribution to establishing cost effective diagnostic thresholds for GDM and will hopefully lead to more research to clarify this important but vexed area of clinical diagnosis.

Conclusions

The results presented in this analysis, based on a UK setting, do not suggest that the diagnostic thresholds for GDM adopted by the WHO are cost-effective. On the other hand they do provide some support for the cost-effectiveness of the diagnostic criteria adopted by NICE when compared to either no screening/treatment and to WHO 2013 diagnostic criteria. Furthermore, according to this analysis, universal screening would seem to offer poor value for money and does not appear cost-effective compared to the current NICE guidance of targeting high risk women.

Acknowledgements

We are grateful to Professor DR McCance and Professor HD McIntyre for allowing us to use their local datasets from the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) trial and to Professor F Dunne for allowing us to use her Atlantic DiP dataset

We are also grateful to Professor David James who provided clinical support during the development of the updated NICE guideline on Diabetes in Pregnancy

Competing interests

All authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/coi_disclosure.pdf and declare:

Funding: Some of this work was undertaken by the now defunct National Collaborating Centre for Women's and Children's Health (subsumed within the National Guideline Alliance from 1st April 2016), which received funding from NICE. The views expressed in this publication are those of the authors and not necessarily those of the institute. Revisions to the guideline model after the guideline was published and drafting of the manuscript was done in the author's own time and was not funded

National Institute for Health and Care Excellence (2015). Diabetes in pregnancy: management from preconception to the postnatal period. Available from https://www.nice.org.uk/guidance/ng3

PBJ and SBR are employees of the National Guideline Alliance (part of the RCOG), which receives its funding from NICE.

MJAM, KS, AD and RWB received travel expenses from NICE for attending clinical guideline development meetings

Author contribution

Paul Jacklin designed and developed the health economic model, undertook the health economic analysis, wrote the first draft of the manuscript and incorporated edits from co-authors. Mike Maresh provided clinical input into the design of the health economic model; read, commented and edited various draft of the manuscripts. Katharine Stanley supplied the Norwich dataset, provided clinical input into the design of the health economic model; read, commented and edited various draft of the manuscripts. Anne Dornhorst provided clinical input into the design of the health economic model; read, commented and edited various draft of the manuscripts. Chris Patterson provided statistical advice, undertook statistical analysis of the HAPO dataset; read, commented and edited various drafts of the manuscript. Shona Burman Roy reviewed the clinical literature, contributed to discussions of model design; read, commented and edited various drafts of the manuscript. Rudy Bilous chaired the NICE guideline, provided clinical input into the design of the health economic model; read, commented and edited various draft of the manuscripts

Transparency declaration

The lead author, Paul Jacklin, affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Exclusive License

"I Paul Jacklin The Corresponding Author of this article contained within the original manuscript which includes any diagrams & photographs within and any related or stand alone film submitted (the Contribution") has the right to grant on behalf of all authors and does grant

on behalf of all authors, a licence to the BMJ Publishing Group Ltd and its licencees, to permit this Contribution (if accepted) to be published in the BMJ and any other BMJ Group products and to exploit all subsidiary rights, as set out in our licence set out at: http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/copyright-open-access-and-permission-reuse."

Data sharing Statement

Potential for data sharing (the health economic model) can be discussed with study investigators.

References

1. Metzger BE, Lowe LP, Dyer AR, et al. Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med* 2008;358:1991-2002.
2. 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.
3. WHO Health organisation, 2013. Diagnostic criteria and classification of hyperglycaemia first detected in pregnancy WHO/NMH/MND/13.2 http://apps.who.int/iris/bitstream/10665/85975/1/WHO_NMH_MND_13.2_eng.pdf.
4. American College of Obstetricians and Gynecologists. *Practice Bulletin 137: Mellitus. Obstet Gynecol* 2013; 122: 406-16
5. National Institute for Health and Care Excellence (NICE) (2015) Diabetes in pregnancy: management of diabetes and its complications from preconception to the postnatal period. Clinical guideline NG3 (2015). Available from www.nice.org.uk/guidance/ng3/resources/diabetes-in-pregnancy-management-of-diabetes-and-itscomplications-from-preconception-to-the-postnatal-period-51038446021, accessed February 2016
6. Meek CL, Lewis HB, Patient C,et al. Diagnosis of gestational diabetes: falling through the net. *Diabetologia* 2015;Sep;58(9):2003-12
7. Duran A, Sáenz S, Torrejón MJ et al. Introduction of IADPSG criteria for the screening and diagnosis of gestational diabetes mellitus results in improved pregnancy outcomes at a lower cost in a large cohort of pregnant women: the St. Carlos Gestational Diabetes Study. *Diabetes Care* 2014;37(9):2442–2450. doi: 10.2337/dc14-0179.
8. Herman WH. Insights offered by economic analyses. *Diabetes Care*. 2014 Sep;37(9):2424-6. doi: 10.2337/dc14-1232.

9. 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.
10. 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
11. Horvath K, Koch K, Jeitler K et al. Effects of treatment in women with gestational diabetes mellitus: systematic review and meta-analysis. *BMJ (Clin Res Ed)* 2010;340:c1395
12. British National Formulary. July 2016. https://www.medicinescomplete.com/mc/bnf/current/ (accessed 4 Aug 2016).
13. Unit Costs of Health and Social Care 2015. Personal Social Services Research Unit, The University of Kent, 2015.
14. Department of Health. NHS reference costs: financial year 2014–2015. https://www.gov.uk/government/publications/nhs-reference-costs-2014-to-2015 , Department of Health, 2015.
15. NHS Electronic Drug Tariff, August 2016. http://www.drugtariff.nhsbsa.nhs.uk/#/00336026-DD_1/DD00336022/Home (accessed 4 Aug 2016).
16. Round,J.A., Jacklin,P., Fraser,R.B., Hughes,R.G., Mugglestone,M.A., Holt,R.I., Screening for gestational diabetes mellitus: cost-utility of different screening strategies based on a woman's individual risk of disease, *Diabetologia* 2011,54(2), 256-263. doi: 10.1007/s00125-010-1881-y
17. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. October 2014 (https://www.nice.org.uk/media/default/about/what-we-do/our-programmes/developing-nice-guidelines-the-manual.pdf)
18. Briggs A, Claxton K, Sculpher M. Decision Modelling for Health Economic Evaluation. Oxford: Oxford University Press; 2006
19. Avalos GE, Owens LA, Dunne F et al. Applying Current Screening Tools for Gestational Diabetes Mellitus to a European Population: Is It Time for Change? *Diabetes Care*. 2013 Oct;36(10):3040-4. doi: 10.2337/dc12-2669. Epub 2013 Jun 11.

20. Simmons D, Moses RG. Gestational Diabetes Mellitus: To Screen or Not to Screen? Is this really still a question? Diabetes Care. 2013 Oct;36(10):2877-2878

21. Armstrong R, Waters E, Doyle J (editors). Chapter 21: Reviews in health promotion and public health. In Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org. Accessed June 2015

22. Werner EF, Pettker CM, Zucerwise et al. Screening for gestational diabetes mellitus: are the criteria proposed by the international association of the Diabetes and Pregnancy Study Groups cost-effective? Diabetes Care. 2012 Mar;35(3):529-35. doi: 10.2337/dc11-1643.

23. Donovan LE, Cundy T. Does exposure to hyperglycaemia in utero increase the risk of obesity and diabetes in the offspring? A critical reappraisal. Diabetic Medicine. 2015 Mar;32(3):295-304. doi: 10.1111/dme.12625. Epub 2014 Dec 17.

Supplementary Report

This supplementary document provides further details about model parameter estimates and model results.

Multivariable prediction models to estimate baseline risk

Model 1 includes all three blood glucose regression coefficients while Model 2 includes only blood glucose regression coefficients which remained significant after performing variable selection by backward elimination.

Table x1. Logistic regression models to predict neonatal shoulder dystocia

	Co-efficient b (Standard error (SE(b))			
	Model with blood glucose covariates		Model with all covariates	
Variable	**Model 1[a]**	**Model 2[b]**	**Model 1[a]**	**Model 2[b]**
Centre (Manchester v Belfast)	-	-	1.151 (0.424)	1.151 (0.423)
(Brisbane v Belfast)	-	-	0.562 (0.491)	0.505 (0.489)
(Newcastle v Belfast)	-	-	1.622 (0.472)	1.604 (0.472)
Age at OGTT (yr)	-	-	-0.022 (0.024)	-0.023 (0.024)
BMI at OGTT (kg/m^2)	-	-	-0.011 (0.024)	-0.006 (0.023)
Smoker (Yes v No)	-	-	-0.477 (0.409)	-0.480 (0.409)
Drinker (Yes v No)	-	-	-0.107 (0.317)	-0.101 (0.317)
Family history DM (Yes v No)	-	-	-0.008 (0.187)	-0.006 (0.184)
Gestational age at OGTT (wk)	-	-	-0.114 (0.092)	-0.111 (0.091)
Neonate gender (F v M)	-	-	-1.316 (0.292)	-1.321 (0.292)
Family history HBP (Yes v No)[c]	-	-	-	-
Maternal UTI (Yes v No)[c]	-	-	-	-
Mean Blood Pressure (mmHg)	-	-	-0.007 (0.015)	-0.006 (0.015)
Hospital admission before delivery (Yes v No)	-	-	0.175 (0.267)	0.173 (0.266)
Parity (1 v 0)	-	-	-0.108 (0.420)	-0.118 (0.420)
(2+ v 0)	-	-	0.469 (0.414)	0.456 (0.412)
(Unknown v 0)	-	-	-0.013 (0.399)	-0.026 (0.399)
Fasting blood glucose[d]	0.166 (0.110)	-	0.151 (0.112)	-
1-hr blood glucose[d]	-0.152 (0.163)	-	-0.138 (0.165)	-
2-hr blood glucose[d]	0.265 (0.151)	0.267 (0.097)	0.222 (0.152)	0.223 (0.100)
Constant	-4.475 (0.122)	-4.467 (0.122)	1.139 (3.508)	0.925 (3.025)

(a) Sensitivity analysis
(b) Base case analysis
(c) Omitted from HAPO model for shoulder dystocia
(d) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for shoulder dystocia arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x2. Logistic regression models to predict caesarean section

	Co-efficient b (Standard error (SE(b))			
	Model with blood glucose covariates		Model with all covariates	
Variable	**Model 1[a]**	**Model 1[a]**	**Model 1[a]**	**Model 1[a]**
Centre (Manchester v Belfast)	-	-	-0.495 (0.092)	-0.494 (0.092)
(Brisbane v Belfast)	-	-	-0.114 (0.100)	-0.099 (0.098)
(Newcastle v Belfast)	-	-	-0.692 (0.141)	-0.681 (0.140)
Age at OGTT (yr)	-	-	0.034 (0.007)	0.034 (0.007)
BMI at OGTT (kg/m^2)	-	-	0.039 (0.007)	0.039 (0.007)
Smoker (Yes v No)	-	-	-0.292 (0.106)	-0.304 (0.106)
Drinker (Yes v No)	-	-	-0.025 (0.087)	-0.028 (0.087)
Family history DM (Yes v No)	-	-	0.052 (0.057)	0.050 (0.057)
Gestational age at OGTT (wk)	-	-	0.004 (0.029)	0.004 (0.029)
Neonate gender (F v M)	-	-	-0.205 (0.071)	-0.205 (0.071)
Family history HBP (Yes v No)[c]	-	-	-	-
Maternal UTI (Yes v No)[c]	-	-	-	-
Mean Blood Pressure (mmHg)	-	-	0.003 (0.004)	0.003 (0.004)
Hospital admission before delivery (Yes v No)	-	-	0.510 (0.079)	0.514 (0.079)
Parity (1 v 0)[c]	-	-	-	-
(2+ v 0)[c]	-	-	-	-
(Unknown v 0)[c]	-	-	-	-
Fasting blood glucose[d]	0.053 (0.040)	-	-0.009 (0.044)	-
1-hr blood glucose[d]	0.119 (0.048)	0.138 (0.046)	0.101 (0.051)	0.144 (0.037)
2-hr blood glucose[d]	0.113 (0.046)	0.123 (0.046)	0.071 (0.048)	
Constant	-1.433 (0.035)	-1.435 (0.035)	-3.509 (0.950)	-3.518 (0.947)

(a) Sensitivity analysis
(b) Base case analysis
(c) Omitted from HAPO model for caesarean section
(d) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for caesarean section arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x3. Logistic regression models to predict neonatal intensive care unit admissions

	Co-efficient b (Standard error (SE(b))			
	Model with blood glucose covariates		Model with all covariates	
Variable	**Model 1[a]**	**Model 1[a]**	**Model 1[a]**	**Model 1[a]**
Centre (Manchester v Belfast)	-	-	0.889 (0.159)	0.894 (0.159)
(Brisbane v Belfast)	-	-	1.400 (0.163)	1.393 (0.161)
(Newcastle v Belfast)	-	-	1.163 (0.191)	1.153 (0.161)
Age at OGTT (yr)	-	-	0.012 (0.009)	0.013 (0.009)
BMI at OGTT (kg/m^2)	-	-	0.024 (0.009)	0.025 (0.009)
Smoker (Yes v No)	-	-	0.201 (0.130)	0.209 (0.130)
Drinker (Yes v No)	-	-	-0.023 (0.117)	-0.025 (0.117)
Family history DM (Yes v No)	-	-	0.038 (0.069)	0.033 (0.069)
Gestational age at OGTT (wk)	-	-	-0.052 (0.038)	-0.050 (0.038)
Neonate gender (F v M)	-	-	-0.302 (0.094)	-0.304 (0.094)
Family history HBP (Yes v No)[c]	-	-	-	-
Maternal UTI (Yes v No)[c]	-	-	-	-
Mean Blood Pressure (mmHg)	-	-	0.006 (0.006)	0.006 (0.006)
Hospital admission before delivery (Yes v No)	-	-	0.792 (0.097)	0.794 (0.148)
Parity (1 v 0)	-	-	-0.474 (0.148)	-0.474 (0.148)
(2+ v 0)	-	-	-0.493 (0.157)	-0.490 (0.157)
(Unknown v 0)	-	-	-0.086 (0.135)	-0.084 (0.135)
Fasting blood glucose[d]	-0.025 (0.050)	-	-0.003 (0.054)	-
1-hr blood glucose[d]	0.078 (0.064)	-	0.082 (0.067)	-
2-hr blood glucose[d]	0.167 (0.060)	0.208 (0.041)	0.107 (0.063)	0.159
Constant	-2.375 (0.046)	-2.374 (0.046)	-3.061 (1.243)	-3.181 (1.236)

(a) Sensitivity analysis
(b) Base case analysis
(c) Omitted from HAPO model for neonatal intensive care unit admissions
(d) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for neonatal intensive care unit admissions arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x4. Logistic regression models to predict jaundice

	Co-efficient b (Standard error (SE(b))			
	Model with blood glucose covariates		Model with all covariates	
Variable	**Model 1[a]**	**Model 1[a]**	**Model 1[a]**	**Model 1[a]**
Centre (Manchester v Belfast)	-	-	0.410 (0.157)	0.407 (0.157)
(Brisbane v Belfast)	-	-	0.420 (0.173)	0.449 (0.171)
(Newcastle v Belfast)	-	-	-0.332 (0.259)	-0.315 (0.259)
Age at OGTT (yr)	-	-	-0.005 (0.011)	0.005 (0.011)
BMI at OGTT (kg/m^2)	-	-	-0.009 (0.012)	-0.011 (0.011)
Smoker (Yes v No)	-	-	-0.093 (0.162)	0.082 (0.161)
Drinker (Yes v No)	-	-	-0.508 (0.163)	-0.514 (0.163)
Family history DM (Yes v No)	-	-	-0.060 (0.094)	-0.060 (0.094)
Gestational age at OGTT (wk)	-	-	-0.077 (0.047)	-0.078 (0.047)
Neonate gender (F v M)	-	-	-0.115 (0.113)	-0.116 (0.113)
Family history HBP (Yes v No)[c]	-	-	-	-
Maternal UTI (Yes v No)[c]	-	-	-	-
Mean Blood Pressure (mmHg)	-	-	0.018 (0.007)	0.018 (0.007)
Hospital admission before delivery (Yes v No)	-	-	0.865 (0.116)	0.867 (0.116)
Parity (1 v 0)	-	-	-0.380 (0.185)	-0.382 (0.185)
(2+ v 0)	-	-	-0.526 (0.200)	-0.526 (0.200)
(Unknown v 0)	-	-	0.078 (0.165)-	0.078 (0.165)
Fasting blood glucose[d]	-0.063 (0.061)	-	-0.055 (0.066)	-
1-hr blood glucose[d]	0.199 (0.078)	0.237 (0.052)	0.192 (0.079)	0.216 (0.056)
2-hr blood glucose[d]	0.102 (0.072)	-	0.073 (0.074)	
Constant	-2.850 (0.057)	-2.846 (0.057)	-2.014 (1.526)	-1.927 (1.522)

(a) Sensitivity analysis
(b) Base case analysis
(c) Omitted from HAPO model for jaundice
(d) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for jaundice arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x5. Logistic regression models to predict pre-eclampsia

	Co-efficient b (Standard error (SE(b))			
	Model with blood glucose covariates		Model with all covariates	
Variable	**Model 1[a]**	**Model 1[a]**	**Model 1[a]**	**Model 1[a]**
Centre (Manchester v Belfast)	-	-	-0.800 (0.193)	-0.794 (0.192)
(Brisbane v Belfast)	-	-	-0.277 (0.202)	-0.308 (0.200)
(Newcastle v Belfast)	-	-	-0.667 (0.278)	-0.685 (0.278)
Age at OGTT (yr)	-	-	-0.011 (0.015)	-0.009 (0.015)
BMI at OGTT (kg/m^2)	-	-	0.097 (0.012)	0.101 (0.011)
Smoker (Yes v No)	-	-	-0.569 (0.246)	-0.556 (0.245)
Drinker (Yes v No)	-	-	-0.168 (0.194)	-0.170 (0.194)
Family history DM (Yes v No)	-	-	0.006 (0.127)	-0.004 (0.127)
Gestational age at OGTT (wk)	-	-	-0.096 (0.059)	-0.092 (0.059)
Neonate gender (F v M)	-	-	0.174 (0.147)	0.173 (0.147)
Family history HBP (Yes v No)	-	-	0.230 (0.150)	0.233 (0.150)
Maternal UTI (Yes v No)	-	-	0.721 (0.211)	0.734 (0.211)
Mean Blood Pressure (mmHg) [c]	-	-	-	-
Hospital admission before delivery (Yes v No) [c]	-	-	-	-
Parity (1 v 0)	-	-	-0.292 (0.240)	-0.291 (0.240)
(2+ v 0)	-	-	-0.703 (0.271)	-0.701 (0.271)
(Unknown v 0)	-	-	0.023 (0.224)	0.026 (0.224)
Fasting blood glucose[d]	0.183 (0.068)	0.201 (0.065)	0.062 (0.078)	-
1-hr blood glucose[d]	0.083 (0.098)	-	0.065 (0.104)	-
2-hr blood glucose[d]	0.150 (0.090)	0.196 (0.072)	0.195 (0.096)	0.272 (0.067)
Constant	-3.455 (0.075)	-3.453 (0.075)	-3.107 (1.855)	-3.370 (1.842)

(a) Sensitivity analysis
(b) Base case analysis
(c) Omitted from HAPO model for pre-eclampsia
(d) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for pre-eclampsia arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x6. Logistic regression models to predict induction of labour

	Co-efficient b (Standard error (SE(b))		
	Model with blood glucose covariates	Model with all covariates	
Variable	**Model 1[a]**	**Model 1[a]**	**Model 1[a]**
Centre (Manchester v Belfast)	-	-0.476 (0.077)	-0.476 (0.077)
(Brisbane v Belfast)	-	-0.333 (0.087)	-0.337 (0.085)
(Newcastle v Belfast)	-	-0.384 (0.110)	-0.387 (0.109)
Age at OGTT (yr)	-	0.006 (0.006)	0.006 (0.006)
BMI at OGTT (kg/m^2)	-	0.039 (0.006)	0.039 (0.006)
Smoker (Yes v No)	-	0.051 (0.082)	0.051 (0.082
Drinker (Yes v No)	-	0.079 (0.072)	0.079 (0.072)
Family history DM (Yes v No)	-	0.016 (0.048)	0.016 (0.048)
Gestational age at OGTT (wk)	-	0.011 (0.024)	0.011 (0.024)
Neonate gender (F v M)	-	-0.038 (0.059)	-0.038 (0.059)
Family history HBP (Yes v No)[c]	-	-	-
Maternal UTI (Yes v No)[c]	-	-	-
Mean Blood Pressure (mmHg)	-	0.008 (0.004)	0.008 (0.004)
Hospital admission before delivery (Yes v No)	-	0.608 (0.066)	0.608 (0.066)
Parity (1 v 0)	-	-0.363 (0.101)	-0.363 (0.101)
(2+ v 0)	-	-0.193 (0.105)	-0.193 (0.105)
(Unknown v 0)	-	0.141 (0.094)	0.141 (0.094)
Fasting blood glucose[d]	0.079 (0.033)	0.009 (0.037)	-
1-hr blood glucose[d]	-0.093 (0.041)	-0.111 (0.043)	-0.108 (0.041)
2-hr blood glucose[d]	0.100 (0.040)	0.094 (0.041)	0.096 (0.041)
Constant	-1.032 (0.029)	-3.037 (0.796)	-3.050 (0.794)

(a) Sensitivity analysis
(b) Base case analysis
(c) Omitted from HAPO model for induction of labour
(d) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for induction of labour arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x7: Model unit costs

Category	Cost	Standard Error	Distribution	Source
2 sample OGTT	£8.07	n/a	n/a	NICE 2015[5]
3 sample OGTT	£12.11	n/a	n/a	NICE 2015[5]
Rapilose OGTT solution	£3.48	n/a	n/a	BNF July 2016[12]
Health Care Assistant Band 3 (per hour)	£25	n/a	n/a	Unit Costs of Health and Social Care 2015[13]
Nurse Band 7 (per hour of patient contact)	£147	n/a	n/a	Unit Costs of Health and Social Care 2015[13]
Dietician	£38	n/a	n/a	Unit Costs of Health and Social Care 2015[13]
Ante-natal appointment	£96	£9.07	Normal	NHS Reference Costs 2014-15[14]
Ultrasound scan	£112	£7.65	Normal	NHS Reference Costs 2014-15[14]
Rapid acting insulin	£0.02	n/a	n/a	BNF June 2016[12]
Regular insulin	£0.02	n/a	n/a	BNF June 2016[12]
Needles	£0.10	n/a	n/a	NHS Drugs Tariff June 2016[15]
Lancets	£0.03	n/a	n/a	NHS Drugs Tariff June 2016[15]
Strips	£0.18	n/a	n/a	NHS Drugs Tariff June 2016[15]
Treatment of GDM	£987	n/a	n/a	Calculated
Severe hypoglycaemia	£650	n/a	n/a	NHS Reference Costs 2014-15[14]
Admission to NICU	£1,176	£38	Normal	NHS Reference Costs 2014-15[14]
Caesarean section	£982	£80	Normal	NHS Reference Costs 2014-15[14]
Neonatal death	£777	£39	Normal	NHS Reference Costs 2014-15[14]
Shoulder dystocia	£1,394	£79	Normal	NHS Reference Costs 2014-15[14]
Birth trauma	£1,394	£79	Normal	NHS Reference Costs 2014-15[14]
Serious perinatal complication (death, shoulder dystocia, birth trauma)	£1,347	n/a	n/a	Calculated
Phototherapy	£788	£72	Normal	NHS Reference Costs 2014-15[14]
Pre-eclampsia	£4,750	n/a	n/a	NICE 2015[5]

QALYs

A QALY loss was estimated for each individual component (shoulder dystocia, death and birth trauma) of the composite serious perinatal outcome, which was used in the ACHOIS study.[9] A weighting for each individual component was derived according to their relative frequency in the selected studies to assess treatment effectiveness.[9, 10] These were then used in order to derive a weighted average for a serious perinatal complication as shown in Table x8. QALY losses from a serious perinatal complication could be experienced over a lifetime and therefore an annual discount rate of 3.5% was applied in line with NICE methods.[13] For each patient, an expected QALY decrement is calculated based on their risk of serious perinatal complications. These individual patient QALY decrements are then summed across all patients to give the total QALY decrement for the patient dataset for each different diagnostic threshold.

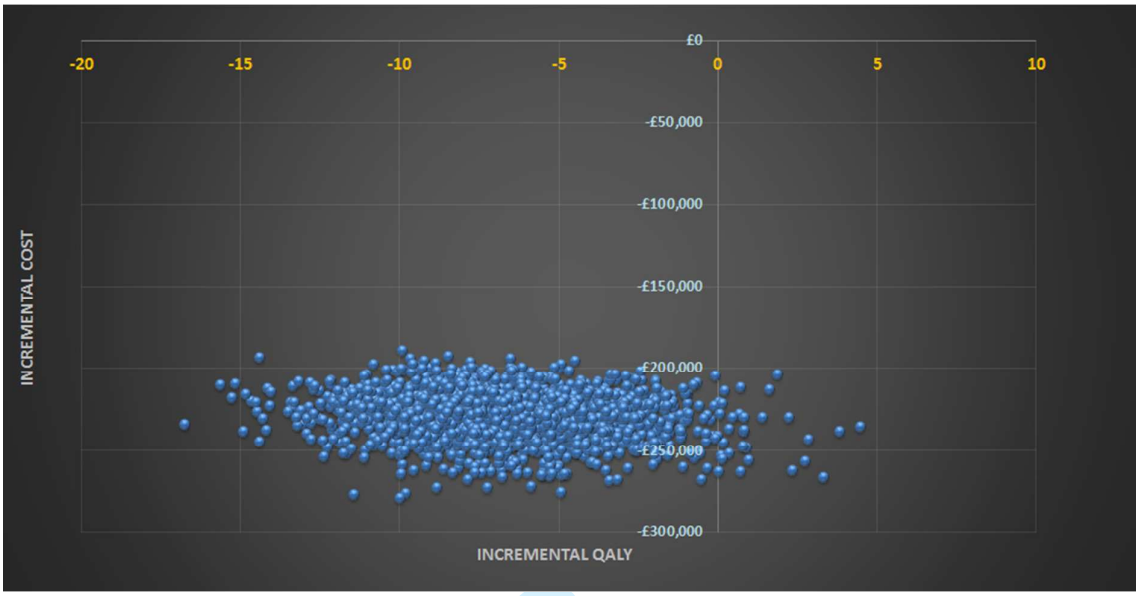
Table x8: QALY losses and weights from individual components of the composite outcome of serious perinatal complications

Complication	Weight	QALY	Weighted QALY
Death	0.08	25	2.00
Shoulder dystocia	0.73	0.2	0.15
Birth trauma	0.20	0.2	0.04

The analyses presented in this paper include a maternal health state utility which was estimated from quality of life data collected as part of the ACHOIS study. Whilst treatment conferred a small benefit in maternal health state utility, this was small in comparison to QALYs derived from infant outcomes. The value of the maternal health state utility with and without treatment is the same as has been used previously[5].

Results for the HAPO (4) population with risk factors

Figure x1: Cost-effectiveness plane for NICE 2015 compared with WHO 2013 for HAPO (4) with risk factors



Summary of Results for each model population

Table x9: Summary of deterministic ICERs for each population with backward elimination of plasma glucose variables with non-significant coefficients

	All covariates		Plasma glucose covariates				
Diagnostic threshold	**HAPO Risk factor (n=3,549)**	**HAPO No Risk factor (n=2,614)**	**HAPO Risk factor (n=3,549)**	**HAPO No Risk factor (n=2,614)**	**Atlantic DiP Risk factor (n=1,988)**	**Atlantic DiP No Risk factor (n=3,302)**	**Norwich (n=12,754)**
No Treatment	-	-	-	-	-	-	-
NICE 2015	£23,073	£43,845	£25,434	£35,230	£23,755	£35,732	£33,177
WHO 2013	£37,669	£220,638	£41,631	£97,941	£42,457	£45,075	£42,931

Table x10: Probability that a threshold is cost-effective at a threshold of £30,000 per QALY and the net monetary benefit in each population using regression models with backward elimination of plasma glucose variables with non-significant coefficients

	All covariates		Plasma glucose covariates				
Diagnostic threshold	**HAPO Risk factor (NMB)**	**HAPO No Risk factor (NMB)**	**HAPO Risk factor (NMB)**	**HAPO No Risk factor (NMB)**	**Atlantic DiP Risk factor (NMB)**	**Atlantic DiP No Risk factor (NMB)**	**Norwich (NMB)**
No Treatment	21.9% (£391)	78.7% (£223)	33.7% (£845)	66.9% (£392)	30.2% (£518)	69.0% (£402)	59.7% (£1,141)
NICE 2015	43.4% (£233,192)	20.9% (-£57,742)	45.6% (£111,502)	31.0% (-£33,767)	50.8% (£116,178)	18.0% (-£36,481)	23.2% (-£76,289)
WHO 2013	34.7% (£201,384)	0.5% (-£94,754)	20.8% (£45,208)	2.2% (-£61,385)	19.1% (£53,129)	13.1% (-£88,283)	17.2% (-£300,254)

Results for the HAPO (4) population without risk factors

Table x11: Clinical outcomes for HAPO (4) population without NICE risk factors (n=2,614)

Diagnostic threshold	Diagnosed	SD	SPC	CS	NICU	Jaund	PE	IOL
No Treatment	0	24	34	466	188	126	55	647
NICE 2015	208	23	31	460	184	124	51	655
WHO 2013	253	23	31	459	184	123	51	657

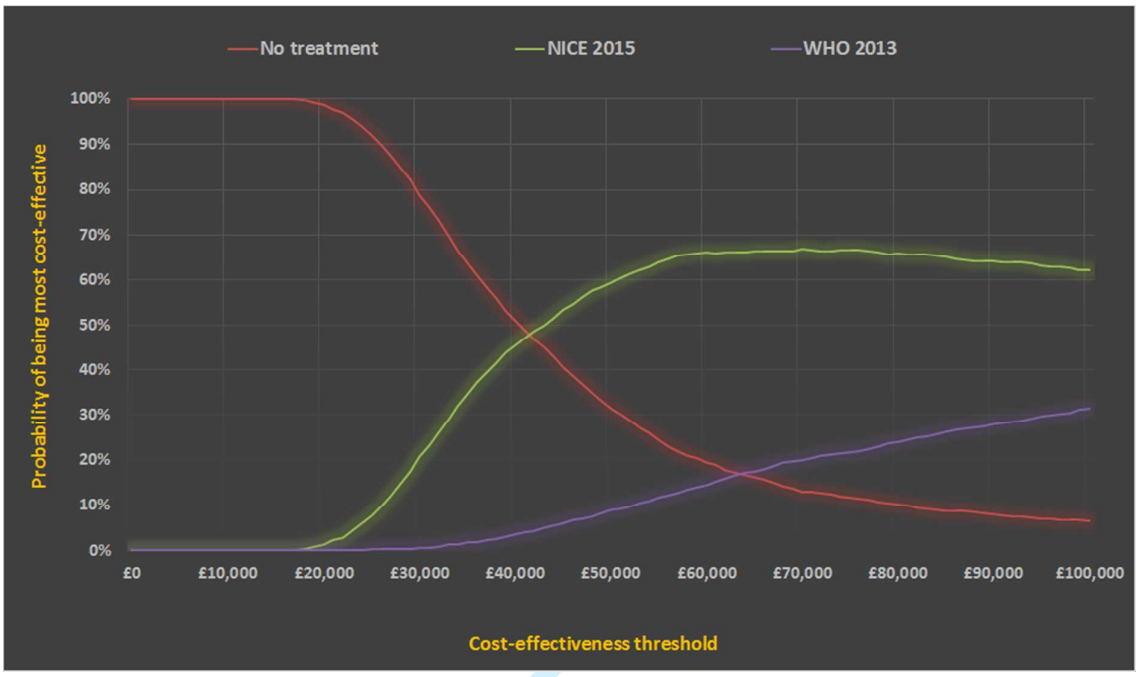
Table x12: Deterministic analysis for HAPO (4) population without NICE risk factors (n=2,614)

Diagnostic threshold	Cost	QALY	Incremental cost	Incremental QALY	ICER
No Treatment	£0	0.00	n/a	n/a	n/a
NICE 2015	£238,074	5.43	£238,074	5.43	£43,845
WHO 2013	£281,357	5.63	£43,283	0.20	£220,638

Table x13: Probabilistic sensitivity analysis for HAPO (4) in a population without NICE risk factors

Diagnostic threshold	NMB CE threshold £30,000 per QALY	Probability cost-effective CE threshold £30,000 per QALY
No Treatment	£223	78.9%
NICE 2015	-£57,742	20.9%
WHO 2013	-£94,754	0.5%

Figure x2: Cost-effectiveness acceptability curve indicating the probability of a threshold or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for HAPO (4) population without risk factors



Results for the Atlantic DiP population with risk factors

Table 14: Clinical outcomes for Atlantic DiP population with NICE risk factors (n=1,988)

Diagnostic threshold	Diagnosed	SD	SPC	CS	NICU	Jaund	PE	IOL
No Treatment	0	25	34	408	177	122	73	522
NICE 2015	497	19	26	391	163	116	56	545
WHO 2013	749	17	24	385	158	112	51	555

Table x15: Deterministic analysis for the Atlantic DiP population with NICE risk factors (n=1,988)

Diagnostic threshold	Cost	QALY	Incremental cost	Incremental QALY	ICER
No Treatment	£0	0.00	n/a	n/a	n/a
NICE 2015	£414,714	17.46	£414,714	17.46	£23,755
WHO 2013	£626,417	22.44	£211,703	4.98	£42,457

Table x16: Probabilistic sensitivity analysis for Atlantic in a population with NICE risk factors

Diagnostic threshold	NMB CE threshold £30,000 per QALY	Probability cost-effective CE threshold £30,000 per QALY
No Treatment	£518	30.2%
NICE 2015	£116,178	50.8%
WHO 2013	£53,129	19.1%

Figure x3: Cost-effectiveness acceptability curve indicating the probability of a threshold or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for the Atlantic DiP centres population with risk factors

Results for the Atlantic DiP population without risk factors

Table x17: Clinical outcomes for Atlantic DiP population without NICE risk factors (n=3,302)

Diagnostic threshold	Diagnosed	SD	SPC	CS	NICU	Jaund	PE	IOL
No Treatment	0	33	45	575	254	168	84	828
NICE 2015	194	31	42	569	248	166	79	837
WHO 2013	371	30	41	564	245	163	76	844

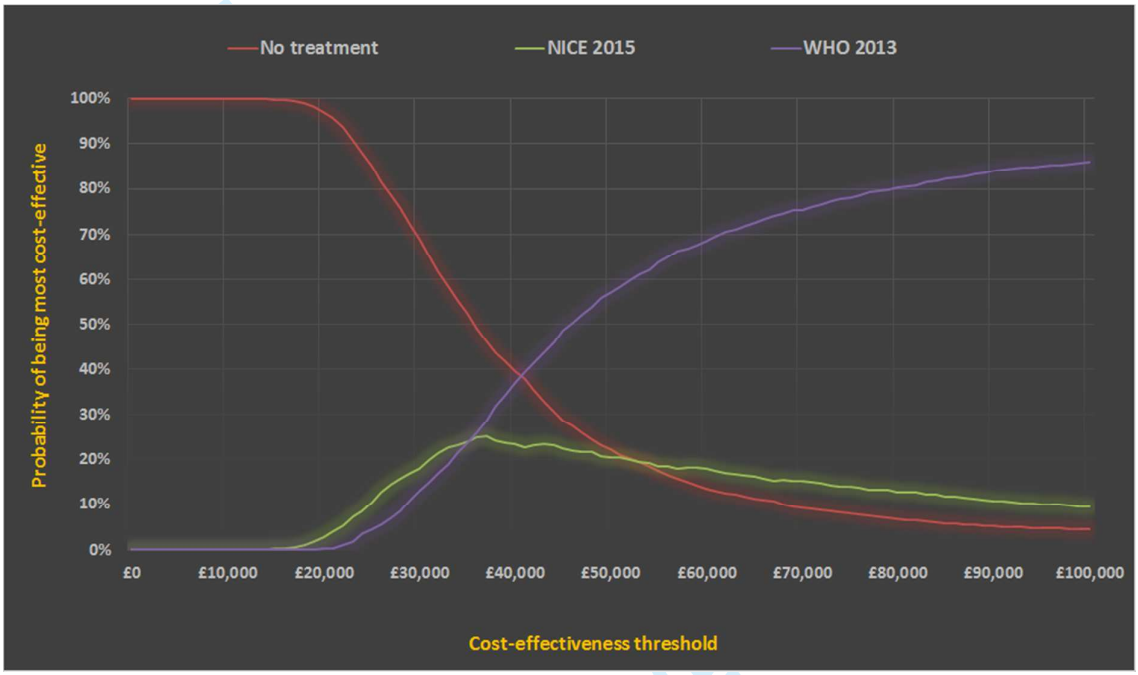
Table x18: Deterministic analysis for the Atlantic DiP population without NICE risk factors (n=3,302)

Diagnostic threshold	Cost	QALY	Incremental cost	Incremental QALY	ICER
No Treatment	£0	0.00	n/a	n/a	n/a
NICE 2015	£231,633	6.48	£231,633	6.48	£35,732
WHO 2013	£381,795	9.81	£150,162	3.33	£45,075

Table x19: Probabilistic sensitivity analysis for the Atlantic DiP population without NICE risk factors

Diagnostic threshold	NMB CE threshold £30,000 per QALY	Probability cost-effective CE threshold £30,000 per QALY
No Treatment	£402	69.0%
NICE 2015	-£36,481	18.0%
WHO 2013	-£88,283	13.1%

Figure x4: Cost-effectiveness acceptability curve indicating the probability of a threshold or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for the Atlantic DiP centres population without risk factors



Results for the Norwich population

Table x20: Clinical outcomes for Norwich population (n=12,754)

Diagnostic threshold	Diagnosed	SD	SPC	CS	NICU	Jaund	PE	IOL
No Treatment	0	132	182	2,333	1,005	699	346	3,173
NICE 2015	888	122	168	2,305	981	687	318	3,214
WHO 2013	1,771	117	161	2,283	965	676	301	3,248

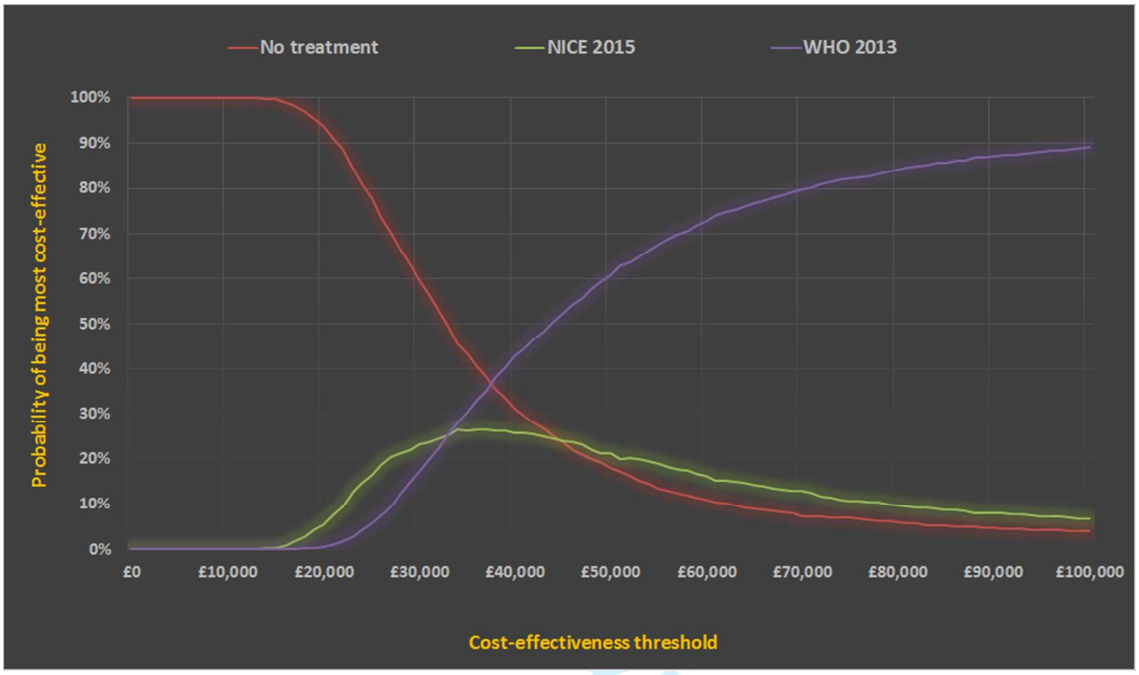
Table x21: Deterministic analysis for the Norwich population (n=12,754)

Diagnostic threshold	Cost	QALY	Incremental cost	Incremental QALY	ICER
No Treatment	£0	0.00	n/a	n/a	n/a
NICE 2015	£979,903	29.54	£979,903	29.54	£33,177
WHO 2013	£1,725,098	46.89	£745,195	17.35	£42,931

Table x22: Probabilistic sensitivity analysis for the Norwich population

Diagnostic threshold	NMB CE threshold £30,000 per QALY	Probability cost-effective CE threshold £30,000 per QALY
No Treatment	£1,141	59.7%
NICE 2015	-£76,289	23.2%
WHO 2013	-£300,524	17.2%

Figure x5: Cost-effectiveness acceptability curve indicating the probability of a threshold or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for the Norwich population



One-way sensitivity analysis

The cost-effectiveness of universal screening was not generally affected when the model was re-run using the regression models without backward elimination of non-significant variables with no screening/no treatment continuing to be the cost-effective option in populations not selected on the basis of NICE risk factors (see Table x23). In the Norwich population, universal screening was borderline cost-effective compared to no screening/no treatment at £30,000 per QALY but the same point remains that a risk factor subset in this population would have a lower ICER than that reported, and that a subset without risk factors, (i.e. those additionally incorporated as a result of universal screening compared to risk factor screening), would have a higher ICER. In populations with NICE risk factors the NICE 2015 diagnostic thresholds were still found to be cost-effective at a threshold of £30,000 per QALY, with broadly similar ICERs as previously. Similarly, the WHO 2013 diagnostic threshold was never found to be cost effective even in a population with risk factors.

Table x23: Summary of deterministic ICERs for each population without backward elimination of non-significant coefficients

	All covariates		Plasma glucose covariates				
Diagnostic threshold	**HAPO Risk factor (n=3,549)**	**HAPO No Risk factor (n=2,614)**	**HAPO Risk factor (n=3,549)**	**HAPO No Risk factor (n=2,614)**	**Atlantic DiP Risk factor (n=1,988)**	**Atlantic DiP No Risk factor (n=3,302)**	**Norwich (n=12,754)**
No Treatment	-	-	-	-	-	-	-
NICE 2015	£22,786	£46,677	£24,802	£39,338	£22,126	£37,887	£31,191
WHO 2013	£33,876	£107,247	£35,852	£54,288	£41,652	£43,106	£43,694

Mean plasma glucose values according to risk factor status

Table x241: Mean plasma glucose values in HAPO (4) and Atlantic DiP population according to their risk factor status

	HAPO 4			Atlantic DiP		
	Fasting	1-hour	2-hour	Fasting	1-hour	2-hour
True Positives	5.24	9.90	7.89	5.21	10.21	7.61
False Positives	4.50	7.20	5.95	4.33	6.75	5.33
True Negatives	4.44	6.95	5.78	3.92	5.99	4.76
False Negatives	4.89	9.52	7.41	4.90	9.51	7.12

CHEERS Statement

CHEERS checklist—Items to include when reporting economic evaluations of health interventions

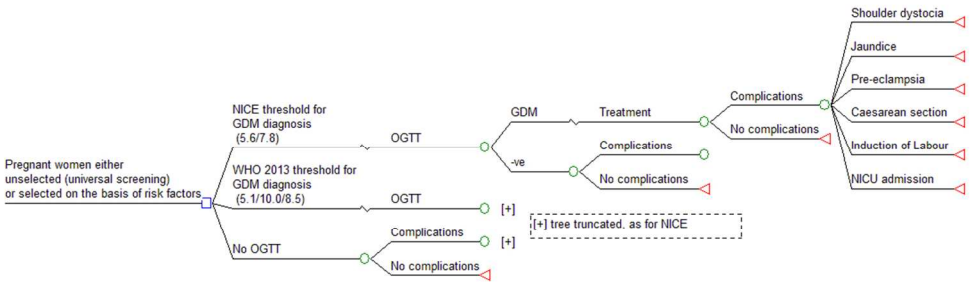
Section/item	Item No	Recommendation	Reported on page No/ line No
Title and abstract			
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	Yes Page 1 Line 2
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.	Yes Page 2 Lines 29-47
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study.	Yes Page 3-4 Lines 70-80
		Present the study question and its relevance for health policy or practice decisions.	Yes Page 3 Lines 53-68
Methods			
Target population and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	Yes Page 5 Lines 94-122
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	Yes Page 5 Lines 94-99
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	Yes Page 9 Line 189
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	Yes Page 4 Line 77-78; 84-85
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	Yes Page 10 Line 200 Supp. Report Page 9 Line 74

Section/item	Item No	Recommendation	Reported on page No/ line No
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	Yes Page 10 Line 200 Supp. Report Page 9 Line 74
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	Yes Pages 6-7 Lines 124-145
Measurement of effectiveness	11a	*Single study-based estimates:* Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	Yes Pages 8 Lines 174-177
	11b	*Synthesis-based estimates:* Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	Yes Pages 8 Lines 171-174
Measurement and valuation of preference based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	
Estimating resources and costs	13a	*Single study-based economic evaluation:* Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	
	13b	*Model-based economic evaluation:* Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	Yes Page 10-11 Lines 188-213 Supp. Report Page 8 Line 65
Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	Yes Page 10 Lines 199

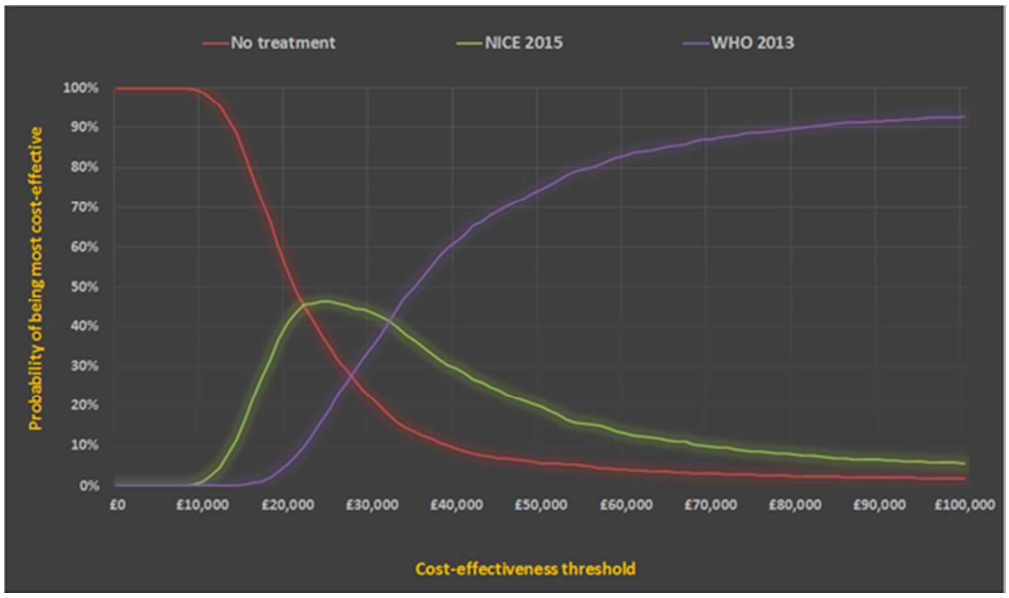
Section/item	Item No	Recommendation	Reported on page No/ line No
Choice of model	15	Describe and give reasons for the specific type of decision-analytical model used. Providing a figure to show model structure is strongly recommended.	Yes Page 4 Lines 84-85; 89-91
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	Yes Page 6-7 Lines 124-145 Yes Page 11 Lines 214-218 Supp. Report Page 2-7 Supp. Report Page 9 Lines 68-86 +References to other sources
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	Yes Page 7 Lines 147-164 Yes Page 11 Lines 225-232 Supp. Report Page 2-7
Results			
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	Yes Page 9 Lines 185-186 Page 11 Lines 212-213

Section/item	Item No	Recommendation	Reported on page No/ line No
			Supp. Report Page 2-9
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	Yes Page 13 Lines 246-248 Supp. Report Page 11-19
Characterising uncertainty	20a	*Single study-based economic evaluation:* Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	
	20b	*Model-based economic evaluation:* Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	Yes Supp. Report Page 11 Lines 95-103 Supp. Report Page 20 Lines 181-182
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	Yes Supp. Report Page 11 Lines 95-103
Discussion			
Study findings, limitations, generalisability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	Yes Pages 16-22
Other			
Source of funding	23	Describe how the study was funded and the role	Yes

Section/item	Item No	Recommendation	Reported on page No/ line No
		of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	Page 23 Lines 476-489
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	Yes Page 23 Lines 472-474



Model Schematic



Cost-effectiveness acceptability curve indicating the probability of a threshold or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for the HAPO 4 centres population with risk factors

BMJ Open

A COST EFFECTIVENESS COMPARISON OF THE NICE 2015 AND WHO 2013 DIAGNOSTIC CRITERIA FOR WOMEN WITH GESTATIONAL DIABETES WITH AND WITHOUT RISK FACTORS

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-016621.R1
Article Type:	Research
Date Submitted by the Author:	28-Apr-2017
Complete List of Authors:	Jacklin, Paul; Royal College of Obstetricians and Gynaecologists, National Guideline Alliance Maresh, Michael; Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre Patterson, Chris; Queen's University Belfast, Centre for Public Health Stanley, Katharine; Norfolk and Norwich University Hospitals NHS Foundation Trust, Department of Obstetrics and Gynaecology Dornhorst, Anne; Hammersmith Hospital, Department of Investigative Medicine Burman-ROY, Shona; Royal College of Obstetricians and Gynaecologists, National Guideline Alliance Bilous, Rudy; Newcastle University Medicine Malaysia
Primary Subject Heading:	Health economics
Secondary Subject Heading:	Diabetes and endocrinology, Obstetrics and gynaecology, Diagnostics
Keywords:	HEALTH ECONOMICS, DIABETES & ENDOCRINOLOGY, OBSTETRICS

Title:

A COST EFFECTIVENESS COMPARISON OF THE NICE 2015 AND WHO 2013 DIAGNOSTIC CRITERIA FOR WOMEN WITH GESTATIONAL DIABETES WITH AND WITHOUT RISK FACTORS

Authors: PB Jacklin[1], MJA Maresh[2], CC Patterson[3], KP Stanley[4], A Dornhorst[5], S Burman-Roy[1], RW Bilous[6]

Institutions:

1. Royal College of Obstetricians and Gynaecologists, London NW1 4RG, UK
2. St. Mary's Hospital, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester M13 9WL, UK
3. Centre for Public Health, Queen's University Belfast, Room 3.014, ICS Block B, Grosvenor Road, Belfast BT12 6BJ, UK
4. Department of Obstetrics and Gynaecology, Norfolk and Norwich University Hospitals NHS Foundation Trust, Colney Ln, Norwich NR4 7UY, UK
5. Faculty of Medicine, Department of Investigative Medicine, Hammersmith Hospital, Imperial College London, London, UK
6. Newcastle University Medicine Malaysia, Johor, Malaysia

Corresponding author:

Paul Jacklin
e-mail: pjacklin@rcog.org.uk

Abstract: 278 words

Main text: 4618 words

Abstract

Objectives To compare the cost effectiveness of The National Institute for Health and Care Excellence (NICE) 2015 and the World Health Organisation (WHO) 2013 diagnostic thresholds for gestational diabetes (GDM).

Setting: The analysis was from the perspective of the National Health Service (NHS) in England and Wales.

Participants: 6,221 patients from four of the Hyperglycaemia and Adverse Pregnancy Outcomes (HAPO) study centres (2 UK, 2 Australian), 6,308 patients from the Atlantic Diabetes in Pregnancy (DiP) study and 12,755 patients from UK clinical practice

Primary and secondary outcome measures planned: The incremental cost per quality adjusted life year (QALY), net monetary benefit (NMB) and the probability of being cost-effective at cost-effectiveness thresholds of £20,000 and £30,000 per QALY

Results. In a population of pregnant women from the four HAPO study centres, and utilising NICE defined risk factors for GDM, diagnosing GDM using NICE 2015 criteria had an incremental cost effectiveness ratio (ICER) of £20,400 per QALY gained (relative to no treatment) compared to £33,596 per QALY gained (relative to NICE 2015 criteria) using WHO 2013 diagnostic criteria. At a cost-effectiveness threshold of £30,000 per QALY the NICE 2015 criteria had a 53.5% probability of being cost-effective compared to the WHO 2013 diagnostic criteria which had a 26.8% probability of being cost-effective (no treatment had a 19.8% probability of being cost-effective). The ICERs for women without NICE risk factors in this population were £36,878 and £141,812 per QALY for NICE and WHO diagnostic criteria, respectively.

Conclusion The NICE 2015 diagnostic criteria for GDM can be considered cost-effective relative to the WHO 2013 alternative at a cost-effectiveness (CE) threshold of £30,000 per

QALY. Universal screening for GDM was not found to be cost-effective relative to screening based on NICE risk factors.

Keywords: Cost Effectiveness, Gestational Diabetes, Screening, Risk Factors, Diagnosis

Strengths and limitations of this study

- This economic evaluation addresses an important clinical and policy issue. The existing economic evidence is limited and WHO have stated that studies of this type are needed to inform a future update of their guideline
- Our paper has used patient-level data from the influential HAPO study for an economic analysis which has not been previously been published in a peer reviewed journal
- This analysis provides clear evidence that universal screening is not cost-effective in the UK
- This analysis suggests that the NICE diagnostic criteria for GDM are more cost-effective than the WHO criteria in the UK context
- Model conclusions are sensitive to uncertainties with respect to valuation of health outcomes and the possible long term metabolic consequences for offspring for which the evidence is debated and which are hard to quantify

Introduction

The diagnostic glycaemic thresholds for GDM remain the subject of considerable debate. The original definition was based upon maternal risk for developing postpartum diabetes, but subsequent thresholds have concentrated on complications during pregnancy and the health of the offspring. The publication of the HAPO study[1] demonstrated a linear association between increasing levels of maternal hyperglycaemia and adverse perinatal outcomes with no obvious threshold, an association that has also been observed in subsequent analyses.[2] The discussion around the diagnostic criteria that should define GDM has intensified. New diagnostic thresholds were proposed by the International Association of Diabetes in Pregnancy Study Group (IADPSG)[3] based upon the HAPO study levels of plasma glucose when fasting, and at 1 and 2 hours after an oral 75g glucose load that were associated with covariate adjusted odds ratio of 1.75 relative to the mean glucose value in the whole HAPO cohort on three offspring outcomes: exceeding the 90th centile for birth weight, for cord serum C-peptide concentration and for percent fetal body fat. These diagnostic criteria have been subsequently adopted by the WHO.[4] However, they remain controversial and have not been supported by bodies such as the National Institutes for Health and the American College of Obstetricians.[5] Furthermore, WHO has acknowledged that they will have to be revisited in the near future in the light of new studies reporting their cost-effectiveness.[4]

In 2015 NICE published updated guidance on Diabetes in Pregnancy[6] which included recommendations on diagnostic thresholds for GDM which differ from those adopted by WHO. These NICE thresholds were informed by an economic evaluation of the type that WHO considered important to inform future recommendations, but have attracted criticism in the UK[7] and elsewhere. Data from a published Spanish study[8] have been widely cited[7, 9] in support of the

cost effectiveness of the WHO criteria, although a UK analysis has more recently suggested that it is not cost-effective to identify gestational diabetes for treatment.[10]

In this paper we compared the cost-effectiveness of NICE 2015 and WHO 2013 diagnostic thresholds for GDM, as these are new thresholds proposed by national and international bodies. The analysis was undertaken using a revised version of the health economic model developed for the NICE guideline and was based upon data from the UK and Australian HAPO Study centres.

Methods

Model description

A decision analytic framework was used to evaluate the cost effectiveness of two recently proposed diagnostic thresholds for GDM, together with a no diagnosis/no treatment option (See Table 1). A schematic of the model is shown in Figure 1. Cost-effectiveness was evaluated using both deterministic and probabilistic sensitivity analysis.

Table 1: Diagnostic thresholds for plasma glucose evaluated in the economic model

Threshold name	Fasting (mmol/L)	1-hour (mmol/L)	2-hour (mmol/L)
No diagnosis/no treatment	-	-	-
NICE 2015	≥5.6	-	≥7.8
WHO 2013	≥5.1	≥10.0	≥8.5

Population

The model population comprised women of gestational age 24-28 weeks without pre-existing diabetes. The analysis utilised individual patient data from three datasets which, although not

restricted to the UK, provide a representative cross section of the demographic and patient characteristics that would be found in the UK (Table x1 in the Supplementary Report provides a breakdown of ethnic groups in each of our datasets) . The analyses were run separately for each dataset and, where possible, for subgroups with and without risk factors for GDM within a dataset.

i. HAPO – a dataset from the two UK (Manchester and Belfast) and two Australian (Brisbane and Newcastle) centres of the HAPO Study, referred to as HAPO (4)

ii. Norwich – these data were routinely collected between 2008 and February 2014 on women who had an oral glucose tolerance test (OGTT) on the basis of the presence of one or more risk factors for GDM. The results were obtained from laboratory records with no identifiers. Risk factors in addition to those recommended by NICE were used e.g. women with polycystic ovary syndrome, previous stillbirth or recurrent glycosuria.

iii. Atlantic Diabetes in Pregnancy (Atlantic DiP) – these data were collected between 2007 and 2013 as part of a research initiative in the Republic of Ireland intended to improve pregnancy outcomes for women with diabetes before, during and after pregnancy.

For the HAPO (4) and Atlantic DiP datasets the populations were stratified according to whether or not they had NICE risk factors for GDM (body mass index (BMI) above 30 kg/m^2, previous baby with birthweight ≥4.5 kg, previous GDM, first-degree relative with diabetes and minority ethnic family origin with a high prevalence of diabetes). This facilitated a comparison of the cost-effectiveness of universal screening for GDM when compared with a risk factor approach.

The NICE risk factor approach could not be replicated exactly because the patient data used in the model do not include information on previous offspring birth weight, and the HAPO (4) dataset does not provide information on previous GDM. Similarly the Atlantic DIP dataset does not include data on previous macrosomia or previous GDM. Therefore, the comparison in the model was between universal screening and a subset of NICE risk factors. Our Norwich dataset only included the plasma glucose values from a three point (fasting, 1 and 2 hour) OGTT and therefore it was not possible to assess cost-effectiveness according to the presence of risk factors in this group.

Clinical outcomes

The agreed outcomes for the economic model were selected prior to model development by the NICE Guideline Development Group. They were:

i. Shoulder dystocia (SD) – this was used to estimate serious perinatal complications (SPC), a broader composite outcome (death, shoulder dystocia and birth trauma) used as a primary outcome in clinical trials. The estimation of SPC from shoulder dystocia has been described elsewhere.[6]
ii. Caesarean section (CS)
iii. Neonatal intensive care unit (NICU) admission
iv. Jaundice requiring phototherapy (Jaund)
v. Pre-eclampsia (PE)
vi. Induction of labour (IOL)

Outcomes were prioritised for inclusion in the model if they had a direct impact on health related quality of life and/or cost. Birth weight was not included because there were few long-term outcome data for modelling any risk benefit of a reduction in birth weight for future diabetes and other health outcomes in the offspring.

In addition, outcomes were only included if the relationship with plasma glucose levels had been established in the HAPO study, and also that they had been assessed in intervention studies used to derive treatment effect size estimates. Possible double counting of certain outcomes was taken into account (e.g. preterm birth and NICU admission). The final list of outcomes included in the model was therefore a pragmatic one.

Baseline risk

Logistic regression analyses of patient data from HAPO (4) were used to predict a baseline risk for all six outcomes for each woman, based on their characteristics including their OGTT results. In the HAPO study the OGTT was blinded to the carers, unless there was overt diabetes, thus allowing direct comparison of the OGTT with perinatal outcomes without intermediate treatment effects for those meeting the new diagnostic criteria for GDM.

For each of the six outcomes, 2 logistic analyses to predict risk were assessed:

i. Prediction based on OGTT plasma glucose results and including the same covariates as used for Model 2 in the original analysis of the HAPO data[1] – this could not be applied to the Norwich and Atlantic DiP datasets as information on all HAPO covariates was not available

ii. Prediction based only on OGTT plasma glucose results

Backward elimination of plasma glucose variables with non-significant coefficients was undertaken to arrive at a 'final' logistic regression analysis to predict baseline risk for each outcome for the base case analysis, although a sensitivity analysis is also presented where the model was run with plasma glucose variables with non-significant coefficients retained. The logistic regression analyses used to predict the baseline risk for each outcome are shown in the Supplementary Report, Tables x2 to x7.

Clinical effectiveness

For each evaluated diagnostic threshold inTable 1 the model determined whether a woman would be identified as having GDM based on her OGTT. If the woman was not identified as having GDM then outcome probabilities were based on the predicted baseline risk, but for women identified as having GDM the predicted baseline risk was modified to take account of the effects of treatment. Treatment effectiveness for most outcomes was estimated from a random-effects meta-analysis of two studies, the Australian Carbohydrate Intolerance Study (ACHOIS) and the Landon et al. trial.[11, 12] Other published studies of treatment for GDM were adjudged to lack adequate randomisation.[13] For the NICU outcome only the Landon et al. trial data were used as it was considered to more closely represent UK practice as all neonatal nursery admissions were utilised. Similarly, the incidence of pre-eclampsia seemed high in ACHOIS in both arms, and again only Landon et al. trial data were utilised. The treatment effects for each of the model's clinical outcomes are shown in Table 2 along with parameters for probabilistic sampling. The model assumes that the relative treatment effect will be the same irrespective of the absolute baseline risk. For deterministic analyses the point estimate of relative risk was used but in order to account for uncertainty in these point estimates, these relative risks were sampled from a log-normal distribution in the simulations undertaken for probabilistic sensitivity analysis (PSA).

Table 2: Relative treatment effects for model outcomes

Outcome	Relative risk (RR)	Standard error (log RR)	Source

Shoulder dystocia	0.41	0.316	ACHOIS (2005), Landon (2009)
Caesarean section	0.88	0.095	ACHOIS (2005), Landon (2009)
NICU	0.77	0.194	Landon (2009)
Jaundice requiring phototherapy	0.83	0.136	ACHOIS (2005), Landon (2009)
Pre-eclampsia	0.46	0.345	Landon (2009)
Induction of Labour	1.16	0.126	ACHOIS (2005), Landon (2009)

214

215

Costs

Costing was undertaken from the perspective of the NHS, was calculated for each woman in the dataset being analysed and was made up of three components;

- the costs of the diagnostic test – not applied in the *no test/no treat* strategy
- the costs of treatment- applied to every woman diagnosed with GDM at a particular threshold
- the costs associated with the various outcomes – with the cost for each woman being the expected (or average) cost of the outcome based on her estimated risk

The costs calculated for each woman were then summed across the entire patient dataset to give a total cost for a particular diagnostic threshold.

Costs are presented in pounds sterling and were taken from published UK sources where possible (cost year 2015). They have not been discounted as they are all assumed to occur within 12 months of diagnosis. Model unit costs are reported in the Supplementary Report, Table x14. The costing methodology and assumptions are described in greater detail elsewhere.[6]

Other event probabilities

Probabilities in decision analysis were used to calculate the expected costs and benefits of the various comparators. Many of these probabilities stemmed from relative treatment effects but a few additional event probabilities were included in the model in order to estimate certain costs. These probabilities are shown in Table 3 and their source is described elsewhere.[6]

Table 3: *Model event probability not derived from patient level regression*

Event	Probability
Not requiring hypoglycaemic therapy when treated for GDM	36%
Risk of hypoglycaemia if taking hypoglycaemic therapy	20%
Risk of hypoglycaemia being severe (requiring hospitalisation)	5%

Quality Adjusted Life Years (QALYs)

Following previous studies[6, 14] a QALY decrement of 2.2 was assigned to serious perinatal complications (SPC), defined as per the ACHOIS study as a composite outcome of shoulder dystocia, death and birth trauma.[11] More detail on the derivation of this QALY loss is provided in the Supplementary Report. The cost-effectiveness of a healthcare intervention is determined by the opportunity cost of the health foregone on the basis that with a fixed health budget any newly funded intervention would displace the least cost-effective treatment currently provided. In the UK, NICE typically uses a threshold of £20,000 to £30,000 per QALY as a benchmark[15] for the opportunity cost of health foregone and this paper assesses cost-effectiveness accordingly.

Sensitivity analysis

Probabilistic sensitivity analysis, using Monte Carlo simulation (with 2,000 iterations for each analysis), was undertaken in order to assess the impact of sampling uncertainty on model inputs. Parameters and distributions for the probabilistic sensitivity analysis are given in Table 2 and Table x14 in the supplementary report. For the logistic regression coefficients used to predict baseline risk, the Cholesky decomposition method[16] was used to sample from a multivariate normal distribution in order to reflect correlations between the coefficients. The Cholesky

decomposition of the variance covariance matrices from the regression analyses used in the base case probabilistic sensitivity analysis are given in Table x8 to x13 in the Supplementary Report.

Results

Table 4 shows the percentage of women diagnosed with GDM in the three populations using both of the evaluated diagnostic thresholds. In addition, for the HAPO (4) and Atlantic DiP datasets this is additionally broken down in the subgroups with and without NICE risk factors (RF).

Table 4: Percentage of women identified with GDM by threshold and population

Threshold name	Norwich (n=12,754)	HAPO all (n=6,163)	HAPO RF (n=3,549)	HAPO No RF (n=2,614)	DiP All (n=5,290)	DiP RF (n=1,988)	DiP No RF (n=3,302)
NICE 2015	7.0%	13.6%	17.7%	8.0%	13.1%	25.0%	5.9%
WHO 2013	13.9%	18.9%	25.7%	9.7%	21.2%	37.7%	11.2%

Detailed deterministic and probabilistic results for HAPO (4) with risk factors are shown in Table 5, Table 6, Table 7 and Figure 2.

Table 5: *Clinical outcomes for HAPO (4) population with NICE risk factors (n=3,549)*

Diagnostic threshold	Diagnosed	SD	SPC	CS	NICU	Jaund	PE	IOL
No Treatment	0	49	67	759	345	219	146	974
NICE 2015	629	41	56	739	326	210	123	1,004
WHO 2013	912	39	54	731	321	207	117	1,016

Table 6: *Deterministic analysis for the HAPO (4 centres) population with NICE risk factors (n=3,549)*

Diagnostic threshold	Cost[a]	QALY[a]	Incremental cost	Incremental QALY	ICER
No Treatment	£0	0.00	n/a	n/a	n/a
NICE 2015	£546,349	26.78	£546,349	26.78	£23,073
WHO 2013	£778,993	34.35	£254,376	7.57	£37,669

a) Costs and QALYs are measured relative to a baseline of No Treatment

Table 5 indicates that there was a relatively small difference in clinical outcomes contrasting NICE and WHO diagnostic criteria, despite there being a 45% increase in women diagnosed with GDM. Using the WHO 2013 criteria, instead of the NICE 2015 criteria, an additional 142 women would have been diagnosed with GDM, and treated in order to prevent 1 case of shoulder dystocia.

In the deterministic analysis the NICE 2015 diagnostic criteria would be considered cost-effective at a cost-effectiveness threshold of £30,000 per QALY (Table 6).

The probabilistic sensitivity analysis reached a similar conclusion, with the NICE 2015 diagnostic threshold having the highest probability of being the most cost-effective treatment and the highest NMB using a cost-effectiveness threshold of £30,000 per QALY (Table 7 and Figure 2). The analysis also suggested that no diagnosis/no treatment might be considered the most likely to be cost-effective when using a lower cost-effectiveness threshold of £20,000 per QALY. The probability of no diagnosis/no treatment being cost-effective falls sharply in the cost-effectiveness threshold range of £20,000 - £30,000 *per* QALY. As shown in the cost-effectiveness acceptability curve in Figure 2, the WHO 2013 diagnostic threshold becomes

more cost-effective as the cost-effectiveness threshold increases. Nevertheless, this would have to exceed £30,000 per QALY before becoming cost-effective, indicating that the further reduction in adverse outcomes, are achieved at an unacceptably high opportunity cost. The Supplementary Report plots the incremental cost and QALY outcomes of 2,000 simulations from the probabilistic analysis on the cost-effectiveness plane (see Figure x1). Whilst most points fall in the south-western quadrant, suggesting that WHO 2013 diagnostic criteria are likely to lead to additional QALYs when compared with NICE 2015 criteria, all points show that NICE 2015 criteria were associated with markedly lower costs.

Table 7: *Probabilistic sensitivity analysis for HAPO (4) in a population with NICE risk factors*

Diagnostic threshold	**NMB[a] CE threshold £30,000 per QALY**	**Probability cost-effective CE threshold £20,000 per QALY**	**Probability cost-effective CE threshold WTP = £30,000 per QALY**
No Treatment	£486	55.5%	19.8%
NICE 2015	£230,798	42.1%	53.5%
WHO 2013	£178,231	2.4%	26.8%

a) NMB is measured relative to the least costly and least effective strategy in each simulation

Summaries of results for all of the model populations and more detailed results are provided in the Supplementary Report.

Tables x16 and x17 in the Supplementary Report show that in both the HAPO (4) and Atlantic DiP populations with NICE risk factors, the NICE diagnostic threshold is the most cost-effective strategy at a cost-effectiveness threshold of £30,000 per QALY. The NICE 2015

diagnostic threshold has ICERs of less than £30,000 per QALY, and in the probabilistic sensitivity analysis it has the highest net monetary benefit and the highest probability of being the most cost-effective. For HAPO (4) the results are similar if baseline risks are estimated using logistic regression based on all covariates or a logistic regression just using plasma glucose levels.

The results also suggested that universal screening would not be cost-effective as, when compared to risk factor screening (as recommended in NICE guidelines), the additional women included in such an approach would be those without risk factors and the model demonstrates that the ICERs for diagnosis and treatment are all well in excess of £30,000 per QALY; markedly so when using WHO 2013 diagnostic thresholds. These conclusions were supported by an analysis of the Norwich dataset (see Supplementary Report).

It was not possible to stratify the Norwich dataset according to risk factors, and therefore the ICERs presented relate to a comparison between no screening/treatment and universal screening and treatment. However, the results were consistent with those for HAPO (4) and Atlantic DiP. First, they showed that universal screening was not cost-effective even when compared to an alternative of no screening/no treatment. Second, the ICERs for the whole population were a weighted average of the populations with and without risk factors. The ICER for the population without risk factors would be higher than the ICER for the entire population, which was only marginally below the £30,000 per QALY threshold.

Deterministic sensitivity analysis

As part of a sensitivity analysis the deterministic models were re-run using the logistic regression models without backward elimination of glucose variables with non-significant coefficients, and these analyses are summarised in the Supplementary Report.

Discussion

In the NICE guideline analysis, 14 alternative diagnostic thresholds were compared and there was no single optimal diagnostic threshold which clearly emerged[6]. This is not surprising given the small differences in patient outcomes between them. In that analysis the previous WHO 1999 criteria emerged as a relatively cost-effective strategy. However, the Guideline Development Group rejected a fasting threshold of 7.0 mmol/L as there was a wide clinical consensus that this was too high, as 6.1-7.0 mmol/L is diagnostic of impaired fasting glycaemia in the non-pregnant population. Intervention studies had used a lower fasting threshold than 7.0 mmol/L as a basis for inclusion, and therefore made a case for intervention at lower levels. Based upon detailed cost effectiveness analysis of all the options, the Guideline Development Group ultimately decided on recommending a fasting plasma glucose of 5.6 mmol/L and a 2 hour plasma glucose of 7.8 mmol/L. In this paper, we have restricted our analysis of cost-effectiveness to the WHO 2013 and NICE 2015 criteria (with a no screening/treatment baseline also included) as these two recommendations have the most clinical currency at present.

All of the analyses presented in this paper suggest that, in a population with NICE risk factors, the NICE 2015 diagnostic criteria for GDM could be considered cost-effective relative to no screening/no treatment and to WHO 2013 diagnostic thresholds when using a cost-effectiveness threshold of £30,000 per QALY. The analyses also show that no screening/no treatment is cost-effective in populations without NICE risk factors, suggesting that universal screening does not represent value for money, at least in a UK setting.

One of the limitations of our analysis was that the 2-hour threshold was restricted to the historical WHO 1999 2-hour definition of 7.8mmol/l, or the new WHO 2013 criteria of 8.5 mmol/l. It is conceivable that a 2-hour threshold lying between these values might outperform both. Our greater focus, though was on the optimal fasting level as this is where the greatest controversy lies with respect to potentially missed treatment opportunities.

As noted by the proponents of WHO 2013 diagnostic criteria for GDM, using a lower fasting plasma glucose threshold would by definition detect more cases. Furthermore, because we assumed in the model that the relative treatment effect would be the same in additionally diagnosed cases, it follows that such a threshold could potentially yield the lowest number of adverse outcomes and the greatest QALY gain. However, our analysis suggests that the relatively small additional gains are not justified by the substantially higher costs that such lower thresholds would require.

A key driver of our results were the logistic regression models which were used to predict baseline risk. For the outcomes included in this study these regression models suggested that the 2-hour plasma glucose was a much more important predictor of adverse outcomes than the fasting plasma glucose, something we were unaware of when selecting the model's clinical outcomes. For the regression models fitted to predict baseline risk in the HAPO (4) dataset with covariates and backward elimination of the OGTT plasma glucose variables (Model 1 base case analysis regressions in Supplementary Tables x2 to x7), the Hosmer-Lemeshow Goodness of Fit Test did not indicate evidence of poor fit ($p > 0.05$). However, there was evidence of poor fit ($p < 0.05$) for the regression models of caesarean section and NICU admission where the prediction was based only on OGTT plasma glucose results (Model 2 base case analysis

regressions in Supplementary Tables x2 to x7). Nevertheless, as indicated in Supplementary Table x16 and x17, the choice of prediction model did not have a large bearing on cost-effectiveness.

We consider that our analysis which builds on previous modelling[6, 14] is one of the most comprehensive assessments of the cost-effectiveness of diagnostic thresholds for GDM yet undertaken, and will hopefully contribute to the WHO's expectation "that a substantial body of new data will emerge in the near future, providing currently scarce health and economic evaluation of the recommended criteria applied to various populations and with different approaches (universal screening, screening only women at high risk, diagnostic testing only)".[4]

A number of commentators [17, 18] have recently advocated universal screening for GDM. The essence of the argument is based upon the number of cases of GDM that would be missed with selective screening, and the subsequent reduced opportunity to prevent a serious perinatal outcome. Of course it is true that universal screening will detect more cases, although the absolute numbers will depend upon the thresholds used to define GDM. Table 5 shows that many more women would need to be diagnosed in order to prevent a single adverse outcome. However, in the context of finite health care resources, it must be accepted that it may be cost-effective to miss some cases. Epidemiological measures such as number needed to treat (or number needed to screen in this case) implicitly recognise that a goal of health care systems cannot be to maximize health gain without any consideration of cost. Identifying missed cases carries an opportunity cost and it may be that those resources would achieve greater benefit if employed elsewhere in the health care system. If a population is divided into those with risk factors and those without risk factors, then the prevalence of GDM must be lower in the group without risk factors (and the number needed to screen higher) with concomitantly lower cost-

effectiveness. However, the comparative cost-effectiveness of screening in those with and without risk factors is not only affected by the respective prevalence in the two groups, but also differences in severity. In those diagnosed with GDM and who had risk factors there were, as anticipated, greater levels of hyperglycaemia than in those without risk factors. As shown in Table x31 in the Supplementary Report, 'true positives' or identified cases (risk factor present and GDM) had higher plasma glucose values than 'false negatives' or missed cases (risk factors absent and GDM) when defining GDM positives according to WHO 2013 diagnostic thresholds.

We would therefore expect the women with risk factors and GDM to be at greater risk of adverse outcomes than the women with GDM without risk factors as a result of their higher plasma glucose levels. So the "cases" missed with selective screening would have, on average, fewer adverse outcomes than in "cases" in a population with risk factors. So the ICER would be greater in the population without risk factors because prevalence is lower and cases have fewer adverse outcomes.

Our analysis, by splitting the HAPO (4) and Atlantic DiP datasets into those with and without risk factors, was able to evaluate the cost-effectiveness of moving from risk factor screening to universal screening. Whilst diagnosis in populations with risk factors was shown to be cost-effective at a threshold of £30,000 per QALY, it was never cost-effective to diagnose and treat in those without risk factors. Table 4 indicates the large differences that exist in prevalence between the populations with and without risk factors. Our analysis suggests that the cost-effectiveness threshold would have to substantially exceed currently accepted UK norms for universal screening to be considered cost-effective. Although the NICE risk factor approach could not be replicated exactly, we felt that the approximation used was acceptable, as the only

women who would be omitted from the model risk factor population were multiparous and would have had a large baby previously and/or a past history of GDM. This approximation would over-estimate slightly the benefits of universal screening, as the baseline risk in a group designated as being without NICE risk factors present would be over-stated.

A previous study[8] from Spain using WHO 2013 diagnostic criteria suggested cost effectiveness compared with a two-step protocol using the Carpenter – Coustan thresholds. However, this was largely based upon estimates of reduction of caesarean section rates of 50% which we find implausible based upon changes in diagnostic criteria alone, noting that ACHOIS and Landon et al. found only a 4% and 21% reduction in caesarean section respectively as a result of treating gestational diabetes. The Spanish study did not consider other alternative thresholds, and was a retrospective, before and after analysis which has been criticised by the Cochrane Collaboration as it does not control for possible changes in important variables, such as clinical management, over time.[19]

A recently published UK Health Technology Assessment (HTA) [10] suggested that the identification of gestational diabetes for treatment is not cost-effective, in which case finding a cost-effective threshold becomes somewhat redundant. Although the HTA followed a similar approach to our analysis there were some differences which could explain the different conclusions. In our analysis, jaundice was included as an outcome and the relative treatment effect would have tended to lower the incremental costs of intervention as a result of reduced rates of phototherapy. This was not included as an outcome in the HTA. Instrumental delivery was included as an outcome in the HTA but not in our analysis. While instrumental delivery rates could in theory be increased by treatment, as there will be more vaginal births, this could be counteracted by those mothers not treated delivering larger babies vaginally requiring

assistance; this would be in accord with the HTA meta-analysis which failed to demonstrate a treatment effect on instrumental delivery rates. In addition the HTA reported smaller treatment effects for NICU admission and pre-eclampsia. Unlike our analysis, the HTA did not assume 100% uptake of the OGTT and that may also have led to a smaller estimate of treatment benefit. However, the differences should not be over-emphasised. Like the HTA our results would not support the identification and treatment of gestational diabetes if a cost-effectiveness threshold of £20,000 per QALY was used. However, it was the view of the Guideline Development Group that the clinical benefit of identifying and treating women with GDM is widely practiced, and that a no identification/no treat policy would not be acceptable to patients or health care providers. As such, the Group felt that the higher cost threshold of £30,000 was justified.

Our model has a number of limitations particularly with respect to the valuation of health outcomes. We did not include large for gestational age as an outcome because it was felt that shoulder dystocia was the relevant immediate complication of interest, and that possible long term metabolic consequences for the offspring were hard to quantify and therefore difficult to incorporate within the model. As previously noted, the QALY loss from a serious perinatal complication used in this analysis is likely to be overstated because of the relatively large weight given to death based on the intervention studies.[14] HAPO failed to show an association between perinatal mortality and plasma glucose levels, which may mean that perinatal mortality reduction is less amenable to reduction by treatment than other serious perinatal complications. In this respect the cost-effectiveness of diagnosing and treating GDM may be over-stated. On the other hand, the model does not take account of any potential long term effects on the offspring (e.g. adiposity and the likelihood of subsequent pathology) as these effects are difficult to quantify but may under-estimate the QALY gain from diagnosis and treatment. A US study[20] considered the potential long-term benefits to the mother whereby a diagnosis of

GDM averts or delays onset of Type 2 diabetes mellitus, but this was not incorporated into our model as we did not consider that the relationship was sufficiently well established at this time. However, to the extent that such a relationship does exist our model would also underestimate the QALY gain from a diagnosis of GDM. A recent review has, however, questioned the association between maternal glycaemia and subsequent cardio-metabolic outcomes in offspring in humans[21] and a recent follow-up study failed to find evidence of a reduction in childhood obesity or metabolic dysfunction at five years in the offspring of women treated for mild gestational diabetes in the study of Landon et al [12, 22].

Despite these caveats, we feel our analysis represents a robust analysis of the cost-effectiveness of the NICE versus the WHO 2013 diagnostic thresholds for GDM based upon our current understanding of the impact of intervention in women with GDM in the UK population. We acknowledge completely that this analysis cannot be the final word on the subject, and that further health economic evaluation is required to either corroborate our findings or to challenge them. Nevertheless, our analysis represents a constructive and evidence based contribution to establishing cost effective diagnostic thresholds for GDM and will hopefully lead to more research to clarify this important but vexed area of clinical diagnosis.

Conclusions

The results presented in this analysis, based on a UK setting, do not suggest that the diagnostic thresholds for GDM adopted by the WHO are cost-effective. On the other hand they do provide some support for the cost-effectiveness of the diagnostic criteria adopted by NICE when compared to either no screening/treatment and to WHO 2013 diagnostic criteria. Furthermore, according to this analysis, universal screening would seem to offer poor value for money and

does not appear cost-effective compared to the current NICE guidance of targeting high risk women.

Acknowledgements

We are grateful to Professor DR McCance and Professor HD McIntyre for allowing us to use their local datasets from the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) trial and to Professor F Dunne for allowing us to use her Atlantic DiP dataset.

We are also grateful to Professor David James who provided clinical support during the development of the updated NICE guideline on Diabetes in Pregnancy.

Competing interests

All authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/coi_disclosure.pdf and declare:

Funding: Some of this work was undertaken by the now defunct National Collaborating Centre for Women's and Children's Health (subsumed within the National Guideline Alliance from 1st April 2016), which received funding from NICE. The views expressed in this publication are those of the authors and not necessarily those of the institute. Revisions to the guideline model after the guideline was published and drafting of the manuscript was done in the author's own time and was not funded.

National Institute for Health and Care Excellence (2015). Diabetes in pregnancy: management from preconception to the postnatal period. Available from https://www.nice.org.uk/guidance/ng3

PBJ and SBR are employees of the National Guideline Alliance (part of the RCOG), which receives its funding from NICE.

MJAM, KS, AD and RWB received travel expenses from NICE for attending clinical guideline development meetings

Author contribution

Paul Jacklin designed and developed the health economic model, undertook the health economic analysis, wrote the first draft of the manuscript and incorporated edits from co-authors. Mike Maresh provided clinical input into the design of the health economic model; read, commented and edited various draft of the manuscripts. Katharine Stanley supplied the Norwich dataset, provided clinical input into the design of the health economic model; read, commented and edited various draft of the manuscripts. Anne Dornhorst provided clinical input into the design of the health economic model; read, commented and edited various draft of the manuscripts. Chris Patterson provided statistical advice, undertook statistical analysis of the HAPO dataset; read, commented and edited various drafts of the manuscript. Shona Burman-Roy reviewed the clinical literature, contributed to discussions of model design; read, commented and edited various drafts of the manuscript. Rudy Bilous chaired the NICE guideline, provided clinical input into the design of the health economic model; read, commented and edited various draft of the manuscripts.

Transparency declaration

The lead author, Paul Jacklin, affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Exclusive License

"I Paul Jacklin The Corresponding Author of this article contained within the original manuscript which includes any diagrams & photographs within and any related or stand alone film submitted (the Contribution") has the right to grant on behalf of all authors and does grant

on behalf of all authors, a licence to the BMJ Publishing Group Ltd and its licencees, to permit this Contribution (if accepted) to be published in the BMJ and any other BMJ Group products and to exploit all subsidiary rights, as set out in our licence set out at: http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/copyright-open-access-and-permission-reuse."

Data sharing Statement

Potential for data sharing (the health economic model) can be discussed with study investigators.

References

1. Metzger BE, Lowe LP, Dyer AR, et al. Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med* 2008;358:1991-2002.
2. Farrar D, Simmonds M, Bryant M, et al. Hyperglycaemia and risk of adverse perinatal outcomes: systematic review and meta-analysis. *BMJ*. 2016;354:i4694. doi: 10.1136/bmj.i4694.
3. Metzger BE, Gabbe SG, Persson B, 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.
4. WHO Health organisation, 2013. Diagnostic criteria and classification of hyperglycaemia first detected in pregnancy WHO/NMH/MND/13.2 http://apps.who.int/iris/bitstream/10665/85975/1/WHO_NMH_MND_13.2_eng.pdf.
5. American College of Obstetricians and Gynecologists. *Practice Bulletin 137: Mellitus. Obstet Gynecol* 2013; 122: 406-16
6. National Institute for Health and Care Excellence (NICE) (2015) Diabetes in pregnancy: management of diabetes and its complications from preconception to the postnatal period. Clinical guideline NG3 (2015). Available from www.nice.org.uk/guidance/ng3/resources/diabetes-in-pregnancy-management-of-diabetes-and-itscomplications-from-preconception-to-the-postnatal-period-51038446021, accessed February 2016
7. Meek CL, Lewis HB, Patient C, Murphy HR, Simmons D. Diagnosis of gestational diabetes: falling through the net. *Diabetologia* 2015;Sep;58(9):2003-12
8. Duran A, Sáenz S, Torrejón MJ et al. Introduction of IADPSG criteria for the screening and diagnosis of gestational diabetes mellitus results in improved pregnancy outcomes at a lower cost in a large cohort of pregnant women: the St. Carlos Gestational Diabetes Study. *Diabetes Care* 2014;37(9):2442–2450. doi: 10.2337/dc14-0179.

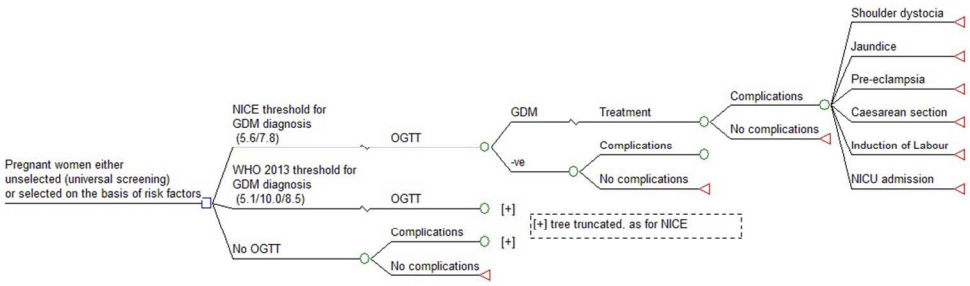
9. Herman WH. Insights offered by economic analyses. *Diabetes Care*. 2014 Sep;37(9):2424-6. doi: 10.2337/dc14-1232.
10. Farrar D, Simmonds M, Griffin S et al. The identification and treatment of women with hyperglycaemia in pregnancy: an analysis of individual participant data, systematic reviews, meta-analyses and an economic evaluation.Health Technol Assess. 2016 Nov;20(86):1-348.
11. 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.
12. Landon MB, Spong CY, Thom E, et al. A multicenter, randomized trial of treatment for mild gestational diabetes. *N Engl J Med* 2009;361:1339-48
13. Horvath K, Koch K, Jeitler K et al. Effects of treatment in women with gestational diabetes mellitus: systematic review and meta-analysis. *BMJ (Clin Res Ed)* 2010;340:c1395
14. Round,J.A., Jacklin,P., Fraser,R.B., Hughes,R.G., Mugglestone,M.A., Holt,R.I. Screening for gestational diabetes mellitus: cost-utility of different screening strategies based on a woman's individual risk of disease, *Diabetologia* 2011,54(2), 256-263. doi: 10.1007/s00125-010-1881-y
15. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. October 2014 (https://www.nice.org.uk/media/default/about/what-we-do/our-programmes/developing-nice-guidelines-the-manual.pdf)
16. Briggs A, Claxton K, Sculpher M. Decision Modelling for Health Economic Evaluation. Oxford: Oxford University Press; 2006
17. Avalos GE, Owens LA, Dunne F. Applying Current Screening Tools for Gestational Diabetes Mellitus to a European Population: Is It Time for Change? *Diabetes Care*. 2013 Oct;36(10):3040-4. doi: 10.2337/dc12-2669. Epub 2013 Jun 11.
18. Simmons D, Moses RG. Gestational Diabetes Mellitus: To Screen or Not to Screen? Is this really still a question? Diabetes Care. 2013 Oct;36(10):2877-2878
19. Armstrong R, Waters E, Doyle J (editors). Chapter 21: Reviews in health promotion and public health. In Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of

Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org. Accessed June 2015

20. Werner EF, Pettker CM, Zucerwise et al. Screening for gestational diabetes mellitus: are the criteria proposed by the international association of the Diabetes and Pregnancy Study Groups cost-effective? Diabetes Care. 2012 Mar;35(3):529-35. doi: 10.2337/dc11-1643.

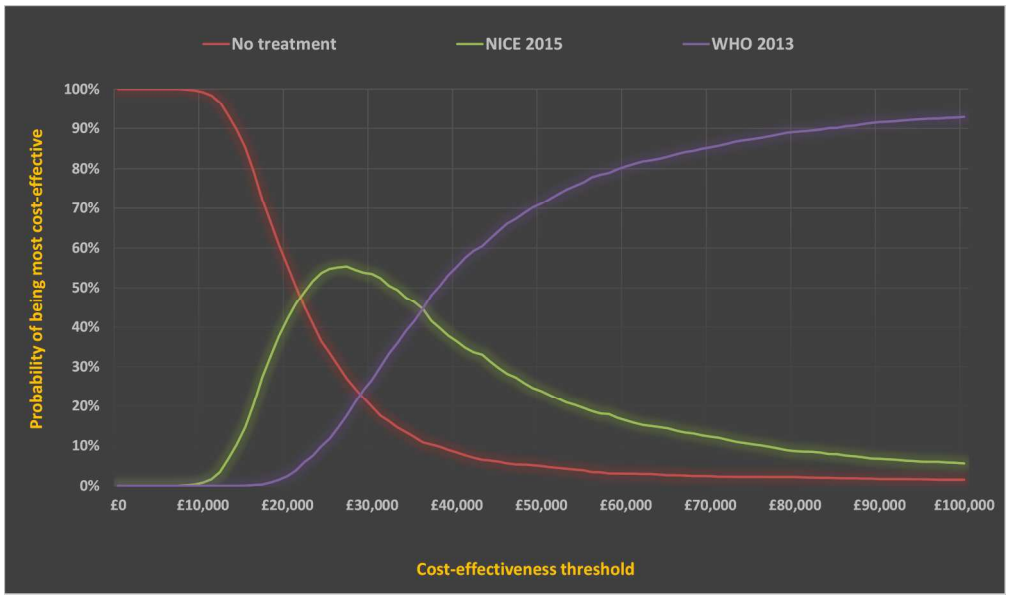
21. Donovan LE, Cundy T. Does exposure to hyperglycaemia in utero increase the risk of obesity and diabetes in the offspring? A critical reappraisal. Diabetic Medicine. 2015 Mar;32(3):295-304. doi: 10.1111/dme.12625. Epub 2014 Dec 17.

22. Landon MB, Rice MM, Varner MW et al. Mild gestational diabetes mellitus and long-term child health. Diabetes Care. 2015 Mar;38(3):445-52. doi: 10.2337/dc14-2159. Epub 2014 Nov 20.



Model Schematic

95x31mm (300 x 300 DPI)



Cost-effectiveness acceptability curve indicating the probability of a threshold or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for the HAPO (4) centres population with risk factors

267x157mm (300 x 300 DPI)

Supplementary Report

This supplementary document provides further details about model parameter estimates and model results.

Table x1. Ethnicity of women in patient datasets and of UK

Ethnic group	HAPO (4) centres	Atlantic DiP	Norfolk[a]	UK[b]
White	79%	93%	96.5%	87%
Black	2%	2%	0.5%	3%
Asian	13%	4%	1.6%	7%
Other	6%	1%	1.6%	3%

(a) Our Norwich dataset did not include data on ethnicity and the values presented here are census data for Norfolk (Estimated from 2011 Census: Ethnic group, local authorities in the United Kingdom. Office for National Statistics. 11 October 2013)

(b) Included for comparative purposes (2011 Census: Ethnic group, local authorities in the United Kingdom. Office for National Statistics. 11 October 2013)

Multivariable prediction models to estimate baseline risk

Model 1 includes the covariates used in the original analysis of the HAPO data whilst Model 2 is restricted to plasma glucose variables (Tables x2 to Tables x7). In the base case analysis, backward elimination of plasma glucose variables with non-significant coefficients from the prediction models was undertaken. A sensitivity analysis was undertaken retaining all plasma glucose variables. For each model Hosmer-Lemeshow goodness-of-fit statistics are presented and predicted probabilities are used to derive the area under the receiver-operating characteristic (ROC) curve as an indicator of the model's discriminatory ability.

Table x2. Logistic regression models to predict neonatal shoulder dystocia

	Co-efficient b (Standard error (SE(b))			
	Model 1 (all covariates)		Model 2 (blood glucose covariates)	
Variable	**Base case analysis**	**Sensitivity analysis**	**Base case analysis**	**Sensitivity analysis**
Centre (Manchester v Belfast)	1.151 (0.423)	1.151 (0.424)	-	-
(Brisbane v Belfast)	0.505 (0.489)	0.562 (0.491)	-	-
(Newcastle v Belfast)	1.604 (0.472)	1.622 (0.472)	-	-
Age at OGTT (yr)	-0.023 (0.024)	-0.022 (0.024)	-	-
BMI at OGTT (kg/m^2)	-0.006 (0.023)	-0.011 (0.024)	-	-
Smoker (Yes v No)	-0.480 (0.409)	-0.477 (0.409)	-	-
Drinker (Yes v No)	-0.101 (0.317)	-0.107 (0.317)	-	-
Family history DM (Yes v No)	-0.006 (0.184)	-0.008 (0.187)	-	-
Gestational age at OGTT (wk)	-0.111 (0.091)	-0.114 (0.092)	-	-
Neonate gender (F v M)	-1.321 (0.292)	-1.316 (0.292)	-	-
Family history HBP (Yes v No)[a]	-	-	-	-
Maternal UTI (Yes v No)[a]	-	-	-	-
Mean Blood Pressure (mmHg)	-0.006 (0.015)	-0.007 (0.015)	-	-
Hospital admission before delivery (Yes v No)	0.173 (0.266)	0.175 (0.267)	-	-
Parity (1 v 0)	-0.118 (0.420)	-0.108 (0.420)	-	-
(2+ v 0)	0.456 (0.412)	0.469 (0.414)	-	-
(Unknown v 0)	-0.026 (0.399)	-0.013 (0.399)	-	-
Fasting blood glucose[b]	-	0.151 (0.112)	-	0.166 (0.110)
1-hr blood glucose[b]	-	-0.138 (0.165)	-	-0.152 (0.163)
2-hr blood glucose[b]	0.223 (0.100)	0.222 (0.152)	0.267 (0.097)	0.265 (0.151)
Constant	0.925 (3.025)	1.139 (3.508)	-4.467 (0.122)	-4.475 (0.122)
Hosmer-Lemeshow goodness-of-fit test	χ^2 = 2.94, df=8; P=0.94	χ^2 = 6.36, df=8; P=0.61	χ^2 = 4.99, df=8; P=0.76	χ^2 = 11.51, df=8; P=0.18
Area under the ROC curve (95% CI)	0.75 (0.70, 0.80)	0.76 (0.70, 0.81)	0.58 (0.51, 0.65)	0.60 (0.53, 0.67)

(a) Omitted from HAPO model for shoulder dystocia

(b) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for shoulder dystocia arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x3. Logistic regression models to predict caesarean section

	Co-efficient b (Standard error (SE(b))			
	Model 1 (all covariates)		Model 2 (blood glucose covariates)	
Variable	**Base case analysis**	**Sensitivity analysis**	**Base case analysis**	**Sensitivity analysis**
Centre (Manchester v Belfast)	-0.494 (0.092)	-0.495 (0.092)	-	-
(Brisbane v Belfast)	-0.099 (0.098)	-0.114 (0.100)	-	-
(Newcastle v Belfast)	-0.681 (0.140)	-0.692 (0.141)	-	-
Age at OGTT (yr)	0.034 (0.007)	0.034 (0.007)	-	-
BMI at OGTT (kg/m^2)	0.039 (0.007)	0.039 (0.007)	-	-
Smoker (Yes v No)	-0.304 (0.106)	-0.292 (0.106)	-	-
Drinker (Yes v No)	-0.028 (0.087)	-0.025 (0.087)	-	-
Family history DM (Yes v No)	0.050 (0.057)	0.052 (0.057)	-	-
Gestational age at OGTT (wk)	0.004 (0.029)	0.004 (0.029)	-	-
Neonate gender (F v M)	-0.205 (0.071)	-0.205 (0.071)	-	-
Family history HBP (Yes v No)[a]	-	-	-	-
Maternal UTI (Yes v No)[a]	-	-	-	-
Mean Blood Pressure (mmHg)	0.003 (0.004)	0.003 (0.004)	-	-
Hospital admission before delivery (Yes v No)	0.514 (0.079)	0.510 (0.079)	-	-
Parity (1 v 0)[a]	-	-	-	-
(2+ v 0)[a]	-	-	-	-
(Unknown v 0)[a]	-	-	-	-
Fasting blood glucose[b]	-	-0.009 (0.044)	-	0.053 (0.040)
1-hr blood glucose[b]	0.144 (0.037)	0.101 (0.051)	0.138 (0.046)	0.119 (0.048)
2-hr blood glucose[b]	-	0.071 (0.048)	0.123 (0.046)	0.113 (0.046)
Constant	-3.518 (0.947)	-3.509 (0.950)	-1.435 (0.035)	-1.433 (0.035)
Hosmer-Lemeshow goodness-of-fit test	χ^2 = 1.88, df=8; P=0.99	χ^2 = 5.11, df=8; P=0.75	χ^2 = 16.56, df=8; P=0.04	χ^2 = 17.66, df=8; P=0.02
Area under the ROC curve (95% CI)	0.65 (0.63, 0.66)	0.65 (63, 0.66)	0.58 (0.56, 0.60)	0.58 (0.57, 0.60)

(a) Omitted from HAPO model for caesarean section

(b) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for caesarean section arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x4. Logistic regression models to predict neonatal intensive care unit admissions

	Co-efficient b (Standard error (SE(b))			
	Model 1 (all covariates)		Model 2 (blood glucose covariates)	
Variable	**Base case analysis**	**Sensitivity analysis**	**Base case analysis**	**Sensitivity analysis**
Centre (Manchester v Belfast)	0.894 (0.159)	0.889 (0.159)	-	-
(Brisbane v Belfast)	1.393 (0.161)	1.400 (0.163)	-	-
(Newcastle v Belfast)	1.153 (0.190)	1.163 (0.191)	-	-
Age at OGTT (yr)	0.013 (0.009)	0.012 (0.009)	-	-
BMI at OGTT (kg/m^2)	0.025 (0.009)	0.024 (0.009)	-	-
Smoker (Yes v No)	0.209 (0.130)	0.201 (0.130)	-	-
Drinker (Yes v No)	-0.025 (0.117)	-0.023 (0.117)	-	-
Family history DM (Yes v No)	0.033 (0.069)	0.038 (0.069)	-	-
Gestational age at OGTT (wk)	-0.050 (0.038)	-0.052 (0.038)	-	-
Neonate gender (F v M)	-0.304 (0.094)	-0.302 (0.094)	-	-
Family history HBP (Yes v No)[a]	-	-	-	-
Maternal UTI (Yes v No)[a]	-	-	-	-
Mean Blood Pressure (mmHg)	0.006 (0.006)	0.006 (0.006)	-	-
Hospital admission before delivery (Yes v No)	0.794 (0.097)	0.792 (0.097)	-	-
Parity (1 v 0)	-0.474 (0.148)	-0.474 (0.148)	-	-
(2+ v 0)	-0.490 (0.157)	-0.493 (0.157)	-	-
(Unknown v 0)	-0.084 (0.135)	-0.086 (0.135)	-	-
Fasting blood glucose[b]	-	-0.003 (0.054)	-	-0.025 (0.050)
1-hr blood glucose[b]	-	0.082 (0.067)	-	0.078 (0.064)
2-hr blood glucose[b]	0.159 (0.045)	0.107 (0.063)	0.208 (0.041)	0.167 (0.060)
Constant	-3.181 (1.236)	-3.061 (1.243)	-2.374 (0.046)	-2.375 (0.046)
Hosmer-Lemeshow goodness-of-fit test	χ^2 = 14.18, df=8; P=0.08	χ^2 = 11.41, df=8; P=0.18	χ^2 = 22.16, df=8; P=0.005	χ^2 = 12.72, df=8; P=0.12
Area under the ROC curve (95% CI)	0.71 (0.69, 0.73)	0.71 (0.69, 0.73)	0.57 (0.55, 0.60)	0.57 (0.55, 0.60)

(a) Omitted from HAPO model for neonatal intensive care unit admissions

(b) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for neonatal intensive care unit admissions arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x5. Logistic regression models to predict jaundice

	Co-efficient b (Standard error (SE(b))			
	Model 1 (all covariates)		Model 2 (blood glucose covariates)	
Variable	**Base case analysis**	**Sensitivity analysis**	**Base case analysis**	**Sensitivity analysis**
Centre (Manchester v Belfast)	0.407 (0.157)	0.410 (0.157)	-	-
(Brisbane v Belfast)	0.449 (0.171)	0.420 (0.173)	-	-
(Newcastle v Belfast)	-0.315 (0.259)	-0.332 (0.259)	-	-
Age at OGTT (yr)	0.005 (0.011)	0.005 (0.011)	-	-
BMI at OGTT (kg/m^2)	-0.011 (0.011)	-0.009 (0.012)	-	-
Smoker (Yes v No)	0.082 (0.161)	0.093 (0.162)	-	-
Drinker (Yes v No)	-0.514 (0.163)	-0.508 (0.163)	-	-
Family history DM (Yes v No)	-0.060 (0.094)	-0.060 (0.094)	-	-
Gestational age at OGTT (wk)	-0.078 (0.047)	-0.077 (0.047)	-	-
Neonate gender (F v M)	-0.116 (0.113)	-0.115 (0.113)	-	-
Family history HBP (Yes v No)[a]	-	-	-	-
Maternal UTI (Yes v No)[a]	-	-	-	-
Mean Blood Pressure (mmHg)	0.018 (0.007)	0.018 (0.007)	-	-
Hospital admission before delivery (Yes v No)	0.867 (0.116)	0.865 (0.116)	-	-
Parity (1 v 0)	-0.382 (0.185)	-0.380 (0.185)	-	-
(2+ v 0)	-0.526 (0.200)	-0.526 (0.200)	-	-
(Unknown v 0)	0.078 (0.165)	0.078 (0.165)	-	-
Fasting blood glucose[b]	-	-0.055 (0.066)	-	-0.063 (0.061)
1-hr blood glucose[b]	0.216 (0.056)	0.192 (0.079)	0.237 (0.052)	0.199 (0.078)
2-hr blood glucose[b]	-	0.073 (0.074)	-	0.102 (0.072)
Constant	-1.927 (1.522)	-2.014 (1.526)	-2.846 (0.057)	-2.850 (0.057)
Hosmer-Lemeshow goodness-of-fit test	χ^2 = 8.42, df=8; P=0.39	χ^2 = 7.96, df=8; P=0.44	χ^2 = 2.47, df=8; P=0.96	χ^2 = 10.40, df=8; P=0.24
Area under the ROC curve (95% CI)	0.68 (0.65, 0.71)	0.68 (0.65, 0.71)	0.57 (0.54, 0.60)	0.58 (0.55, 0.61)

(a) Omitted from HAPO model for jaundice

(b) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for jaundice arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x6. Logistic regression models to predict pre-eclampsia

	Co-efficient b (Standard error (SE(b))			
	Model 1 (all covariates)		Model 2 (blood glucose covariates)	
Variable	**Base case analysis**	**Sensitivity analysis**	**Base case analysis**	**Sensitivity analysis**
Centre (Manchester v Belfast)	-0.784 (0.192)	-0.800 (0.193)	-	-
(Brisbane v Belfast)	-0.308 (0.200)	-0.277 (0.202)	-	-
(Newcastle v Belfast)	-0.685 (0.278)	-0.667 (0.278)	-	-
Age at OGTT (yr)	-0.009 (0.015)	-0.011 (0.015)	-	-
BMI at OGTT (kg/m^2)	0.101 (0.011)	0.097 (0.012)	-	-
Smoker (Yes v No)	-0.556 (0.245)	-0.569 (0.246)	-	-
Drinker (Yes v No)	-0.170 (0.194)	-0.168 (0.194)	-	-
Family history DM (Yes v No)	-0.004 (0.127)	0.006 (0.127)	-	-
Gestational age at OGTT (wk)	-0.092 (0.059)	-0.096 (0.059)	-	-
Neonate gender (F v M)	0.173 (0.147)	0.174 (0.147)	-	-
Family history HBP (Yes v No)	0.233 (0.150)	0.230 (0.150)	-	-
Maternal UTI (Yes v No)	0.734 (0.211)	0.721 (0.211)	-	-
Mean Blood Pressure (mmHg) [a]	-	-	-	-
Hospital admission before delivery (Yes v No) [a]	-	-	-	-
Parity (1 v 0)	-0.291 (0.240)	-0.292 (0.240)	-	-
(2+ v 0)	-0.701 (0.271)	-0.703 (0.271)	-	-
(Unknown v 0)	0.026 (0.224)	0.023 (0.224)	-	-
Fasting blood glucose[b]	-	0.062 (0.078)	0.201 (0.065)	0.183 (0.068)
1-hr blood glucose[b]	-	0.065 (0.104)	-	0.083 (0.098)
2-hr blood glucose[b]	0.272 (0.067)	0.195 (0.096)	0.196 (0.072)	0.150 (0.090)
Constant	-3.370 (1.842)	-3.107 (1.855)	-3.453 (0.075)	-3.455 (0.075)
Hosmer-Lemeshow goodness-of-fit test	χ^2 = 5.46, df=8; P=0.71	χ^2 = 8.02, df=8; P=0.43	χ^2 = 12.00, df=8; P=0.15	χ^2 = 15.98, df=8; P=0.04
Area under the ROC curve (95% CI)	0.75 (0.72, 0.78)	0.75 (0.72, 0.79)	0.65 (0.61, 0.68)	0.65 (0.61, 0.68)

(a) Omitted from HAPO model for pre-eclampsia

(b) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for pre-eclampsia arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x7. Logistic regression models to predict induction of labour

	Co-efficient b (Standard error (SE(b))		
	Model 1 (all covariates)		Model 2 (blood glucose covariates)
Variable	**Base case analysis**	**Sensitivity analysis**	**Base case analysis**
Centre (Manchester v Belfast)	-0.476 (0.077)	-0.476 (0.077)	-
(Brisbane v Belfast)	-0.337 (0.085)	-0.333 (0.087)	-
(Newcastle v Belfast)	-0.387 (0.109)	-0.384 (0.110)	-
Age at OGTT (yr)	0.006 (0.006)	0.006 (0.006)	-
BMI at OGTT (kg/m^2)	0.039 (0.006)	0.039 (0.006)	-
Smoker (Yes v No)	0.051 (0.082	0.051 (0.082)	-
Drinker (Yes v No)	0.079 (0.072)	0.079 (0.072)	-
Family history DM (Yes v No)	0.016 (0.048)	0.016 (0.048)	-
Gestational age at OGTT (wk)	0.011 (0.024)	0.011 (0.024)	-
Neonate gender (F v M)	-0.038 (0.059)	-0.038 (0.059)	-
Family history HBP (Yes v No)[a]	-	-	-
Maternal UTI (Yes v No)[a]	-	-	-
Mean Blood Pressure (mmHg)	0.008 (0.004)	0.008 (0.004)	-
Hospital admission before delivery (Yes v No)	0.608 (0.066)	0.608 (0.066)	-
Parity (1 v 0)	-0.363 (0.101)	-0.363 (0.101)	-
(2+ v 0)	-0.193 (0.105)	-0.193 (0.105)	-
(Unknown v 0)	0.141 (0.094)	0.141 (0.094)	-
Fasting blood glucose[b]	-	0.009 (0.037)	0.079 (0.033)
1-hr blood glucose[b]	-0.108 (0.041)	-0.111 (0.043)	-0.093 (0.041)
2-hr blood glucose[b]	0.096 (0.041)	0.094 (0.041)	0.100 (0.040)
Constant	-3.050 (0.794)	-3.037 (0.796)	-1.032 (0.029)
Hosmer-Lemeshow goodness-of-fit test	χ^2 = 9.08, df=8; P=0.34	χ^2 = 9.42 df=8; P=0.31	χ^2 = 9.83 df=8; P=0.28
Area under the ROC curve (95% CI)	0.63 (0.61, 0.65)	0.63 (0.61, 0.65)	0.53 (0.51, 0.55)

(a) Omitted from HAPO model for induction of labour

(b) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for induction of labour arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x8. Cholesky decomposition of shoulder dystocia variance covariance matrix (Model 1, base case)

	Constant	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Age at OGTT (yr)	BMI AT OGTT (kg/m2)	Smoker	Drinker	Family History DM	Gestational age at OGTT (wk)	Neonatal gender	Mean Blood Pressure	Parity (1 v 0)	Parity (2+ v 0)	Parity (Unknown v 0)	Hospital admission before	2-hr blood glucose
Constant	3.025																
Centre (Manchester v Belfast)	-0.104	0.410															
(Brisbane v Belfast)	-0.129	0.335	0.331														
(Newcastle v Belfast)	-0.135	0.334	0.074	0.295													
Age at OGTT (yr)	-0.005	-0.001	0.000	0.000	0.024												
BMI AT OGTT (kg/m2)	-0.001	-0.001	-0.001	-0.001	0.001	0.023											
Smoker	-0.012	0.031	0.001	-0.003	0.047	0.006	0.404										
Drinker	0.001	0.014	-0.006	-0.020	-0.051	0.008	-0.018	0.311									
Family History DM	-0.023	0.012	-0.022	-0.018	-0.005	0.000	-0.009	-0.012	0.179								
Gestational age at OGTT (wk)	-0.080	-0.009	-0.001	0.006	-0.017	-0.007	-0.002	-0.001	-0.008	0.037							
Neonatal gender (F v M)	-0.037	-0.010	-0.006	-0.010	-0.002	0.009	0.003	-0.001	-0.005	-0.080	0.278						
Mean Blood Pressure (mmHg)	-0.004	-0.001	0.000	-0.002	-0.002	-0.006	0.000	0.000	-0.001	-0.011	-0.004	0.004					
Parity (1 v 0)	-0.038	0.014	0.011	-0.029	-0.024	-0.017	0.009	-0.013	0.006	-0.084	-0.020	-0.318	0.253				
(2+ v 0)	-0.026	0.007	0.025	-0.018	-0.082	-0.043	0.000	0.006	0.011	-0.072	-0.027	-0.317	0.021	0.229			
(Unknown v 0)	-0.052	0.019	0.005	-0.020	0.028	0.004	0.015	0.001	0.006	-0.090	-0.026	-0.311	0.020	0.026	0.219		
Hospital admission before delivery	-0.008	-0.007	-0.016	-0.006	0.005	-0.033	-0.006	0.007	-0.002	-0.002	0.004	-0.063	-0.061	-0.071	-0.079	0.225	
2-hr blood glucose	0.012	-0.003	-0.003	0.005	-0.016	-0.015	0.009	0.006	0.013	0.004	0.004	-0.003	-0.007	-0.021	-0.002	-0.015	0.091

Table x9. Cholesky decomposition of caesarean section variance covariance matrix (Model 1, base case)

	Constant	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Age at OGTT (yr)	BMI AT OGTT (kg/m2)	Smoker	Drinker	Family History DM	Gestational age at OGTT	Neonatal gender	Mean Blood Pressure	Hospital admission	1-hr blood glucose
Constant	0.947													
Centre (Manchester v Belfast)	-0.020	0.089												
(Brisbane v Belfast)	-0.028	0.047	0.082											
(Newcastle v Belfast)	-0.028	0.048	0.025	0.126										
Age at OGTT (yr)	-0.001	0.000	0.000	0.000	0.007									
BMI AT OGTT (kg/m^2)	-0.001	-0.001	0.000	0.000	0.000	0.007								
Smoker	-0.004	0.010	0.004	0.001	0.013	0.000	0.104							
Drinker	0.001	0.004	0.000	-0.002	-0.011	0.004	-0.006	0.086						
Family History DM	-0.006	0.006	-0.005	-0.002	-0.003	0.000	-0.003	-0.003	0.056					
Gestational age at OGTT (wk)	-0.026	-0.001	0.000	0.001	-0.005	-0.002	-0.001	-0.001	-0.003	0.011				
Neonate gender	-0.009	-0.001	-0.002	0.000	-0.002	0.002	0.002	-0.001	-0.001	-0.020	0.067			
Mean Blood Pressure (mmHg)	-0.001	0.000	0.000	0.000	-0.001	-0.002	0.000	0.000	0.000	-0.003	-0.001	0.001		
Hospital admission before delivery	-0.003	-0.004	-0.007	-0.001	0.004	-0.009	-0.003	0.004	-0.001	0.000	0.000	-0.042	0.065	
1-hr blood glucose	0.005	-0.001	0.002	0.002	-0.006	-0.006	0.000	0.001	0.005	0.003	0.002	-0.003	-0.003	0.035

Table x10. Cholesky decomposition of neonatal intensive care admission variance covariance matrix (Model 1, base case)

	Constant	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Age at OGTT (yr)	BMI AT OGTT (kg/m2)	Smoker	Drinker	Family History DM	Gestational age at OGTT (wk)	Neonatal gender	Mean Blood Pressure	Parity (1 v 0)	Parity (2+ v 0)	Parity (Unknown v 0)	Hospital admission	2-hr blood glucose
Constant	1.236																
Centre (Manchester v Belfast)	-0.037	0.154															
(Brisbane v Belfast)	-0.047	0.115	0.102														
(Newcastle v Belfast)	-0.050	0.115	0.042	0.137													
Age at OGTT (yr)	-0.002	0.000	0.000	0.000	0.009												
BMI AT OGTT (kg/m2)	-0.001	0.000	0.000	0.000	0.000	0.009											
Smoker	-0.007	0.012	0.002	-0.001	0.018	-0.001	0.128										
Drinker	0.002	0.004	-0.002	-0.004	-0.017	0.006	-0.007	0.115									
Family History DM	-0.008	0.004	-0.008	-0.003	-0.003	0.000	-0.004	-0.003	0.068								
Gestational age at OGTT (wk)	-0.034	-0.003	-0.001	0.002	-0.007	-0.002	-0.001	-0.001	-0.004	0.015							
Neonatal gender (F v M)	-0.012	-0.003	-0.004	-0.002	-0.002	0.003	0.000	-0.001	-0.001	-0.025	0.090						
Mean Blood Pressure (mmHg)	-0.002	0.000	0.000	-0.001	-0.001	-0.002	0.000	0.000	0.000	-0.004	-0.001	0.001					
Parity (1 v 0)	-0.011	0.006	0.004	-0.009	-0.009	-0.005	0.002	-0.003	0.002	-0.024	-0.004	-0.103	0.102				
(2+ v 0)	-0.006	0.002	0.011	-0.005	-0.032	-0.015	-0.002	0.003	0.003	-0.021	-0.006	-0.102	0.012	0.111			
(Unknown v 0)	-0.016	0.008	0.003	-0.005	0.010	0.002	0.006	0.001	0.002	-0.025	-0.006	-0.101	0.011	0.011	0.081		
Hospital admission before delivery	-0.005	-0.001	-0.005	0.000	0.004	-0.012	-0.003	0.005	0.000	-0.001	0.000	-0.033	-0.028	-0.025	-0.032	0.075	
2-hr blood glucose	0.004	0.000	0.000	0.001	-0.007	-0.006	0.004	0.002	0.005	0.003	0.002	-0.002	-0.002	-0.006	0.000	-0.006	0.042

Table x11. Cholesky decomposition of jaundice variance covariance matrix (Model 1, base case)

	Constant	Centre (Manchester v Belfast)	Centre (Brisbane v Belfast)	Centre (Newcastle v Belfast)	Age at OGTT (yr)	BMI AT OGTT (kg/m2)	Smoker	Drinker	Family History DM	Gestational age at OGTT (wk)	Neonatal gender	Mean Blood Pressure	Parity (1 v 0)	Parity (2+ v 0)	Parity (Unknown v 0)	Hospital admission	1-hr blood glucose
Constant	1.522																
Centre (Manchester v Belfast)	-0.038	0.153															
(Brisbane v Belfast)	-0.049	0.102	0.128														
(Newcastle v Belfast)	-0.055	0.102	0.041	0.228													
Age at OGTT (yr)	-0.002	0.000	0.000	0.000	0.011												
BMI AT OGTT (kg/m2)	-0.001	-0.001	-0.001	0.000	0.000	0.011											
Smoker	-0.011	0.017	0.003	-0.001	0.023	0.000	0.158										
Drinker	0.003	0.004	0.000	-0.002	-0.020	0.006	-0.011	0.161									
Family History DM	-0.011	0.010	-0.010	-0.003	-0.005	-0.001	-0.004	-0.005	0.092								
Gestational age at OGTT (wk)	-0.042	-0.003	0.000	0.002	-0.008	-0.004	-0.001	-0.001	-0.005	0.018							
Neonatal gender (F v M)	-0.014	-0.003	-0.003	-0.002	-0.002	0.004	0.002	-0.001	0.000	-0.031	0.108						
Mean Blood Pressure (mmHg)	-0.002	0.000	0.000	0.000	-0.001	-0.003	0.000	0.000	0.000	-0.005	-0.002	0.002					
Parity (1 v 0)	-0.016	0.010	0.008	-0.006	-0.011	-0.007	0.001	-0.001	0.003	-0.032	-0.007	-0.128	0.126				
(2+ v 0)	-0.008	0.002	0.014	-0.004	-0.039	-0.016	-0.003	0.005	0.006	-0.030	-0.009	-0.127	0.016	0.144			
(Unknown v 0)	-0.021	0.013	0.006	-0.004	0.012	0.003	0.005	0.002	0.002	-0.034	-0.008	-0.126	0.015	0.013	0.094		
Hospital admission before delivery	-0.006	-0.003	-0.008	-0.002	0.005	-0.016	-0.004	0.003	-0.001	-0.002	0.001	-0.040	-0.034	-0.028	-0.042	0.089	
1-hr blood glucose	0.007	-0.001	0.004	0.003	-0.008	-0.009	0.003	0.001	0.008	0.004	0.003	-0.004	-0.004	-0.009	-0.001	-0.008	0.051

Table x12. Cholesky decomposition of pre-clampsia variance covariance matrix (Model 1, base case)

	Constant	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Age at OGTT (yr)	BMI AT OGTT (kg/m2)	Smoker	Drinker	Family History DM	Gestational age at OGTT (wk)	Neonatal gender (F v M)	Parity (1 v 0)	Parity (2+ v 0)	Parity (Unknown v 0)	Family History HBP	Maternal UTI	2-hr blood glucose
Constant	1.842																
Centre (Manchester v Belfast)	-0.045	0.187															
(Brisbane v Belfast)	-0.059	0.083	0.173														
(Newcastle v Belfast)	-0.072	0.085	0.053	0.249													
Age at OGTT (yr)	-0.004	0.000	0.000	0.000	0.015												
BMI AT OGTT (kg/m2)	-0.002	-0.001	-0.001	-0.002	0.000	0.010											
Smoker	-0.010	0.019	0.002	0.001	0.026	0.001	0.243										
Drinker	0.002	0.007	-0.002	-0.004	-0.023	0.007	-0.009	0.192									
Family History DM	-0.017	0.013	-0.009	-0.006	-0.009	-0.001	-0.004	-0.004	0.124								
Gestational age at OGTT (wk)	-0.054	-0.003	0.000	0.002	-0.013	-0.012	-0.001	-0.001	-0.008	0.011							
Neonatal gender (F v M)	-0.023	-0.004	-0.006	-0.007	-0.003	0.009	0.001	-0.002	-0.001	-0.106	0.099						
Parity (1 v 0)	-0.020	0.013	0.016	-0.006	-0.014	-0.011	0.003	-0.001	0.001	-0.121	-0.123	0.163					
(2+ v 0)	-0.011	-0.001	0.020	-0.005	-0.047	-0.030	0.000	0.000	0.007	-0.117	-0.120	0.042	0.199				
(Unknown v 0)	-0.029	0.014	0.013	-0.004	0.017	0.002	0.006	0.001	0.000	-0.122	-0.125	0.036	0.025	0.127			
Family History HBP	-0.009	0.000	0.006	-0.005	-0.017	-0.014	0.001	-0.001	0.023	-0.034	-0.037	-0.060	-0.038	-0.048	0.108		
Maternal UTI	-0.004	-0.012	0.031	0.002	0.015	-0.001	-0.014	0.004	0.009	-0.021	-0.022	-0.030	-0.018	-0.028	-0.052	0.193	
2-hr blood glucose	0.006	-0.004	0.002	0.001	-0.012	-0.009	0.005	0.004	0.008	-0.006	-0.004	-0.006	-0.011	-0.002	-0.012	-0.009	0.061

78 **Table x13.** Cholesky decomposition of induction of labour variance covariance matrix (Model 1, base case)

	Constant	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Age at OGTT (yr)	BMI AT OGTT (kg/m2)	Smoker	Drinker	Family History DM	Gestational age at OGTT	Neonatal gender (F v M)	Mean Blood Pressure	Parity (1 v 0)	Parity (2+ v 0)	Parity (Unknown v 0)	Hospital admission	1-hr blood glucose	2-hr blood glucose
Constant	0.794																	
Centre (Manchester v Belfast)	-0.018	0.074																
(Brisbane v Belfast)	-0.024	0.039	0.072															
(Newcastle v Belfast)	-0.024	0.039	0.020	0.097														
Age at OGTT (yr)	-0.001	0.000	0.000	0.000	0.006													
BMI AT OGTT (kg/m2)	0.000	0.000	0.000	0.000	0.000	0.006												
Smoker	-0.005	0.008	0.002	0.001	0.013	0.001	0.081											
Drinker	0.001	0.003	0.000	-0.002	-0.010	0.003	-0.006	0.071										
Family History DM	-0.006	0.005	-0.004	-0.002	-0.002	0.000	-0.002	-0.003	0.047									
Gestational age at OGTT (wk)	-0.021	-0.001	0.000	0.001	-0.004	-0.002	-0.001	0.000	-0.003	0.010								
Neonatal gender (F v M)	-0.007	-0.002	-0.002	-0.001	-0.001	0.002	0.001	-0.001	-0.001	-0.016	0.056							
Mean Blood Pressure (mmHg)	-0.001	0.000	0.000	0.000	-0.001	-0.001	0.000	0.000	0.000	-0.003	-0.001	0.001						
Parity (1 v 0)	-0.008	0.008	0.005	-0.004	-0.006	-0.004	0.000	-0.002	0.001	-0.020	-0.005	-0.075	0.061					
(2+ v 0)	-0.004	0.004	0.008	-0.002	-0.020	-0.010	-0.002	0.001	0.002	-0.019	-0.006	-0.075	0.005	0.067				
(Unknown v 0)	-0.012	0.007	0.003	-0.002	0.006	0.001	0.002	0.001	0.001	-0.021	-0.005	-0.075	0.005	0.004	0.050			
Hospital admission before delivery	-0.004	-0.004	-0.005	-0.001	0.002	-0.007	-0.002	0.003	0.000	-0.001	0.000	-0.017	-0.017	-0.016	-0.020	0.055		
1-hr blood glucose	0.004	0.000	0.004	0.002	-0.004	-0.003	-0.001	0.000	0.003	0.001	0.001	0.000	0.001	-0.001	0.001	0.000	0.040	
2-hr blood glucose	0.001	0.000	-0.003	-0.001	-0.001	-0.002	0.004	0.001	0.001	0.002	0.000	-0.001	0.000	-0.002	0.001	-0.001	-0.027	0.030

79

Table x14: Model unit costs

Category	Cost	Standard Error	Distribution[a]	Source
2 sample OGTT	£8.07	n/a	n/a	NICE 2015[b]
3 sample OGTT	£12.11	n/a	n/a	NICE 2015[b]
Rapilose OGTT solution	£3.48	n/a	n/a	BNF July 2016[c]
Health Care Assistant Band 3 (per hour)	£25	n/a	n/a	Unit Costs of Health and Social Care 2015[d]
Nurse Band 7 (per hour of patient contact)	£147	n/a	n/a	Unit Costs of Health and Social Care 2015[d]
Dietician	£38	n/a	n/a	Unit Costs of Health and Social Care 2015[d]
Antenatal appointment	£96	£9.07	Normal	NHS Reference Costs 2014-15[e]
Ultrasound scan	£112	£7.65	Normal	NHS Reference Costs 2014-15[e]
Rapid acting insulin	£0.02	n/a	n/a	BNF June 2016[c]
Regular insulin	£0.02	n/a	n/a	BNF June 2016[c]
Needles	£0.10	n/a	n/a	NHS Drugs Tariff June 2016[f]
Lancets	£0.03	n/a	n/a	NHS Drugs Tariff June 2016[f]
Strips	£0.18	n/a	n/a	NHS Drugs Tariff June 2016[f]
Treatment of GDM	£987	n/a	n/a	Calculated
Severe hypoglycaemia	£650	n/a	n/a	NHS Reference Costs 2014-15[e]
Admission to NICU	£1,176	£38	Normal	NHS Reference Costs 2014-15[e]
Caesarean section	£982	£80	Normal	NHS Reference Costs 2014-15[e]
Neonatal death	£777	£39	Normal	NHS Reference Costs 2014-15[e]
Shoulder dystocia	£1,394	£79	Normal	NHS Reference Costs 2014-15[e]
Birth trauma	£1,394	£79	Normal	NHS Reference Costs 2014-15[e]
Serious perinatal complication (death, shoulder dystocia, birth trauma)	£1,347	n/a	n/a	Calculated
Phototherapy	£788	£72	Normal	NHS Reference Costs 2014-15[e]
Pre-eclampsia	£4,750	n/a	n/a	NICE 2015[b]

(a) The method used to obtain standard errors and the choice of a normal distribution for probabilistic sampling is described in detail in the NICE 2015 guideline[6]

(b) National Institute for Health and Care Excellence (NICE) (2015) Diabetes in pregnancy: management of diabetes and its complications from preconception to the postnatal period. Clinical guideline NG3 (2015).
(c) British National Formulary. July 2016. https://www.medicinescomplete.com/mc/bnf/current/ (accessed 4 Aug 2016).
(d) Unit Costs of Health and Social Care 2015. Personal Social Services Research Unit, The University of Kent, 2015.
(e) Department of Health. NHS reference costs: financial year 2014–2015. https://www.gov.uk/government/publications/nhs-reference-costs-2014-to-2015 , Department of Health, 2015.
(f) NHS Electronic Drug Tariff, August 2016. http://www.drugtariff.nhsbsa.nhs.uk/#/00336026-DD_1/DD00336022/Home (accessed 4 Aug 2016).

QALYs

A QALY loss was estimated for each individual component (shoulder dystocia, death and birth trauma) of the composite serious perinatal outcome, which was used in the ACHOIS study.[11] A weighting for each individual component was derived according to their relative frequency in the selected studies to assess treatment effectiveness.[11, 12] These were then used in order to derive a weighted average for a serious perinatal complication as shown in Table x15. QALY losses from a serious perinatal complication could be experienced over a lifetime and therefore an annual discount rate of 3.5% was applied in line with NICE methods.[19] For each patient, an expected QALY decrement is calculated based on their risk of serious perinatal complications. These individual patient QALY decrements are then summed across all patients to give the total QALY decrement for the patient dataset for each different diagnostic threshold.

Table x15: QALY losses and weights from individual components of the composite outcome of serious perinatal complications

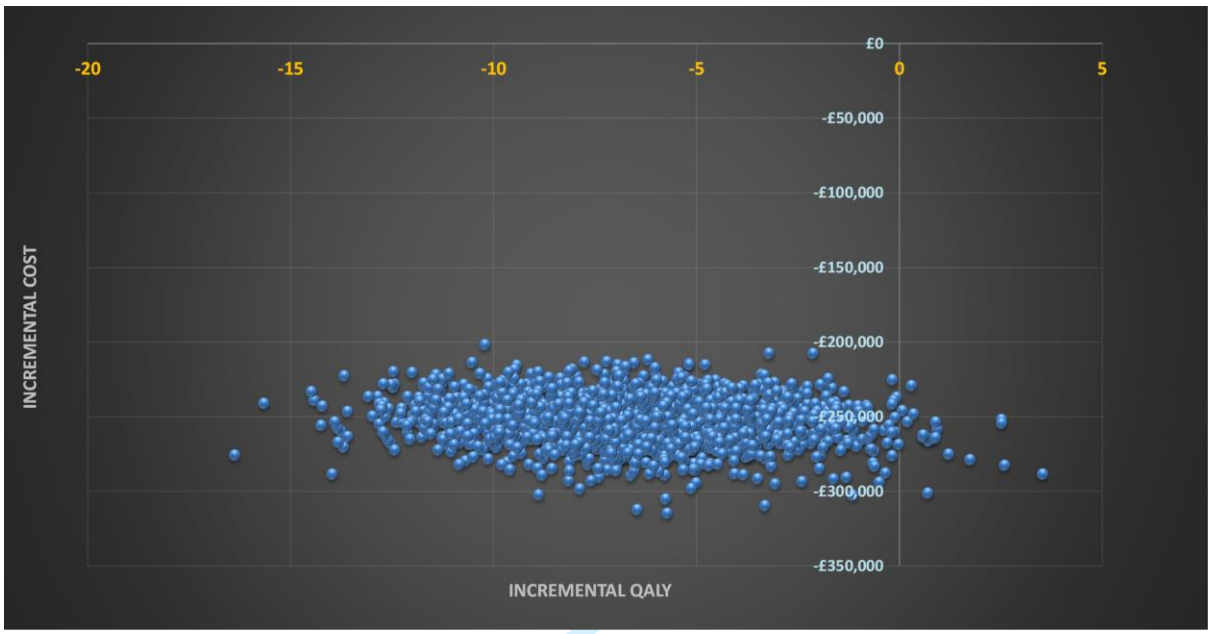
Complication	Weight	QALY	Weighted QALY
Death	0.08	25	2.00
Shoulder dystocia	0.73	0.2	0.15
Birth trauma	0.20	0.2	0.04

The analyses presented in this paper include a maternal health state utility which was estimated from quality of life data collected as part of the ACHOIS study. Whilst treatment conferred a small benefit in maternal health state utility, this was small in comparison to QALYs derived from infant outcomes. The

value of the maternal health state utility with and without treatment is the same as has been used previously.[6]

Results for the HAPO (4) population with risk factors

Figure x1: Cost-effectiveness plane for NICE 2015 compared with WHO 2013 for HAPO (4) with risk factors



Summary of results for each model population

Table x16: Summary of deterministic ICERs for each population with backward elimination of plasma glucose variables with non-significant coefficients

	All covariates		Plasma glucose covariates				
Diagnostic threshold	**HAPO Risk factor (n=3,549)**	**HAPO No Risk factor (n=2,614)**	**HAPO Risk factor (n=3,549)**	**HAPO No Risk factor (n=2,614)**	**Atlantic DiP Risk factor (n=1,988)**	**Atlantic DiP No Risk factor (n=3,302)**	**Norwich (n=12,754)**
No Treatment	-	-	-	-	-	-	-
NICE 2015	£20,400	£36,878	£22,281	£30,449	£20,830	£31,136	£28,893
WHO 2013	£33,596	£141,812	£36,473	£88,661	£35,941	£40,526	£37,918

Table x17: Probability that a threshold is cost-effective at a threshold of £30,000 per QALY and the net monetary benefit in each population using regression models with backward elimination of plasma glucose variables with non-significant coefficients

	All covariates		Plasma glucose covariates				
Diagnostic threshold	**HAPO Risk factor (NMB)**	**HAPO No Risk factor (NMB)**	**HAPO Risk factor (NMB)**	**HAPO No Risk factor (NMB)**	**Atlantic DiP Risk factor (NMB)**	**Atlantic DiP No Risk factor (NMB)**	**Norwich (NMB)**
No Treatment	19.8% (£486)	78.0% (£203)	34.4% (£361)	66.5% (£235)	27.6% (£326)	68.6% (£268)	59.7% (£938)
NICE 2015	53.5% (£230,798)	22.0% (-£57,048)	53.3% (£108,074)	33.4% (-£32,878)	58.3% (£123,6000)	25.3% (-£34,626)	29.6% (-£78,394)
WHO 2013	26.8% (£178,231)	0.1% (-£110,895)	12.4% (£18,317)	0.2% (-£76,674)	14.2% (£48,384)	6.2% (-£106,298)	10.8% (-£380,299)

Results for the HAPO (4) population without risk factors

Table x18: Clinical outcomes for HAPO (4) population without NICE risk factors (n=2,614)

Diagnostic threshold	Diagnosed	SD	SPC	CS	NICU	Jaund	PE	IOL
No Treatment	0	24	34	466	188	126	55	647
NICE 2015	208	23	31	460	184	124	51	655
WHO 2013	253	23	31	459	184	123	51	657

Table x19: Deterministic analysis for HAPO (4) population without NICE risk factors (n=2,614)

Diagnostic threshold	Cost[a]	QALY[a]	Incremental cost	Incremental QALY	ICER
No Treatment	£0	0.00	n/a	n/a	n/a
NICE 2015	£238,074	6.46	£238,074	6.46	£36,878
WHO 2013	£297,364	6.87	£59,290	0.41	£141,812

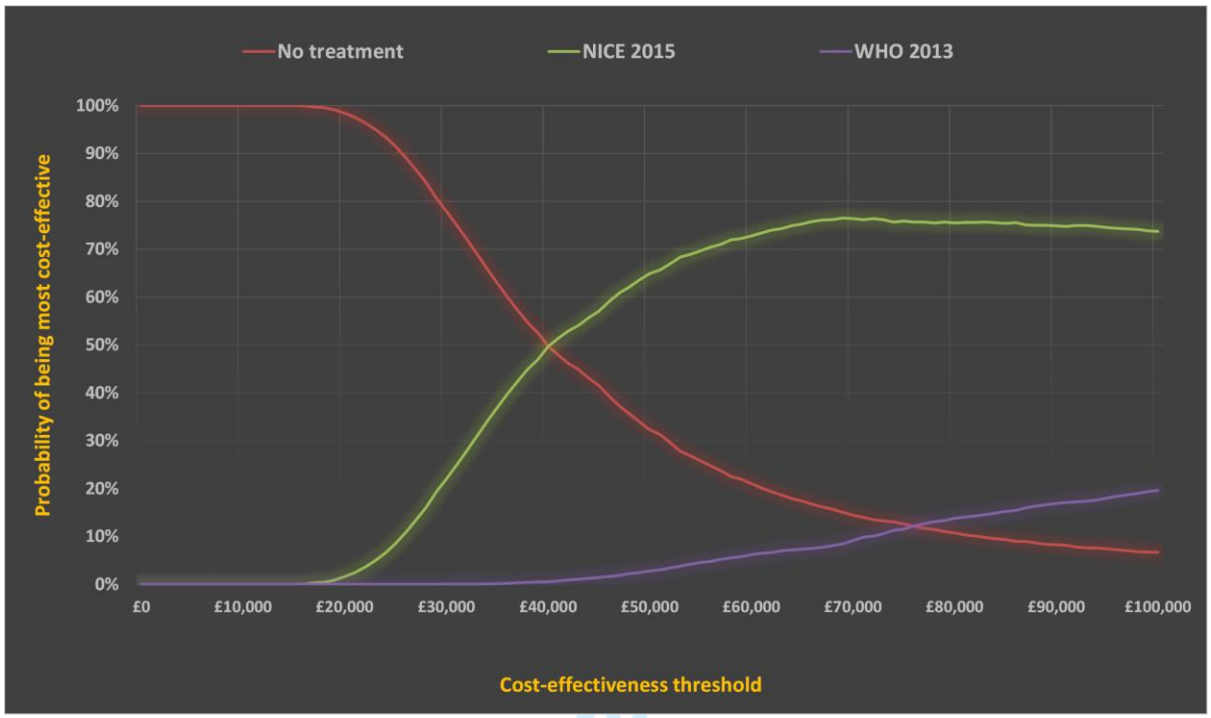
a) *Costs and QALYs are measured relative to a baseline of No Treatment*

Table x20: Probabilistic sensitivity analysis for HAPO (4) in a population without NICE risk factors

Diagnostic threshold	NMB[a] CE threshold £30,000 per QALY	Probability cost-effective CE threshold £30,000 per QALY
No Treatment	£203	78.0%
NICE 2015	-£57,048	22.0%
WHO 2013	-£110,895	0.1%

a) NMB is measured relative to the least costly and least effective strategy in each simulation

Figure x2: Cost-effectiveness acceptability curve indicating the probability of a threshold or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for HAPO (4) population without risk factors



Results for the Atlantic DiP population with risk factors

Table 21: Clinical outcomes for Atlantic DiP population with NICE risk factors (n=1,988)

Diagnostic threshold	Diagnosed	SD	SPC	CS	NICU	Jaund	PE	IOL
No Treatment	0	25	34	408	177	122	73	522
NICE 2015	497	19	26	391	163	116	56	545
WHO 2013	749	17	24	385	158	112	51	555

Table x22: Deterministic analysis for the Atlantic DiP population with NICE risk factors (n=1,988)

Diagnostic threshold	Cost[a]	QALY[a]	Incremental cost	Incremental QALY	ICER
No Treatment	£0	0.00	n/a	n/a	n/a
NICE 2015	£414,714	19.91	£414,714	17.46	£20,830
WHO 2013	£638,590	26.14	£223,876	6.23	£35,941

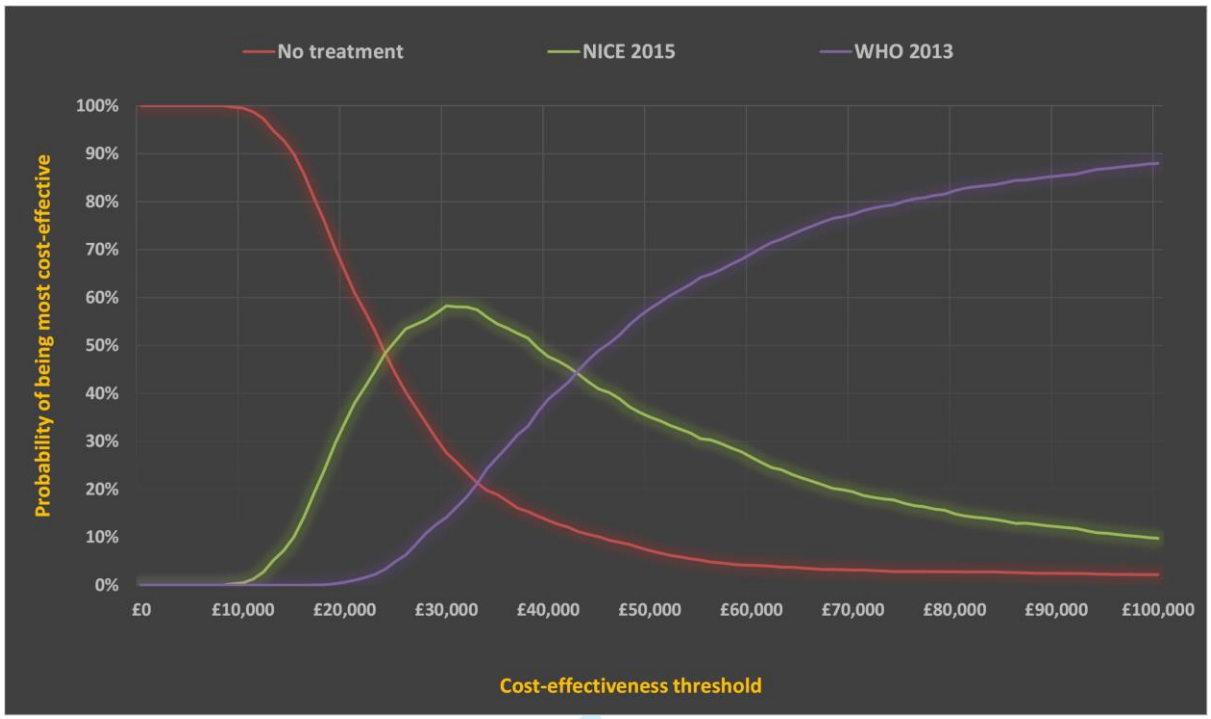
a) *Costs and QALYs are measured relative to a baseline of No Treatment*

Table x23: Probabilistic sensitivity analysis for Atlantic in a population with NICE risk factors

Diagnostic threshold	NMB[a] CE threshold £30,000 per QALY	Probability cost-effective CE threshold £30,000 per QALY
No Treatment	£326	27.6%
NICE 2015	£123,600	58.3%
WHO 2013	£48,384	14.2%

a) NMB is measured relative to the least costly and least effective strategy in each simulation

Figure x3: Cost-effectiveness acceptability curve indicating the probability of a threshold or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for the Atlantic DiP centres population with risk factors



Results for the Atlantic DiP population without risk factors

Table x24: Clinical outcomes for Atlantic DiP population without NICE risk factors (n=3,302)

Diagnostic threshold	Diagnosed	SD	SPC	CS	NICU	Jaund	PE	IOL
No Treatment	0	33	45	575	254	168	84	828
NICE 2015	194	31	42	569	248	166	79	837
WHO 2013	371	30	41	564	245	163	76	844

Table x25: Deterministic analysis for the Atlantic DiP population without NICE risk factors (n=3,302)

Diagnostic threshold	Cost[a]	QALY[a]	Incremental cost	Incremental QALY	ICER
No Treatment	£0	0.00	n/a	n/a	n/a
NICE 2015	£231,633	7.44	£231,633	7.44	£31,136
WHO 2013	£402,014	11.64	£170,381	4.20	£40,526

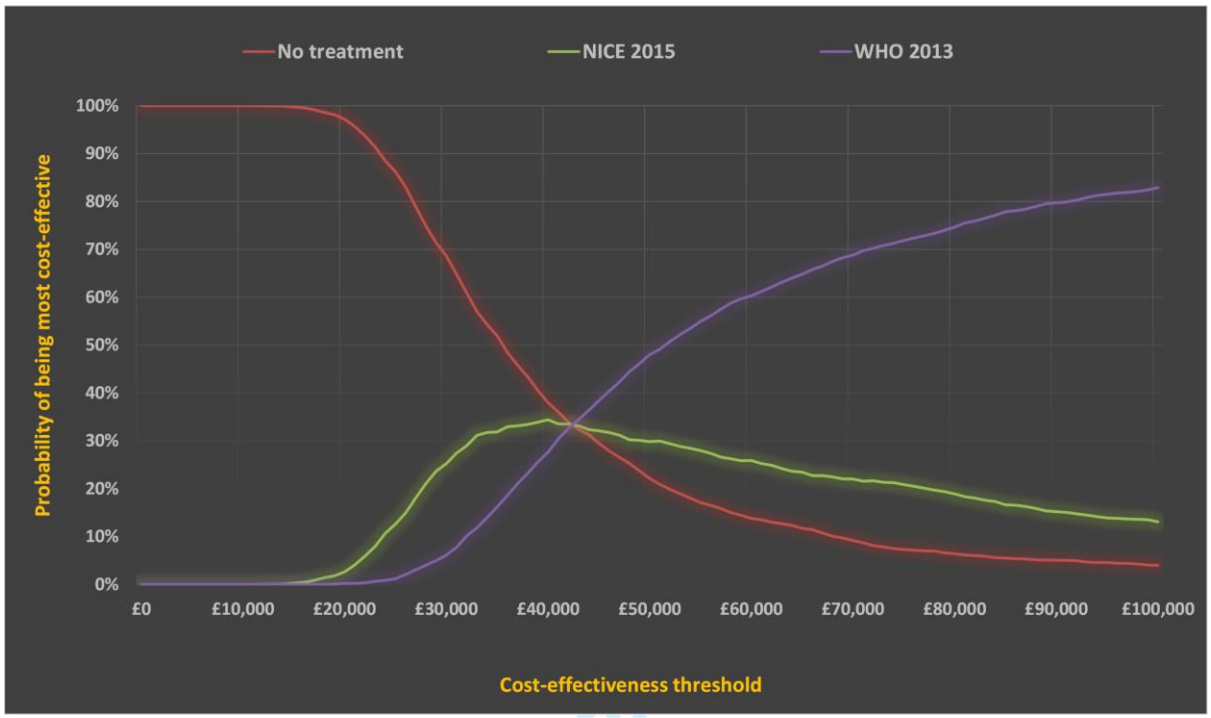
a) *Costs and QALYs are measured relative to a baseline of No Treatment*

Table x26: Probabilistic sensitivity analysis for the Atlantic DiP population without NICE risk factors

Diagnostic threshold	NMB[a] CE threshold £30,000 per QALY	Probability cost-effective CE threshold £30,000 per QALY
No Treatment	£268	68.6%
NICE 2015	-£34,626	25.3%
WHO 2013	-£106,298	6.2%

a) NMB is measured relative to the least costly and least effective strategy in each simulation

Figure x4: Cost-effectiveness acceptability curve indicating the probability of a threshold or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for the Atlantic DiP centres population without risk factors



Results for the Norwich population

Table x27: Clinical outcomes for Norwich population (n=12,754)

Diagnostic threshold	Diagnosed	SD	SPC	CS	NICU	Jaund	PE	IOL
No Treatment	0	132	182	2,333	1,005	699	346	3,173
NICE 2015	888	122	168	2,305	981	687	318	3,214
WHO 2013	1,771	117	161	2,283	965	676	301	3,248

Table x28: Deterministic analysis for the Norwich population (n=12,754)

Diagnostic threshold	Cost[a]	QALY[a]	Incremental cost	Incremental QALY	ICER
No Treatment	£0	0.00	n/a	n/a	n/a
NICE 2015	£979,903	33.91	£979,903	33.91	£28,893
WHO 2013	£1,803,196	55.63	£823,293	21.72	£37,918

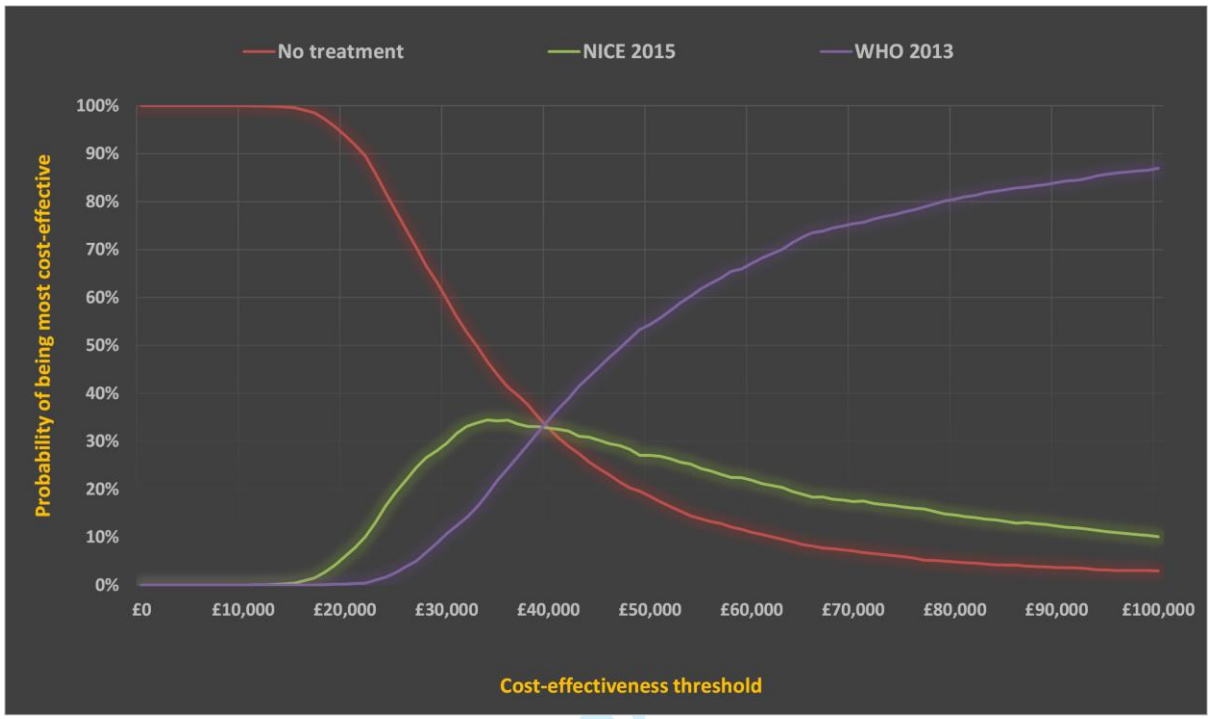
a) *Costs and QALYs are measured relative to a baseline of No Treatment*

Table x29: Probabilistic sensitivity analysis for the Norwich population

Diagnostic threshold	NMB[a] CE threshold £30,000 per QALY	Probability cost-effective CE threshold £30,000 per QALY
No Treatment	£938	59.7%
NICE 2015	-£78,394	29.6%
WHO 2013	-£380,299	10.8%

a) NMB is measured relative to the least costly and least effective strategy in each simulation

Figure x5: Cost-effectiveness acceptability curve indicating the probability of a threshold or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for the Norwich population



Deterministic sensitivity analysis

The cost-effectiveness of universal screening was not generally affected when the model was re-run using the regression models without backward elimination of non-significant variables with no screening/no treatment continuing to be the cost-effective option in populations not selected on the basis of NICE risk factors (see Table x30). In the Norwich population, universal screening was borderline cost-effective compared to no screening/no treatment at £30,000 per QALY but the same point remains that a risk factor subset in this population would have a lower ICER than that reported, and that a subset without risk factors, (i.e. those additionally incorporated as a result of universal screening compared to risk factor screening), would have a higher ICER. In populations with NICE risk factors the NICE 2015 diagnostic thresholds were still found to be cost-effective at a threshold of £30,000 per QALY, with broadly similar ICERs as previously. Similarly, the WHO 2013 diagnostic threshold was never found to be cost effective even in a population with risk factors.

Table x30: Summary of deterministic ICERs for each population without backward elimination of non-significant coefficients

	All covariates		Plasma glucose covariates				
Diagnostic threshold	**HAPO Risk factor (n=3,549)**	**HAPO No Risk factor (n=2,614)**	**HAPO Risk factor (n=3,549)**	**HAPO No Risk factor (n=2,614)**	**Atlantic DiP Risk factor (n=1,988)**	**Atlantic DiP No Risk factor (n=3,302)**	**Norwich (n=12,754)**
No Treatment	-	-	-	-	-	-	-
NICE 2015	£20,162	£38,869	£21,786	£33,473	£19,557	£32,762	£27,354
WHO 2013	£30,734	£94,585	£32,267	£58,604	£35,285	£39,076	£38,402

Mean plasma glucose values according to risk factor status

Table x311: Mean plasma glucose values in HAPO (4) and Atlantic DiP population according to their risk factor status

	HAPO (4)			Atlantic DiP		
	Fasting	1-hour	2-hour	Fasting	1-hour	2-hour
True Positives	5.24	9.90	7.89	5.21	10.21	7.61
False Positives	4.50	7.20	5.95	4.33	6.75	5.33
True Negatives	4.44	6.95	5.78	3.92	5.99	4.76
False Negatives	4.89	9.52	7.41	4.90	9.51	7.12

CHEERS Statement

CHEERS checklist—Items to include when reporting economic evaluations of health interventions

Section/item	Item No	Recommendation	Reported on page No/ line No
Title and abstract			
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	Yes Page 1 Line 2
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.	Yes Page 2 Lines 30-56
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study.	Yes Page 5-6 Lines 93-105
		Present the study question and its relevance for health policy or practice decisions.	Yes Page 5 Lines 75-91
Methods			
Target population and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	Yes Page 7-8 Lines 120-153
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	Yes Page 7 Lines 120-127
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	Yes Page 12 Line 220
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	Yes Page 6 Line 101-102; 109-110
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	Yes Page 12 Line 231 Supp. Report

Section/item	Item No	Recommendation	Reported on page No/ line No
			Page 15 Line 90
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	Yes Page 12 Line 231 Supp. Report Page 15 Line 90
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	Yes Pages 8-9 Lines 155-176
Measurement of effectiveness	11a	*Single study-based estimates:* Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	Yes Pages 10 Lines 205-208
	11b	*Synthesis-based estimates:* Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	Yes Pages 10 Lines 202-204
Measurement and valuation of preference based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	
Estimating resources and costs	13a	*Single study-based economic evaluation:* Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	
	13b	*Model-based economic evaluation:* Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	Yes Page 12-13 Lines 219-245 Supp. Report Page 14 Line 80
Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for	Yes Page 12 Lines 231

Section/item	Item No	Recommendation	Reported on page No/ line No
		converting costs into a common currency base and the exchange rate.	
Choice of model	15	Describe and give reasons for the specific type of decision-analytical model used. Providing a figure to show model structure is strongly recommended.	Yes Page 6 Lines 109-112; 114-116
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	Yes Page 8-9 Lines 155-176 Yes Page 13 Lines 247-250 Supp. Report Page 2-7 Supp. Report Page 15 Lines 84-102 +References to other sources
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	Yes Page 9-10 Lines 178-195 Yes Page 13-14 Lines 259-266 Supp. Report Page 2-7
Results			
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	Yes Page 11 Lines 216-217 Page 13

Section/item	Item No	Recommendation	Reported on page No/ line No
			Lines 244-245 Supp. Report Page 2-15
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	Yes Page 15 Lines 280-282 Supp. Report Page 17-25
Characterising uncertainty	20a	*Single study-based economic evaluation:* Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	
	20b	*Model-based economic evaluation:* Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	Yes Supp. Report Page 17 Lines 116-119 Supp. Report Page 26 Lines 225-227
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	Yes Supp. Report Page 17 Lines 111-114
Discussion			
Study findings, limitations, generalisability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the	Yes Pages 18-25

Section/item	Item No	Recommendation	Reported on page No/ line No
		findings and how the findings fit with current knowledge.	
Other			
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	Yes Page 26 Lines 541-546
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	Yes Page 26 Lines 537-539

BMJ Open

A COST EFFECTIVENESS COMPARISON OF THE NICE 2015 AND WHO 2013 DIAGNOSTIC CRITERIA FOR WOMEN WITH GESTATIONAL DIABETES WITH AND WITHOUT RISK FACTORS

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-016621.R2
Article Type:	Research
Date Submitted by the Author:	09-Jun-2017
Complete List of Authors:	Jacklin, Paul; Royal College of Obstetricians and Gynaecologists, National Guideline Alliance Maresh, Michael; Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre Patterson, Chris; Queen's University Belfast, Centre for Public Health Stanley, Katharine; Norfolk and Norwich University Hospitals NHS Foundation Trust, Department of Obstetrics and Gynaecology Dornhorst, Anne; Hammersmith Hospital, Department of Investigative Medicine Burman-ROY, Shona; Royal College of Obstetricians and Gynaecologists, National Guideline Alliance Bilous, Rudy; Newcastle University Medicine Malaysia
Primary Subject Heading:	Health economics
Secondary Subject Heading:	Diabetes and endocrinology, Obstetrics and gynaecology, Diagnostics
Keywords:	HEALTH ECONOMICS, DIABETES & ENDOCRINOLOGY, OBSTETRICS

Title:

A COST EFFECTIVENESS COMPARISON OF THE NICE 2015 AND WHO 2013 DIAGNOSTIC CRITERIA FOR WOMEN WITH GESTATIONAL DIABETES WITH AND WITHOUT RISK FACTORS

Authors: PB Jacklin[1], MJA Maresh[2], CC Patterson[3], KP Stanley[4], A Dornhorst[5], S Burman-Roy[1], RW Bilous[6]

Institutions:

1. Royal College of Obstetricians and Gynaecologists, London NW1 4RG, UK
2. St. Mary's Hospital, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester M13 9WL, UK
3. Centre for Public Health, Queen's University Belfast, Room 3.014, ICS Block B, Grosvenor Road, Belfast BT12 6BJ, UK
4. Department of Obstetrics and Gynaecology, Norfolk and Norwich University Hospitals NHS Foundation Trust, Colney Ln, Norwich NR4 7UY, UK
5. Faculty of Medicine, Department of Investigative Medicine, Hammersmith Hospital, Imperial College London, London, UK
6. Newcastle University Medicine Malaysia, Johor, Malaysia

Corresponding author:

Paul Jacklin
e-mail: pjacklin@rcog.org.uk

Abstract: 287 words

Main text: 5420 words

Abstract

Objectives To compare the cost effectiveness of The National Institute for Health and Care Excellence (NICE) 2015 and the World Health Organisation (WHO) 2013 diagnostic thresholds for gestational diabetes (GDM).

Setting: The analysis was from the perspective of the National Health Service (NHS) in England and Wales.

Participants: 6,221 patients from four of the Hyperglycaemia and Adverse Pregnancy Outcomes (HAPO) study centres (2 UK, 2 Australian), 6,308 patients from the Atlantic Diabetes in Pregnancy (DiP) study and 12,755 patients from UK clinical practice

Primary and secondary outcome measures planned: The incremental cost per quality adjusted life year (QALY), net monetary benefit (NMB) and the probability of being cost-effective at cost-effectiveness thresholds of £20,000 and £30,000 per QALY

Results. In a population of pregnant women from the four HAPO study centres, and utilising NICE defined risk factors for GDM, diagnosing GDM using NICE 2015 criteria had a NMB of £239,902 (relative to no treatment) at a cost-effectiveness threshold of £30,000 per QALY compared to WHO 2013 criteria which had a NMB of £186,675. NICE 2015 criteria had a 51.5% probability of being cost-effective compared to the WHO 2013 diagnostic criteria which had a 27.6% probability of being cost-effective (no treatment had a 21.0% probability of being cost-effective). For women without NICE risk factors in this population the NMB for NICE 2015 and WHO 2013 criteria were both negative relative to no treatment, and no treatment had a 78.1% probability of being cost effective.

Conclusion The NICE 2015 diagnostic criteria for GDM can be considered cost-effective relative to the WHO 2013 alternative at a cost-effectiveness (CE) threshold of £30,000 per QALY. Universal screening for GDM was not found to be cost-effective relative to screening based on NICE risk factors.

Keywords: Cost Effectiveness, Gestational Diabetes, Screening, Risk Factors, Diagnosis

Strengths and limitations of this study

- This economic evaluation addresses an important clinical and policy issue. The existing economic evidence is limited and WHO have stated that studies of this type are needed to inform a future update of their guideline
- Our paper has used patient-level data from the influential HAPO study for an economic analysis which has not been previously been published in a peer reviewed journal
- This analysis provides clear evidence that universal screening is not cost-effective in the UK
- This analysis suggests that the NICE diagnostic criteria for GDM are more cost-effective than the WHO criteria in the UK context
- Model conclusions are sensitive to uncertainties with respect to valuation of health outcomes and the possible long term metabolic consequences for offspring for which the evidence is debated and which are hard to quantify

Introduction

The diagnostic glycaemic thresholds for GDM remain the subject of considerable debate. The original definition was based upon maternal risk for developing postpartum diabetes, but subsequent thresholds have concentrated on complications during pregnancy and the health of the offspring. The publication of the HAPO study[1] demonstrated a linear association between increasing levels of maternal hyperglycaemia and adverse perinatal outcomes with no obvious threshold, an association that has also been observed in subsequent analyses.[2] The discussion around the diagnostic criteria that should define GDM has intensified. New diagnostic thresholds were proposed by the International Association of Diabetes in Pregnancy Study Group (IADPSG)[3] based upon the HAPO study levels of plasma glucose when fasting, and at 1 and 2 hours after an oral 75g glucose load that were associated with covariate adjusted odds ratio of 1.75 relative to the mean glucose value in the whole HAPO cohort on three offspring outcomes: exceeding the 90th centile for birth weight, for cord serum C-peptide concentration and for percent fetal body fat. These diagnostic criteria have been subsequently adopted by the WHO.[4] However, they remain controversial and have not been supported by bodies such as the National Institutes for Health and the American College of Obstetricians.[5] Furthermore, WHO has acknowledged that they will have to be revisited in the near future in the light of new studies reporting their cost-effectiveness.[4]

In 2015 NICE published updated guidance on Diabetes in Pregnancy[6] which included recommendations on diagnostic thresholds for GDM which differ from those adopted by WHO. These NICE thresholds were informed by an economic evaluation of the type that WHO considered important to inform future recommendations, but have attracted criticism in the UK[7] and elsewhere. Data from a published Spanish study[8] have been widely cited[7, 9] in support of the

cost effectiveness of the WHO criteria, although a UK analysis has more recently suggested that it is not cost-effective to identify gestational diabetes for treatment.[10]

In this paper we compared the cost-effectiveness of NICE 2015 and WHO 2013 diagnostic thresholds for GDM, as these are new thresholds proposed by national and international bodies. The analysis was undertaken using a revised version of the health economic model developed for the NICE guideline and was based upon data from the UK and Australian HAPO Study centres.

Methods

Model description

A decision analytic framework was used to evaluate the cost effectiveness of two recently proposed diagnostic thresholds for GDM, together with a no diagnosis/no treatment option (See Table 1). A schematic of the model is shown in Figure 1. Cost-effectiveness was evaluated using both deterministic and probabilistic sensitivity analysis.

Table 1*:* Diagnostic thresholds for plasma glucose evaluated in the economic model

Threshold name	Fasting (mmol/L)	1-hour (mmol/L)	2-hour (mmol/L)
No diagnosis/no treatment	-	-	-
NICE 2015	≥5.6	-	≥7.8
WHO 2013	≥5.1	≥10.0	≥8.5

Population

The model population comprised women of gestational age 24-28 weeks without pre-existing diabetes. The analysis utilised individual patient data from three datasets which, although not

restricted to the UK, provide a representative cross section of the demographic and patient characteristics that would be found in the UK (Table x1 in the Supplementary Report provides a breakdown of ethnic groups in each of our datasets) . The analyses were run separately for each dataset and, where possible, for subgroups with and without risk factors for GDM within a dataset.

i. HAPO – a dataset from the two UK (Manchester and Belfast) and two Australian (Brisbane and Newcastle) centres of the HAPO Study, referred to as HAPO (4)

ii. Norwich – these data were routinely collected between 2008 and February 2014 on women who had an oral glucose tolerance test (OGTT) on the basis of the presence of one or more risk factors for GDM. The results were obtained from laboratory records with no identifiers. Risk factors in addition to those recommended by NICE were used e.g. women with polycystic ovary syndrome, previous stillbirth or recurrent glycosuria.

iii. Atlantic Diabetes in Pregnancy (Atlantic DiP) – these data were collected between 2007 and 2013 as part of a research initiative in the Republic of Ireland intended to improve pregnancy outcomes for women with diabetes before, during and after pregnancy.

For the HAPO (4) and Atlantic DiP datasets the populations were stratified according to whether or not they had NICE risk factors for GDM (body mass index (BMI) above 30 kg/m^2, previous baby with birthweight ≥4.5 kg, previous GDM, first-degree relative with diabetes and minority ethnic family origin with a high prevalence of diabetes). This facilitated a comparison of the cost-effectiveness of universal screening for GDM when compared with a risk factor approach.

The NICE risk factor approach could not be replicated exactly because the patient data used in the model do not include information on previous offspring birth weight, and the HAPO (4) dataset does not provide information on previous GDM. Similarly the Atlantic DIP dataset does not include data on previous macrosomia or previous GDM. Therefore, the comparison in the model was between universal screening and a subset of NICE risk factors. Our Norwich dataset only included the plasma glucose values from a three point (fasting, 1 and 2 hour) OGTT and therefore it was not possible to assess cost-effectiveness according to the presence of risk factors in this group.

Permission was obtained from the relevant Caldecott Guardian to use anonymised patient OGTT data from the Norfolk and Norwich University Hospitals NHS Foundation Trust for the analysis. The principle investigators from the Australian (Professor HD McIntyre) and British (Professor DR McCance) centres of the HAPO study and the principle investigator of the Atlantic DiP (Professor F Dunne) study gave permission for anonymised patient data from their studies to be used in the analysis.

Clinical outcomes

The agreed outcomes for the economic model were selected prior to model development by the NICE Guideline Development Group. They were:

i. Shoulder dystocia (SD) – this was used to estimate serious perinatal complications (SPC), a broader composite outcome (death, shoulder dystocia and birth trauma) used as a primary outcome in clinical trials. The estimation of SPC from shoulder dystocia has been described elsewhere.[6]
ii. Caesarean section (CS)
iii. Neonatal intensive care unit (NICU) admission

iv. Jaundice requiring phototherapy (Jaund)

v. Pre-eclampsia (PE)

vi. Induction of labour (IOL)

Outcomes were prioritised for inclusion in the model if they had a direct impact on health related quality of life and/or cost. Birth weight was not included because there were few long-term outcome data for modelling any risk benefit of a reduction in birth weight for future diabetes and other health outcomes in the offspring.

In addition, outcomes were only included if the relationship with plasma glucose levels had been established in the HAPO study, and also that they had been assessed in intervention studies used to derive treatment effect size estimates. Possible double counting of certain outcomes was taken into account (e.g. preterm birth and NICU admission). The final list of outcomes included in the model was therefore a pragmatic one.

Baseline risk

Logistic regression analyses of patient data from HAPO (4) were used to predict a baseline risk for all six outcomes for each woman, based on their characteristics including their OGTT results. In the HAPO study the OGTT was blinded to the carers, unless there was overt diabetes, thus allowing direct comparison of the OGTT with perinatal outcomes without intermediate treatment effects for those meeting the new diagnostic criteria for GDM.

For each of the six outcomes, 2 logistic analyses to predict risk were assessed:

i. Prediction based on OGTT plasma glucose results and including the same covariates as used for Model 2 in the original analysis of the HAPO data[1] – this could not be applied to the Norwich and Atlantic DiP datasets as information on all HAPO covariates was not available

ii. Prediction based only on OGTT plasma glucose results

Backward elimination of plasma glucose variables with non-significant coefficients was undertaken to arrive at a 'final' logistic regression analysis to predict baseline risk for each outcome for the base case analysis, although a sensitivity analysis is also presented where the model was run with plasma glucose variables with non-significant coefficients retained. The logistic regression analyses used to predict the baseline risk for each outcome are shown in the Supplementary Report, Tables x2 to x7.

Clinical effectiveness

For each evaluated diagnostic threshold inTable 1 the model determined whether a woman would be identified as having GDM based on her OGTT. If the woman was not identified as having GDM then outcome probabilities were based on the predicted baseline risk, but for women identified as having GDM the predicted baseline risk was modified to take account of the effects of treatment. Treatment effectiveness for most outcomes was estimated from a random-effects meta-analysis of two studies, the Australian Carbohydrate Intolerance Study (ACHOIS) and the Landon et al. trial.[11, 12] Other published studies of treatment for GDM were adjudged to lack adequate randomisation.[13] For the NICU outcome only the Landon et al. trial data were used as it was considered to more closely represent UK practice as all neonatal nursery admissions were utilised. Similarly, the incidence of pre-eclampsia seemed high in ACHOIS in both arms, and again only Landon et al. trial data were utilised. The treatment effects for each of the model's clinical outcomes are shown in Table 2 along with parameters for probabilistic sampling. The model assumes that the relative treatment effect will be the same irrespective of the absolute baseline risk. For deterministic analyses the point estimate of relative risk was used but in order to account for uncertainty in these point estimates, these

relative risks were sampled from a log-normal distribution in the simulations undertaken for probabilistic sensitivity analysis (PSA).

Table 2: Relative treatment effects for model outcomes

Outcome	Relative risk (RR)	Standard error (log RR)	Source
Shoulder dystocia	0.41	0.316	ACHOIS (2005), Landon (2009)
Caesarean section	0.88	0.095	ACHOIS (2005), Landon (2009)
NICU	0.77	0.194	Landon (2009)
Jaundice requiring phototherapy	0.83	0.136	ACHOIS (2005), Landon (2009)
Pre-eclampsia	0.46	0.345	Landon (2009)
Induction of Labour	1.16	0.126	ACHOIS (2005), Landon (2009)

Costs

Costing was undertaken from the perspective of the NHS, was calculated for each woman in the dataset being analysed and was made up of three components;

- the costs of the diagnostic test – not applied in the *no test/no treat* strategy
- the costs of treatment- applied to every woman diagnosed with GDM at a particular threshold
- the costs associated with the various outcomes – with the cost for each woman being the expected (or average) cost of the outcome based on her estimated risk

The costs calculated for each woman were then summed across the entire patient dataset to give a total cost for a particular diagnostic threshold.

Costs are presented in pounds sterling and were taken from published UK sources where possible (cost year 2015). They have not been discounted as they are all assumed to occur within 12 months of diagnosis. Model unit costs are reported in the Supplementary Report, Table x14. The costing methodology and assumptions are described in greater detail elsewhere.[6]

Other event probabilities

Probabilities in decision analysis were used to calculate the expected costs and benefits of the various comparators. Many of these probabilities stemmed from relative treatment effects but a few additional event probabilities were included in the model in order to estimate certain costs. These probabilities are shown in Table 3 and their source is described elsewhere.[6]

Table 3: *Model event probability not derived from patient level regression*

Event	Probability
Not requiring hypoglycaemic therapy when treated for GDM	36%
Risk of hypoglycaemia if taking hypoglycaemic therapy	20%
Risk of hypoglycaemia being severe (requiring hospitalisation)	5%

Quality Adjusted Life Years (QALYs)

Following previous studies[6, 14] a QALY decrement of 2.2 was assigned to serious perinatal complications (SPC), defined as per the ACHOIS study as a composite outcome of shoulder dystocia, death and birth trauma.[11] More detail on the derivation of this QALY loss is provided in the Supplementary Report. The cost-effectiveness of a healthcare intervention is determined by the opportunity cost of the health foregone on the basis that with a fixed health budget any newly funded intervention would displace the least cost-effective treatment currently provided. In the UK, NICE typically uses a threshold of £20,000 to £30,000 per QALY as a benchmark[15] for the opportunity cost of health foregone and this paper assesses cost-effectiveness accordingly.

Sensitivity analysis

Probabilistic sensitivity analysis, using Monte Carlo simulation (with 2,000 iterations for each analysis), was undertaken in order to assess the impact of sampling uncertainty on model inputs. Parameters and distributions for the probabilistic sensitivity analysis are given in Table 2 and Table x14 in the supplementary report. For the logistic regression coefficients used to predict baseline risk, the Cholesky decomposition method[16] was used to sample from a multivariate normal distribution in order to reflect correlations between the coefficients. The Cholesky

decomposition of the variance covariance matrices from the regression analyses used in the base case probabilistic sensitivity analysis are given in Table x8 to x13 in the Supplementary Report.

Results

Table **4** shows the percentage of women diagnosed with GDM in the three populations using both of the evaluated diagnostic thresholds. In addition, for the HAPO (4) and Atlantic DiP datasets this is additionally broken down in the subgroups with and without NICE risk factors (RF).

Table 4: Percentage of women identified with GDM by threshold and population

Threshold name	Norwich (n=12,754)	HAPO all (n=6,163)	HAPO RF (n=3,549)	HAPO No RF (n=2,614)	DiP All (n=5,290)	DiP RF (n=1,988)	DiP No RF (n=3,302)
NICE 2015	7.0%	13.6%	17.7%	8.0%	13.1%	25.0%	5.9%
WHO 2013	13.9%	18.9%	25.7%	9.7%	21.2%	37.7%	11.2%

Detailed deterministic and probabilistic results for HAPO (4) with risk factors are shown in Table 5, Table 6, Table 7 and Figure 2.

Table 5: *Clinical outcomes for HAPO (4) population with NICE risk factors (n=3,549)*

Diagnostic threshold	Diagnosed	SD	SPC	CS	NICU	Jaund	PE	IOL
No Treatment	0	49	67	759	345	219	146	974
NICE 2015	629	41	56	739	326	210	123	1,004
WHO 2013	912	39	54	731	321	207	117	1,016

Table 6: *Deterministic analysis for the HAPO (4 centres) population with NICE risk factors (n=3,549)*

Diagnostic threshold	Cost[a]	QALY[a]	Incremental cost	Incremental QALY	ICER
No Treatment	£0	0.00	n/a	n/a	n/a
NICE 2015	£546,349	26.78	£546,349	26.78	£23,073
WHO 2013	£778,993	34.35	£254,376	7.57	£37,669

a) Costs and QALYs are measured relative to a baseline of No Treatment

Table 5 indicates that there was a relatively small difference in clinical outcomes contrasting NICE and WHO diagnostic criteria, despite there being a 45% increase in women diagnosed with GDM. Using the WHO 2013 criteria, instead of the NICE 2015 criteria, an additional 142 women would have been diagnosed with GDM, and treated in order to prevent 1 case of shoulder dystocia.

In the deterministic analysis the NICE 2015 diagnostic criteria would be considered cost-effective at a cost-effectiveness threshold of £30,000 per QALY (Table 6).

The probabilistic sensitivity analysis reached a similar conclusion, with the NICE 2015 diagnostic threshold having the highest probability of being the most cost-effective treatment and the highest NMB using a cost-effectiveness threshold of £30,000 per QALY (Table 7 and Figure 2). The analysis also suggested that no diagnosis/no treatment might be considered the most likely to be cost-effective when using a lower cost-effectiveness threshold of £20,000 per QALY. The probability of no diagnosis/no treatment being cost-effective falls sharply in the cost-effectiveness threshold range of £20,000 - £30,000 *per* QALY. As shown in the cost-effectiveness acceptability curve in Figure 2, the WHO 2013 diagnostic threshold becomes

more cost-effective as the cost-effectiveness threshold increases. Nevertheless, this would have to exceed £30,000 per QALY before becoming cost-effective, indicating that the further reduction in adverse outcomes, are achieved at an unacceptably high opportunity cost. The Supplementary Report plots the incremental cost and QALY outcomes of 2,000 simulations from the probabilistic analysis on the cost-effectiveness plane (see Figure x1). Whilst most points fall in the south-western quadrant, suggesting that WHO 2013 diagnostic criteria are likely to lead to additional QALYs when compared with NICE 2015 criteria, all points show that NICE 2015 criteria were associated with markedly lower costs.

Table 7: *Probabilistic sensitivity analysis for HAPO (4) in a population with NICE risk factors*

Diagnostic threshold	**NMB[a] CE threshold £30,000 per QALY**	**Probability cost-effective CE threshold £20,000 per QALY**	**Probability cost-effective CE threshold WTP = £30,000 per QALY**
No Treatment	£0	54.1%	21.0%
NICE 2015	£239,902	43.3%	51.5%
WHO 2013	£186,675	2.7%	27.6%

a) NMB is measured relative to a baseline of no treatment

Summaries of results for all of the model populations and more detailed results are provided in the Supplementary Report.

Tables x16 and x17 in the Supplementary Report show that in both the HAPO (4) and Atlantic DiP populations with NICE risk factors, the NICE diagnostic threshold is the most cost-effective strategy at a cost-effectiveness threshold of £30,000 per QALY. The NICE 2015

diagnostic threshold has ICERs of less than £30,000 per QALY, and in the probabilistic sensitivity analysis it has the highest net monetary benefit and the highest probability of being the most cost-effective. For HAPO (4) the results are similar if baseline risks are estimated using logistic regression based on all covariates or a logistic regression just using plasma glucose levels.

The results also suggested that universal screening would not be cost-effective as, when compared to risk factor screening (as recommended in NICE guidelines), the additional women included in such an approach would be those without risk factors and the model demonstrates that the ICERs for diagnosis and treatment are all well in excess of £30,000 per QALY; markedly so when using WHO 2013 diagnostic thresholds. These conclusions were supported by an analysis of the Norwich dataset (see Supplementary Report).

It was not possible to stratify the Norwich dataset according to risk factors, and therefore the ICERs presented relate to a comparison between no screening/treatment and universal screening and treatment. However, the results were consistent with those for HAPO (4) and Atlantic DiP. First, they showed that universal screening was not cost-effective even when compared to an alternative of no screening/no treatment. Second, the ICERs for the whole population were a weighted average of the populations with and without risk factors. The ICER for the population without risk factors would be higher than the ICER for the entire population, which was only marginally below the £30,000 per QALY threshold.

Deterministic sensitivity analysis

As part of a sensitivity analysis the deterministic models were re-run using the logistic regression models without backward elimination of glucose variables with non-significant

coefficients, and these analyses are discussed in the Supplementary Report with the results summarised in Table x30.

Discussion

In the NICE guideline analysis, 14 alternative diagnostic thresholds were compared and there was no single optimal diagnostic threshold which clearly emerged[6]. This is not surprising given the small differences in patient outcomes between them. In that analysis the previous WHO 1999 criteria emerged as a relatively cost-effective strategy. However, the Guideline Development Group rejected a fasting threshold of 7.0 mmol/L as there was a wide clinical consensus that this was too high, as 6.1-7.0 mmol/L is diagnostic of impaired fasting glycaemia in the non-pregnant population. Intervention studies had used a lower fasting threshold than 7.0 mmol/L as a basis for inclusion, and therefore made a case for intervention at lower levels. Based upon detailed cost effectiveness analysis of all the options, the Guideline Development Group ultimately decided on recommending a fasting plasma glucose of 5.6 mmol/L and a 2 hour plasma glucose of 7.8 mmol/L. In this paper, we have restricted our analysis of cost-effectiveness to the WHO 2013 and NICE 2015 criteria (with a no screening/treatment baseline also included) as these two recommendations have the most clinical currency at present.

All of the analyses presented in this paper suggest that, in a population with NICE risk factors, the NICE 2015 diagnostic criteria for GDM could be considered cost-effective relative to no screening/no treatment and to WHO 2013 diagnostic thresholds when using a cost-effectiveness threshold of £30,000 per QALY. The analyses also show that no screening/no treatment is cost-effective in populations without NICE risk factors, suggesting that universal screening does not represent value for money, at least in a UK setting. The slight differences in the costs and QALYs in the current analysis compared to the original NICE guideline are due to a

combination of using updated cost data and a modification of the statistical analysis utilising the Cholesky decomposition (see methods).

One of the limitations of our analysis was that the 2-hour threshold was restricted to the historical WHO 1999 2-hour definition of 7.8mmol/l, or the new WHO 2013 criteria of 8.5 mmol/l. It is conceivable that a 2-hour threshold lying between these values might outperform both. Our greater focus, though was on the optimal fasting level as this is where the greatest controversy lies with respect to potentially missed treatment opportunities.

As noted by the proponents of WHO 2013 diagnostic criteria for GDM, using a lower fasting plasma glucose threshold would by definition detect more cases. Furthermore, because we assumed in the model that the relative treatment effect would be the same in additionally diagnosed cases, it follows that such a threshold could potentially yield the lowest number of adverse outcomes and the greatest QALY gain. However, our analysis suggests that the relatively small additional gains are not justified by the substantially higher costs that such lower thresholds would require.

A key driver of our results were the logistic regression models which were used to predict baseline risk. For the outcomes included in this study these regression models suggested that the 2-hour plasma glucose was a much more important predictor of adverse outcomes than the fasting plasma glucose, something we were unaware of when selecting the model's clinical outcomes. For the regression models fitted to predict baseline risk in the HAPO (4) dataset with covariates and backward elimination of the OGTT plasma glucose variables (Model 1 base case analysis regressions in Supplementary Tables x2 to x7), the Hosmer-Lemeshow Goodness of Fit Test did not indicate evidence of poor fit ($p > 0.05$). However, there was evidence of poor fit

($p < 0.05$) for the regression models of caesarean section and NICU admission where the prediction was based only on OGTT plasma glucose results (Model 2 base case analysis regressions in Supplementary Tables x2 to x7). Nevertheless, as indicated in Supplementary Table x16 and x17, the choice of prediction model did not have a large bearing on cost-effectiveness.

We consider that our analysis which builds on previous modelling[6, 14] is, together with another recently published UK analysis[10], one of the most comprehensive assessments of the cost-effectiveness of diagnostic thresholds for GDM yet undertaken, and will hopefully contribute to the WHO's expectation "that a substantial body of new data will emerge in the near future, providing currently scarce health and economic evaluation of the recommended criteria applied to various populations and with different approaches (universal screening, screening only women at high risk, diagnostic testing only)".[4]

A number of commentators [17, 18] have recently advocated universal screening for GDM. The essence of the argument is based upon the number of cases of GDM that would be missed with selective screening, and the subsequent reduced opportunity to prevent a serious perinatal outcome. Of course it is true that universal screening will detect more cases, although the absolute numbers will depend upon the thresholds used to define GDM. Table 5 shows that many more women would need to be diagnosed in order to prevent a single adverse outcome. However, in the context of finite health care resources, it must be accepted that it may be cost-effective to miss some cases. Epidemiological measures such as number needed to treat (or number needed to screen in this case) implicitly recognise that a goal of health care systems cannot be to maximize health gain without any consideration of cost. Identifying missed cases carries an opportunity cost and it may be that those resources would achieve greater benefit if

employed elsewhere in the health care system. If a population is divided into those with risk factors and those without risk factors, then the prevalence of GDM must be lower in the group without risk factors (and the number needed to screen higher) with concomitantly lower cost-effectiveness. However, the comparative cost-effectiveness of screening in those with and without risk factors is not only affected by the respective prevalence in the two groups, but also differences in severity. In those diagnosed with GDM and who had risk factors there were, as anticipated, greater levels of hyperglycaemia than in those without risk factors. As shown in Table x31 in the Supplementary Report, 'true positives' or identified cases (risk factor present and GDM) had higher plasma glucose values than 'false negatives' or missed cases (risk factors absent and GDM) when defining GDM positives according to WHO 2013 diagnostic thresholds.

We would therefore expect the women with risk factors and GDM to be at greater risk of adverse outcomes than the women with GDM without risk factors as a result of their higher plasma glucose levels. So the "cases" missed with selective screening would have, on average, fewer adverse outcomes than in "cases" in a population with risk factors. So the ICER would be greater in the population without risk factors because prevalence is lower and cases have fewer adverse outcomes.

Our analysis, by splitting the HAPO (4) and Atlantic DiP datasets into those with and without risk factors, was able to evaluate the cost-effectiveness of moving from risk factor screening to universal screening. Whilst diagnosis in populations with risk factors was shown to be cost-effective at a threshold of £30,000 per QALY, it was never cost-effective to diagnose and treat in those without risk factors. Table 4 indicates the large differences that exist in prevalence between the populations with and without risk factors. Our analysis suggests that the cost-

effectiveness threshold would have to substantially exceed currently accepted UK norms for universal screening to be considered cost-effective. Although the NICE risk factor approach could not be replicated exactly, we felt that the approximation used was acceptable, as the only women who would be omitted from the model risk factor population were multiparous and would have had a large baby previously and/or a past history of GDM. This approximation would over-estimate slightly the benefits of universal screening, as the baseline risk in a group designated as being without NICE risk factors present would be over-stated.

A previous study[8] from Spain using WHO 2013 diagnostic criteria suggested cost effectiveness compared with a two-step protocol using the Carpenter – Coustan thresholds. However, this was largely based upon estimates of reduction of caesarean section rates of 50% which we find implausible based upon changes in diagnostic criteria alone, noting that ACHOIS and Landon et al. found only a 4% and 21% reduction in caesarean section respectively as a result of treating gestational diabetes. The Spanish study did not consider other alternative thresholds, and was a retrospective, before and after analysis which has been criticised by the Cochrane Collaboration as it does not control for possible changes in important variables, such as clinical management, over time.[19]

A recently published UK Health Technology Assessment (HTA) [10] suggested that the identification of gestational diabetes for treatment is not cost-effective, in which case finding a cost-effective threshold becomes somewhat redundant. Although the HTA followed a similar approach to our analysis there were some differences which could explain the different conclusions. In our analysis, jaundice was included as an outcome and the relative treatment effect would have tended to lower the incremental costs of intervention as a result of reduced rates of phototherapy. This was not included as an outcome in the HTA. Instrumental delivery

was included as an outcome in the HTA but not in our analysis. While instrumental delivery rates could in theory be increased by treatment, as there will be more vaginal births, this could be counteracted by those mothers not treated delivering larger babies vaginally requiring assistance; this would be in accord with the HTA meta-analysis which failed to demonstrate a treatment effect on instrumental delivery rates. In addition the HTA reported smaller treatment effects for NICU admission and pre-eclampsia. Unlike our analysis, the HTA did not assume 100% uptake of the OGTT and that would have led to a smaller estimate of treatment benefit. We made the simplifying assumption of 100% OGTT uptake because the view of the Guideline Development Group was that uptake would be much higher in a group screened on the basis of risk factors. The HTA also assumed higher uptake of OGTT with risk factor screening compared to universal screening but less than 100%. As we do not find universal screening to be cost-effective then relaxing the assumption of 100% OGTT uptake would only re-inforce that result. We investigated the impact of relaxing the assumption of 100% uptake in groups screened on the basis of risk factors but found that it made a negligible difference to the results. For example, in a deterministic analysis of the HAPO (4) with NICE risk factors, the ICER of NICE 2015 relative to no screening/no treatment only increased from £20,400 per QALY with 100% OGTT uptake to £20,585 per QALY with 90% test uptake.

However, the differences between this analysis and the HTA should not be over-stated. Neither analysis suggests that universal screening for GDM is cost-effective and, like the HTA, our results would not support the identification and treatment of gestational diabetes if a cost-effectiveness threshold of £20,000 per QALY was used. However, it was the view of the Guideline Development Group that the clinical benefit of identifying and treating women with GDM is widely practiced, and that a no identification/no treat policy would not be acceptable to

patients or health care providers. As such, the Group felt that the higher cost threshold of £30,000 was justified.

Our model has a number of limitations particularly with respect to the valuation of health outcomes. We did not include large for gestational age as an outcome because it was felt that shoulder dystocia was the relevant immediate complication of interest, and that possible long term metabolic consequences for the offspring were hard to quantify and therefore difficult to incorporate within the model. As previously noted, the QALY loss from a serious perinatal complication used in this analysis is likely to be overstated because of the relatively large weight given to death based on the intervention studies.[14] HAPO failed to show an association between perinatal mortality and plasma glucose levels, which may mean that perinatal mortality reduction is less amenable to reduction by treatment than other serious perinatal complications. In this respect the cost-effectiveness of diagnosing and treating GDM may be over-stated. On the other hand, the model does not take account of any potential long term effects on the offspring (e.g. adiposity and the likelihood of subsequent pathology) as these effects are difficult to quantify but may under-estimate the QALY gain from diagnosis and treatment. A US study[20] considered the potential long-term benefits to the mother whereby a diagnosis of GDM averts or delays onset of Type 2 diabetes mellitus, but this was not incorporated into our model as we did not consider that the relationship was sufficiently well established at this time. However, to the extent that such a relationship does exist our model would also underestimate the QALY gain from a diagnosis of GDM. A recent review has, however, questioned the association between maternal glycaemia and subsequent cardio-metabolic outcomes in offspring in humans[21] and a recent follow-up study failed to find evidence of a reduction in childhood obesity or metabolic dysfunction at five years in the offspring of women treated for mild gestational diabetes in the study of Landon et al [12, 22].

Despite these caveats, we feel our analysis represents a robust analysis of the cost-effectiveness of the NICE versus the WHO 2013 diagnostic thresholds for GDM based upon our current understanding of the impact of intervention in women with GDM in the UK population. We acknowledge completely that this analysis cannot be the final word on the subject, and that further health economic evaluation is required to either corroborate our findings or to challenge them. Nevertheless, our analysis represents a constructive and evidence based contribution to establishing cost effective diagnostic thresholds for GDM and will hopefully lead to more research to clarify this important but vexed area of clinical diagnosis.

Conclusions

The results presented in this analysis, based on a UK setting, do not suggest that the diagnostic thresholds for GDM adopted by the WHO are cost-effective. On the other hand they do provide some support for the cost-effectiveness of the diagnostic criteria adopted by NICE when compared to either no screening/treatment and to WHO 2013 diagnostic criteria. Furthermore, according to this analysis, universal screening would seem to offer poor value for money and does not appear cost-effective compared to the current NICE guidance of targeting high risk women.

Acknowledgements

We are grateful to Professor DR McCance and Professor HD McIntyre for allowing us to use their local datasets from the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) trial and to Professor F Dunne for allowing us to use her Atlantic DiP dataset.

We are also grateful to Professor David James who provided clinical support during the development of the updated NICE guideline on Diabetes in Pregnancy.

Competing interests

All authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/coi_disclosure.pdf and declare:

Funding: Some of this work was undertaken by the now defunct National Collaborating Centre for Women's and Children's Health (subsumed within the National Guideline Alliance from 1st April 2016), which received funding from NICE. The views expressed in this publication are those of the authors and not necessarily those of the institute. Revisions to the guideline model after the guideline was published and drafting of the manuscript was done in the author's own time and was not funded.

National Institute for Health and Care Excellence (2015). Diabetes in pregnancy: management from preconception to the postnatal period. Available from https://www.nice.org.uk/guidance/ng3

PBJ and SBR are employees of the National Guideline Alliance (part of the RCOG), which receives its funding from NICE.

MJAM, KS, AD and RWB received travel expenses from NICE for attending clinical guideline development meetings

Author contribution

Paul Jacklin designed and developed the health economic model, undertook the health economic analysis, wrote the first draft of the manuscript and incorporated edits from co-authors. Mike Maresh provided clinical input into the design of the health economic model; read, commented and edited various draft of the manuscripts. Katharine Stanley supplied the Norwich dataset, provided clinical input into the design of the health economic model; read, commented and edited various draft of the manuscripts. Anne Dornhorst provided clinical input into the design of the health economic model; read, commented and edited various draft of the manuscripts. Chris Patterson provided statistical advice, undertook statistical analysis of the HAPO dataset; read, commented and edited various drafts of the manuscript. Shona Burman-Roy reviewed the clinical literature, contributed to discussions of model design; read, commented and edited various drafts of the manuscript. Rudy Bilous chaired the NICE guideline, provided clinical input into the design of the health economic model; read, commented and edited various draft of the manuscripts.

Transparency declaration

The lead author, Paul Jacklin, affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Exclusive License

"I Paul Jacklin The Corresponding Author of this article contained within the original manuscript which includes any diagrams & photographs within and any related or stand alone film submitted (the Contribution") has the right to grant on behalf of all authors and does grant

on behalf of all authors, a licence to the BMJ Publishing Group Ltd and its licencees, to permit this Contribution (if accepted) to be published in the BMJ and any other BMJ Group products and to exploit all subsidiary rights, as set out in our licence set out at: http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/copyright-open-access-and-permission-reuse."

Data sharing Statement

Potential for data sharing (the health economic model) can be discussed with study investigators.

References

1. Metzger BE, Lowe LP, Dyer AR, et al. Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med* 2008;358:1991-2002.
2. Farrar D, Simmonds M, Bryant M, et al. Hyperglycaemia and risk of adverse perinatal outcomes: systematic review and meta-analysis. *BMJ*. 2016;354:i4694. doi: 10.1136/bmj.i4694.
3. Metzger BE, Gabbe SG, Persson B, 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.
4. WHO Health organisation, 2013. Diagnostic criteria and classification of hyperglycaemia first detected in pregnancy WHO/NMH/MND/13.2 http://apps.who.int/iris/bitstream/10665/85975/1/WHO_NMH_MND_13.2_eng.pdf.
5. American College of Obstetricians and Gynecologists. *Practice Bulletin 137: Mellitus. Obstet Gynecol* 2013; 122: 406-16
6. National Institute for Health and Care Excellence (NICE) (2015) Diabetes in pregnancy: management of diabetes and its complications from preconception to the postnatal period. Clinical guideline NG3 (2015). Available from www.nice.org.uk/guidance/ng3/resources/diabetes-in-pregnancy-management-of-diabetes-and-itscomplications-from-preconception-to-the-postnatal-period-51038446021, accessed February 2016
7. Meek CL, Lewis HB, Patient C, Murphy HR, Simmons D. Diagnosis of gestational diabetes: falling through the net. *Diabetologia* 2015;Sep;58(9):2003-12
8. Duran A, Sáenz S, Torrejón MJ et al. Introduction of IADPSG criteria for the screening and diagnosis of gestational diabetes mellitus results in improved pregnancy outcomes at a lower cost in a large cohort of pregnant women: the St. Carlos Gestational Diabetes Study. *Diabetes Care* 2014;37(9):2442–2450. doi: 10.2337/dc14-0179.

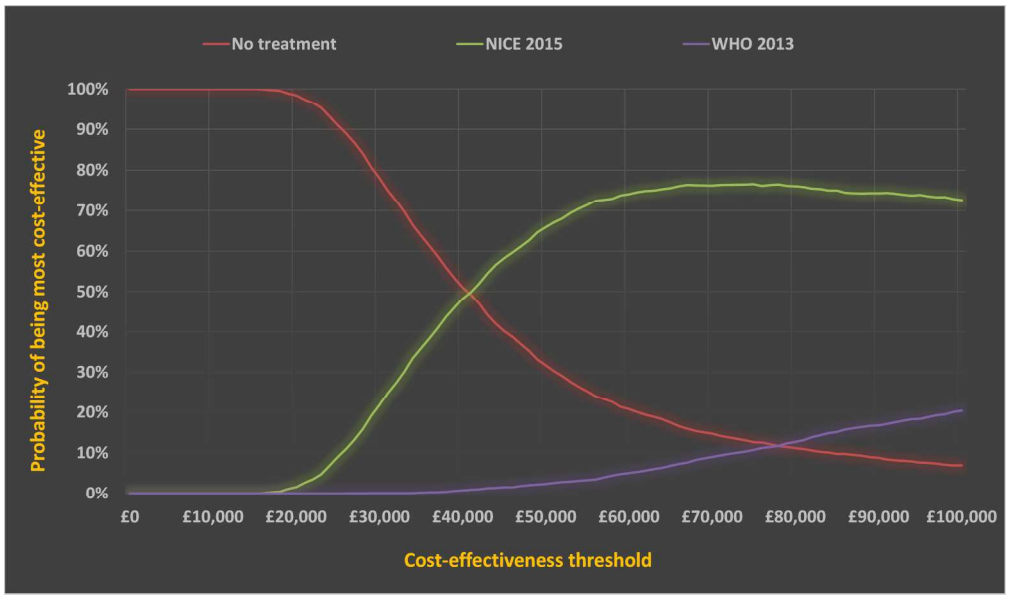
9. Herman WH. Insights offered by economic analyses. *Diabetes Care*. 2014 Sep;37(9):2424-6. doi: 10.2337/dc14-1232.
10. Farrar D, Simmonds M, Griffin S et al. The identification and treatment of women with hyperglycaemia in pregnancy: an analysis of individual participant data, systematic reviews, meta-analyses and an economic evaluation.Health Technol Assess. 2016 Nov;20(86):1-348.
11. 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.
12. Landon MB, Spong CY, Thom E, et al. A multicenter, randomized trial of treatment for mild gestational diabetes. *N Engl J Med* 2009;361:1339-48
13. Horvath K, Koch K, Jeitler K et al. Effects of treatment in women with gestational diabetes mellitus: systematic review and meta-analysis. *BMJ (Clin Res Ed)* 2010;340:c1395
14. Round,J.A., Jacklin,P., Fraser,R.B., Hughes,R.G., Mugglestone,M.A., Holt,R.I. Screening for gestational diabetes mellitus: cost-utility of different screening strategies based on a woman's individual risk of disease, *Diabetologia* 2011,54(2), 256-263. doi: 10.1007/s00125-010-1881-y
15. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. October 2014 (https://www.nice.org.uk/media/default/about/what-we-do/our-programmes/developing-nice-guidelines-the-manual.pdf)
16. Briggs A, Claxton K, Sculpher M. Decision Modelling for Health Economic Evaluation. Oxford: Oxford University Press; 2006
17. Avalos GE, Owens LA, Dunne F. Applying Current Screening Tools for Gestational Diabetes Mellitus to a European Population: Is It Time for Change? *Diabetes Care*. 2013 Oct;36(10):3040-4. doi: 10.2337/dc12-2669. Epub 2013 Jun 11.
18. Simmons D, Moses RG. Gestational Diabetes Mellitus: To Screen or Not to Screen? Is this really still a question? Diabetes Care. 2013 Oct;36(10):2877-2878
19. Armstrong R, Waters E, Doyle J (editors). Chapter 21: Reviews in health promotion and public health. In Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of

Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org. Accessed June 2015

20. Werner EF, Pettker CM, Zucerwise et al. Screening for gestational diabetes mellitus: are the criteria proposed by the international association of the Diabetes and Pregnancy Study Groups cost-effective? Diabetes Care. 2012 Mar;35(3):529-35. doi: 10.2337/dc11-1643.

21. Donovan LE, Cundy T. Does exposure to hyperglycaemia in utero increase the risk of obesity and diabetes in the offspring? A critical reappraisal. Diabetic Medicine. 2015 Mar;32(3):295-304. doi: 10.1111/dme.12625. Epub 2014 Dec 17.

22. Landon MB, Rice MM, Varner MW et al. Mild gestational diabetes mellitus and long-term child health. Diabetes Care. 2015 Mar;38(3):445-52. doi: 10.2337/dc14-2159. Epub 2014 Nov 20.



Cost-effectiveness acceptability curve indicating the probability of a threshold or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for the HAPO (4) centres population with risk factors

267x157mm (300 x 300 DPI)

Supplementary Report

This supplementary document provides further details about model parameter estimates and model results.

Table x1. Ethnicity of women in patient datasets and of UK

Ethnic group	HAPO (4) centres	Atlantic DiP	Norfolk[a]	UK[b]
White	79%	93%	96.5%	87%
Black	2%	2%	0.5%	3%
Asian	13%	4%	1.6%	7%
Other	6%	1%	1.6%	3%

(a) Our Norwich dataset did not include data on ethnicity and the values presented here are census data for Norfolk (Estimated from 2011 Census: Ethnic group, local authorities in the United Kingdom. Office for National Statistics. 11 October 2013)

(b) Included for comparative purposes (2011 Census: Ethnic group, local authorities in the United Kingdom. Office for National Statistics. 11 October 2013)

Multivariable prediction models to estimate baseline risk

Model 1 includes the covariates used in the original analysis of the HAPO data whilst Model 2 is restricted to plasma glucose variables (Tables x2 to Tables x7). In the base case analysis, backward elimination of plasma glucose variables with non-significant coefficients from the prediction models was undertaken. A sensitivity analysis was undertaken retaining all plasma glucose variables. For each model Hosmer-Lemeshow goodness-of-fit statistics are presented and predicted probabilities are used to derive the area under the receiver-operating characteristic (ROC) curve as an indicator of the model's discriminatory ability.

Table x2. Logistic regression models to predict neonatal shoulder dystocia

	Co-efficient b (Standard error (SE(b))			
	Model 1 (all covariates)		Model 2 (blood glucose covariates)	
Variable	**Base case analysis**	**Sensitivity analysis**	**Base case analysis**	**Sensitivity analysis**
Centre (Manchester v Belfast)	1.151 (0.423)	1.151 (0.424)	-	-
(Brisbane v Belfast)	0.505 (0.489)	0.562 (0.491)	-	-
(Newcastle v Belfast)	1.604 (0.472)	1.622 (0.472)	-	-
Age at OGTT (yr)	-0.023 (0.024)	-0.022 (0.024)	-	-
BMI at OGTT (kg/m^2)	-0.006 (0.023)	-0.011 (0.024)	-	-
Smoker (Yes v No)	-0.480 (0.409)	-0.477 (0.409)	-	-
Drinker (Yes v No)	-0.101 (0.317)	-0.107 (0.317)	-	-
Family history DM (Yes v No)	-0.006 (0.184)	-0.008 (0.187)	-	-
Gestational age at OGTT (wk)	-0.111 (0.091)	-0.114 (0.092)	-	-
Neonate gender (F v M)	-1.321 (0.292)	-1.316 (0.292)	-	-
Family history HBP (Yes v No)[a]	-	-	-	-
Maternal UTI (Yes v No)[a]	-	-	-	-
Mean Blood Pressure (mmHg)	-0.006 (0.015)	-0.007 (0.015)	-	-
Hospital admission before delivery (Yes v No)	0.173 (0.266)	0.175 (0.267)	-	-
Parity (1 v 0)	-0.118 (0.420)	-0.108 (0.420)	-	-
(2+ v 0)	0.456 (0.412)	0.469 (0.414)	-	-
(Unknown v 0)	-0.026 (0.399)	-0.013 (0.399)	-	-
Fasting blood glucose[b]	-	0.151 (0.112)	-	0.166 (0.110)
1-hr blood glucose[b]	-	-0.138 (0.165)	-	-0.152 (0.163)
2-hr blood glucose[b]	0.223 (0.100)	0.222 (0.152)	0.267 (0.097)	0.265 (0.151)
Constant	0.925 (3.025)	1.139 (3.508)	-4.467 (0.122)	-4.475 (0.122)
Hosmer-Lemeshow goodness-of-fit test	χ^2 = 2.94, df=8; P=0.94	χ^2 = 6.36, df=8; P=0.61	χ^2 = 4.99, df=8; P=0.76	χ^2 = 11.51, df=8; P=0.18
Area under the ROC curve (95% CI)	0.75 (0.70, 0.80)	0.76 (0.70, 0.81)	0.58 (0.51, 0.65)	0.60 (0.53, 0.67)

(a) Omitted from HAPO model for shoulder dystocia

(b) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for shoulder dystocia arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x3. Logistic regression models to predict caesarean section

	Co-efficient b (Standard error (SE(b))			
	Model 1 (all covariates)		Model 2 (blood glucose covariates)	
Variable	**Base case analysis**	**Sensitivity analysis**	**Base case analysis**	**Sensitivity analysis**
Centre (Manchester v Belfast)	-0.494 (0.092)	-0.495 (0.092)	-	-
(Brisbane v Belfast)	-0.099 (0.098)	-0.114 (0.100)	-	-
(Newcastle v Belfast)	-0.681 (0.140)	-0.692 (0.141)	-	-
Age at OGTT (yr)	0.034 (0.007)	0.034 (0.007)	-	-
BMI at OGTT (kg/m^2)	0.039 (0.007)	0.039 (0.007)	-	-
Smoker (Yes v No)	-0.304 (0.106)	-0.292 (0.106)	-	-
Drinker (Yes v No)	-0.028 (0.087)	-0.025 (0.087)	-	-
Family history DM (Yes v No)	0.050 (0.057)	0.052 (0.057)	-	-
Gestational age at OGTT (wk)	0.004 (0.029)	0.004 (0.029)	-	-
Neonate gender (F v M)	-0.205 (0.071)	-0.205 (0.071)	-	-
Family history HBP (Yes v No)[a]	-	-	-	-
Maternal UTI (Yes v No)[a]	-	-	-	-
Mean Blood Pressure (mmHg)	0.003 (0.004)	0.003 (0.004)	-	-
Hospital admission before delivery (Yes v No)	0.514 (0.079)	0.510 (0.079)	-	-
Parity (1 v 0)[a]	-	-	-	-
(2+ v 0)[a]	-	-	-	-
(Unknown v 0)[a]	-	-	-	-
Fasting blood glucose[b]	-	-0.009 (0.044)	-	0.053 (0.040)
1-hr blood glucose[b]	0.144 (0.037)	0.101 (0.051)	0.138 (0.046)	0.119 (0.048)
2-hr blood glucose[b]	-	0.071 (0.048)	0.123 (0.046)	0.113 (0.046)
Constant	-3.518 (0.947)	-3.509 (0.950)	-1.435 (0.035)	-1.433 (0.035)
Hosmer-Lemeshow goodness-of-fit test	χ^2 = 1.88, df=8; P=0.99	χ^2 = 5.11, df=8; P=0.75	χ^2 = 16.56, df=8; P=0.04	χ^2 = 17.66, df=8; P=0.02
Area under the ROC curve (95% CI)	0.65 (0.63, 0.66)	0.65 (63, 0.66)	0.58 (0.56, 0.60)	0.58 (0.57, 0.60)

(a) Omitted from HAPO model for caesarean section

(b) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for caesarean section arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x4. Logistic regression models to predict neonatal intensive care unit admissions

	Co-efficient b (Standard error (SE(b))			
	Model 1 (all covariates)		Model 2 (blood glucose covariates)	
Variable	**Base case analysis**	**Sensitivity analysis**	**Base case analysis**	**Sensitivity analysis**
Centre (Manchester v Belfast)	0.894 (0.159)	0.889 (0.159)	-	-
(Brisbane v Belfast)	1.393 (0.161)	1.400 (0.163)	-	-
(Newcastle v Belfast)	1.153 (0.190)	1.163 (0.191)	-	-
Age at OGTT (yr)	0.013 (0.009)	0.012 (0.009)	-	-
BMI at OGTT (kg/m^2)	0.025 (0.009)	0.024 (0.009)	-	-
Smoker (Yes v No)	0.209 (0.130)	0.201 (0.130)	-	-
Drinker (Yes v No)	-0.025 (0.117)	-0.023 (0.117)	-	-
Family history DM (Yes v No)	0.033 (0.069)	0.038 (0.069)	-	-
Gestational age at OGTT (wk)	-0.050 (0.038)	-0.052 (0.038)	-	-
Neonate gender (F v M)	-0.304 (0.094)	-0.302 (0.094)	-	-
Family history HBP (Yes v No)[a]	-	-	-	-
Maternal UTI (Yes v No)[a]	-	-	-	-
Mean Blood Pressure (mmHg)	0.006 (0.006)	0.006 (0.006)	-	-
Hospital admission before delivery (Yes v No)	0.794 (0.097)	0.792 (0.097)	-	-
Parity (1 v 0)	-0.474 (0.148)	-0.474 (0.148)	-	-
(2+ v 0)	-0.490 (0.157)	-0.493 (0.157)	-	-
(Unknown v 0)	-0.084 (0.135)	-0.086 (0.135)	-	-
Fasting blood glucose[b]	-	-0.003 (0.054)	-	-0.025 (0.050)
1-hr blood glucose[b]	-	0.082 (0.067)	-	0.078 (0.064)
2-hr blood glucose[b]	0.159 (0.045)	0.107 (0.063)	0.208 (0.041)	0.167 (0.060)
Constant	-3.181 (1.236)	-3.061 (1.243)	-2.374 (0.046)	-2.375 (0.046)
Hosmer-Lemeshow goodness-of-fit test	χ^2 = 14.18, df=8; P=0.08	χ^2 = 11.41, df=8; P=0.18	χ^2 = 22.16, df=8; P=0.005	χ^2 = 12.72, df=8; P=0.12
Area under the ROC curve (95% CI)	0.71 (0.69, 0.73)	0.71 (0.69, 0.73)	0.57 (0.55, 0.60)	0.57 (0.55, 0.60)

(a) Omitted from HAPO model for neonatal intensive care unit admissions

(b) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for neonatal intensive care unit admissions arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x5. Logistic regression models to predict jaundice

	Co-efficient b (Standard error (SE(b))			
	Model 1 (all covariates)		Model 2 (blood glucose covariates)	
Variable	**Base case analysis**	**Sensitivity analysis**	**Base case analysis**	**Sensitivity analysis**
Centre (Manchester v Belfast)	0.407 (0.157)	0.410 (0.157)	-	-
(Brisbane v Belfast)	0.449 (0.171)	0.420 (0.173)	-	-
(Newcastle v Belfast)	-0.315 (0.259)	-0.332 (0.259)	-	-
Age at OGTT (yr)	0.005 (0.011)	0.005 (0.011)	-	-
BMI at OGTT (kg/m^2)	-0.011 (0.011)	-0.009 (0.012)	-	-
Smoker (Yes v No)	0.082 (0.161)	0.093 (0.162)	-	-
Drinker (Yes v No)	-0.514 (0.163)	-0.508 (0.163)	-	-
Family history DM (Yes v No)	-0.060 (0.094)	-0.060 (0.094)	-	-
Gestational age at OGTT (wk)	-0.078 (0.047)	-0.077 (0.047)	-	-
Neonate gender (F v M)	-0.116 (0.113)	-0.115 (0.113)	-	-
Family history HBP (Yes v No)[a]	-	-	-	-
Maternal UTI (Yes v No)[a]	-	-	-	-
Mean Blood Pressure (mmHg)	0.018 (0.007)	0.018 (0.007)	-	-
Hospital admission before delivery (Yes v No)	0.867 (0.116)	0.865 (0.116)	-	-
Parity (1 v 0)	-0.382 (0.185)	-0.380 (0.185)	-	-
(2+ v 0)	-0.526 (0.200)	-0.526 (0.200)	-	-
(Unknown v 0)	0.078 (0.165)	0.078 (0.165)	-	-
Fasting blood glucose[b]	-	-0.055 (0.066)	-	-0.063 (0.061)
1-hr blood glucose[b]	0.216 (0.056)	0.192 (0.079)	0.237 (0.052)	0.199 (0.078)
2-hr blood glucose[b]	-	0.073 (0.074)	-	0.102 (0.072)
Constant	-1.927 (1.522)	-2.014 (1.526)	-2.846 (0.057)	-2.850 (0.057)
Hosmer-Lemeshow goodness-of-fit test	χ^2 = 8.42, df=8; P=0.39	χ^2 = 7.96, df=8; P=0.44	χ^2 = 2.47, df=8; P=0.96	χ^2 = 10.40, df=8; P=0.24
Area under the ROC curve (95% CI)	0.68 (0.65, 0.71)	0.68 (0.65, 0.71)	0.57 (0.54, 0.60)	0.58 (0.55, 0.61)

(a) Omitted from HAPO model for jaundice

(b) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for jaundice arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x6. Logistic regression models to predict pre-eclampsia

	Co-efficient b (Standard error (SE(b))			
	Model 1 (all covariates)		Model 2 (blood glucose covariates)	
Variable	**Base case analysis**	**Sensitivity analysis**	**Base case analysis**	**Sensitivity analysis**
Centre (Manchester v Belfast)	-0.784 (0.192)	-0.800 (0.193)	-	-
(Brisbane v Belfast)	-0.308 (0.200)	-0.277 (0.202)	-	-
(Newcastle v Belfast)	-0.685 (0.278)	-0.667 (0.278)	-	-
Age at OGTT (yr)	-0.009 (0.015)	-0.011 (0.015)	-	-
BMI at OGTT (kg/m^2)	0.101 (0.011)	0.097 (0.012)	-	-
Smoker (Yes v No)	-0.556 (0.245)	-0.569 (0.246)	-	-
Drinker (Yes v No)	-0.170 (0.194)	-0.168 (0.194)	-	-
Family history DM (Yes v No)	-0.004 (0.127)	0.006 (0.127)	-	-
Gestational age at OGTT (wk)	-0.092 (0.059)	-0.096 (0.059)	-	-
Neonate gender (F v M)	0.173 (0.147)	0.174 (0.147)	-	-
Family history HBP (Yes v No)	0.233 (0.150)	0.230 (0.150)	-	-
Maternal UTI (Yes v No)	0.734 (0.211)	0.721 (0.211)	-	-
Mean Blood Pressure (mmHg) [a]	-	-	-	-
Hospital admission before delivery (Yes v No) [a]	-	-	-	-
Parity (1 v 0)	-0.291 (0.240)	-0.292 (0.240)	-	-
(2+ v 0)	-0.701 (0.271)	-0.703 (0.271)	-	-
(Unknown v 0)	0.026 (0.224)	0.023 (0.224)	-	-
Fasting blood glucose[b]	-	0.062 (0.078)	0.201 (0.065)	0.183 (0.068)
1-hr blood glucose[b]	-	0.065 (0.104)	-	0.083 (0.098)
2-hr blood glucose[b]	0.272 (0.067)	0.195 (0.096)	0.196 (0.072)	0.150 (0.090)
Constant	-3.370 (1.842)	-3.107 (1.855)	-3.453 (0.075)	-3.455 (0.075)
Hosmer-Lemeshow goodness-of-fit test	χ^2 = 5.46, df=8; P=0.71	χ^2 = 8.02, df=8; P=0.43	χ^2 = 12.00, df=8; P=0.15	χ^2 = 15.98, df=8; P=0.04
Area under the ROC curve (95% CI)	0.75 (0.72, 0.78)	0.75 (0.72, 0.79)	0.65 (0.61, 0.68)	0.65 (0.61, 0.68)

(a) Omitted from HAPO model for pre-eclampsia

(b) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for pre-eclampsia arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x7. Logistic regression models to predict induction of labour

	Co-efficient b (Standard error (SE(b))		
	Model 1 (all covariates)		Model 2 (blood glucose covariates)
Variable	**Base case analysis**	**Sensitivity analysis**	**Base case analysis**
Centre (Manchester v Belfast)	-0.476 (0.077)	-0.476 (0.077)	-
(Brisbane v Belfast)	-0.337 (0.085)	-0.333 (0.087)	-
(Newcastle v Belfast)	-0.387 (0.109)	-0.384 (0.110)	-
Age at OGTT (yr)	0.006 (0.006)	0.006 (0.006)	-
BMI at OGTT (kg/m^2)	0.039 (0.006)	0.039 (0.006)	-
Smoker (Yes v No)	0.051 (0.082	0.051 (0.082)	-
Drinker (Yes v No)	0.079 (0.072)	0.079 (0.072)	-
Family history DM (Yes v No)	0.016 (0.048)	0.016 (0.048)	-
Gestational age at OGTT (wk)	0.011 (0.024)	0.011 (0.024)	-
Neonate gender (F v M)	-0.038 (0.059)	-0.038 (0.059)	-
Family history HBP (Yes v No)[a]	-	-	-
Maternal UTI (Yes v No)[a]	-	-	-
Mean Blood Pressure (mmHg)	0.008 (0.004)	0.008 (0.004)	-
Hospital admission before delivery (Yes v No)	0.608 (0.066)	0.608 (0.066)	-
Parity (1 v 0)	-0.363 (0.101)	-0.363 (0.101)	-
(2+ v 0)	-0.193 (0.105)	-0.193 (0.105)	-
(Unknown v 0)	0.141 (0.094)	0.141 (0.094)	-
Fasting blood glucose[b]	-	0.009 (0.037)	0.079 (0.033)
1-hr blood glucose[b]	-0.108 (0.041)	-0.111 (0.043)	-0.093 (0.041)
2-hr blood glucose[b]	0.096 (0.041)	0.094 (0.041)	0.100 (0.040)
Constant	-3.050 (0.794)	-3.037 (0.796)	-1.032 (0.029)
Hosmer-Lemeshow goodness-of-fit test	χ^2 = 9.08, df=8; P=0.34	χ^2 = 9.42 df=8; P=0.31	χ^2 = 9.83 df=8; P=0.28
Area under the ROC curve (95% CI)	0.63 (0.61, 0.65)	0.63 (0.61, 0.65)	0.53 (0.51, 0.55)

(a) Omitted from HAPO model for induction of labour

(b) Blood glucose values are 'standardised' – so the exponential of the coefficient represents the odds ratio for induction of labour arising from a 1 Standard Deviation (SD) increase in plasma glucose (fasting plasma glucose mean (SD) = 4.60(0.47); 1-hour plasma glucose mean (SD) = 7.57(1.83); 2-hour plasma glucose mean (SD) = 6.21(1.44)

Table x8. Cholesky decomposition of shoulder dystocia variance covariance matrix (Model 1, base case)

	Constant	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Age at OGTT (yr)	BMI AT OGTT (kg/m2)	Smoker	Drinker	Family History DM	Gestational age at OGTT (wk)	Neonatal gender	Mean Blood Pressure	Parity (1 v 0)	Parity (2+ v 0)	Parity (Unknown v 0)	Hospital admission before	2-hr blood glucose
Constant	3.025																
Centre (Manchester v Belfast)	-0.104	0.410															
(Brisbane v Belfast)	-0.129	0.335	0.331														
(Newcastle v Belfast)	-0.135	0.334	0.074	0.295													
Age at OGTT (yr)	-0.005	-0.001	0.000	0.000	0.024												
BMI AT OGTT (kg/m2)	-0.001	-0.001	-0.001	-0.001	0.001	0.023											
Smoker	-0.012	0.031	0.001	-0.003	0.047	0.006	0.404										
Drinker	0.001	0.014	-0.006	-0.020	-0.051	0.008	-0.018	0.311									
Family History DM	-0.023	0.012	-0.022	-0.018	-0.005	0.000	-0.009	-0.012	0.179								
Gestational age at OGTT (wk)	-0.080	-0.009	-0.001	0.006	-0.017	-0.007	-0.002	-0.001	-0.008	0.037							
Neonatal gender (F v M)	-0.037	-0.010	-0.006	-0.010	-0.002	0.009	0.003	-0.001	-0.005	-0.080	0.278						
Mean Blood Pressure (mmHg)	-0.004	-0.001	0.000	-0.002	-0.002	-0.006	0.000	0.000	-0.001	-0.011	-0.004	0.004					
Parity (1 v 0)	-0.038	0.014	0.011	-0.029	-0.024	-0.017	0.009	-0.013	0.006	-0.084	-0.020	-0.318	0.253				
(2+ v 0)	-0.026	0.007	0.025	-0.018	-0.082	-0.043	0.000	0.006	0.011	-0.072	-0.027	-0.317	0.021	0.229			
(Unknown v 0)	-0.052	0.019	0.005	-0.020	0.028	0.004	0.015	0.001	0.006	-0.090	-0.026	-0.311	0.020	0.026	0.219		
Hospital admission before delivery	-0.008	-0.007	-0.016	-0.006	0.005	-0.033	-0.006	0.007	-0.002	-0.002	0.004	-0.063	-0.061	-0.071	-0.079	0.225	
2-hr blood glucose	0.012	-0.003	-0.003	0.005	-0.016	-0.015	0.009	0.006	0.013	0.004	0.004	-0.003	-0.007	-0.021	-0.002	-0.015	0.091

69 **Table x9.** Cholesky decomposition of caesarean section variance covariance matrix (Model 1, base case)

	Constant	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Age at OGTT (yr)	BMI AT OGTT (kg/m2)	Smoker	Drinker	Family History DM	Gestational age at OGTT	Neonatal gender	Mean Blood Pressure	Hospital admission	1-hr blood glucose
Constant	0.947													
Centre (Manchester v Belfast)	-0.020	0.089												
(Brisbane v Belfast)	-0.028	0.047	0.082											
(Newcastle v Belfast)	-0.028	0.048	0.025	0.126										
Age at OGTT (yr)	-0.001	0.000	0.000	0.000	0.007									
BMI AT OGTT (kg/m^2)	-0.001	-0.001	0.000	0.000	0.000	0.007								
Smoker	-0.004	0.010	0.004	0.001	0.013	0.000	0.104							
Drinker	0.001	0.004	0.000	-0.002	-0.011	0.004	-0.006	0.086						
Family History DM	-0.006	0.006	-0.005	-0.002	-0.003	0.000	-0.003	-0.003	0.056					
Gestational age at OGTT (wk)	-0.026	-0.001	0.000	0.001	-0.005	-0.002	-0.001	-0.001	-0.003	0.011				
Neonate gender	-0.009	-0.001	-0.002	0.000	-0.002	0.002	0.002	-0.001	-0.001	-0.020	0.067			
Mean Blood Pressure (mmHg)	-0.001	0.000	0.000	0.000	-0.001	-0.002	0.000	0.000	0.000	-0.003	-0.001	0.001		
Hospital admission before delivery	-0.003	-0.004	-0.007	-0.001	0.004	-0.009	-0.003	0.004	-0.001	0.000	0.000	-0.042	0.065	
1-hr blood glucose	0.005	-0.001	0.002	0.002	-0.006	-0.006	0.000	0.001	0.005	0.003	0.002	-0.003	-0.003	0.035

70

71

Table x10. Cholesky decomposition of neonatal intensive care admission variance covariance matrix (Model 1, base case)

	Constant	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Age at OGTT (yr)	BMI AT OGTT (kg/m2)	Smoker	Drinker	Family History DM	Gestational age at OGTT (wk)	Neonatal gender	Mean Blood Pressure	Parity (1 v 0)	Parity (2+ v 0)	Parity (Unknown v 0)	Hospital admission	2-hr blood glucose
Constant	1.236																
Centre (Manchester v Belfast)	-0.037	0.154															
(Brisbane v Belfast)	-0.047	0.115	0.102														
(Newcastle v Belfast)	-0.050	0.115	0.042	0.137													
Age at OGTT (yr)	-0.002	0.000	0.000	0.000	0.009												
BMI AT OGTT (kg/m2)	-0.001	0.000	0.000	0.000	0.000	0.009											
Smoker	-0.007	0.012	0.002	-0.001	0.018	-0.001	0.128										
Drinker	0.002	0.004	-0.002	-0.004	-0.017	0.006	-0.007	0.115									
Family History DM	-0.008	0.004	-0.008	-0.003	-0.003	0.000	-0.004	-0.003	0.068								
Gestational age at OGTT (wk)	-0.034	-0.003	-0.001	0.002	-0.007	-0.002	-0.001	-0.001	-0.004	0.015							
Neonatal gender (F v M)	-0.012	-0.003	-0.004	-0.002	-0.002	0.003	0.000	-0.001	-0.001	-0.025	0.090						
Mean Blood Pressure (mmHg)	-0.002	0.000	0.000	-0.001	-0.001	-0.002	0.000	0.000	0.000	-0.004	-0.001	0.001					
Parity (1 v 0)	-0.011	0.006	0.004	-0.009	-0.009	-0.005	0.002	-0.003	0.002	-0.024	-0.004	-0.103	0.102				
(2+ v 0)	-0.006	0.002	0.011	-0.005	-0.032	-0.015	-0.002	0.003	0.003	-0.021	-0.006	-0.102	0.012	0.111			
(Unknown v 0)	-0.016	0.008	0.003	-0.005	0.010	0.002	0.006	0.001	0.002	-0.025	-0.006	-0.101	0.011	0.011	0.081		
Hospital admission before delivery	-0.005	-0.001	-0.005	0.000	0.004	-0.012	-0.003	0.005	0.000	-0.001	0.000	-0.033	-0.028	-0.025	-0.032	0.075	
2-hr blood glucose	0.004	0.000	0.000	0.001	-0.007	-0.006	0.004	0.002	0.005	0.003	0.002	-0.002	-0.002	-0.006	0.000	-0.006	0.042

Table x11. Cholesky decomposition of jaundice variance covariance matrix (Model 1, base case)

	Constant	Centre (Manchester v Belfast)	Centre (Brisbane v Belfast)	Centre (Newcastle v Belfast)	Age at OGTT (yr)	BMI AT OGTT (kg/m2)	Smoker	Drinker	Family History DM	Gestational age at OGTT (wk)	Neonatal gender	Mean Blood Pressure	Parity (1 v 0)	Parity (2+ v 0)	Parity (Unknown v 0)	Hospital admission	1-hr blood glucose
Constant	1.522																
Centre (Manchester v Belfast)	-0.038	0.153															
(Brisbane v Belfast)	-0.049	0.102	0.128														
(Newcastle v Belfast)	-0.055	0.102	0.041	0.228													
Age at OGTT (yr)	-0.002	0.000	0.000	0.000	0.011												
BMI AT OGTT (kg/m2)	-0.001	-0.001	-0.001	0.000	0.000	0.011											
Smoker	-0.011	0.017	0.003	-0.001	0.023	0.000	0.158										
Drinker	0.003	0.004	0.000	-0.002	-0.020	0.006	-0.011	0.161									
Family History DM	-0.011	0.010	-0.010	-0.003	-0.005	-0.001	-0.004	-0.005	0.092								
Gestational age at OGTT (wk)	-0.042	-0.003	0.000	0.002	-0.008	-0.004	-0.001	-0.001	-0.005	0.018							
Neonatal gender (F v M)	-0.014	-0.003	-0.003	-0.002	-0.002	0.004	0.002	-0.001	0.000	-0.031	0.108						
Mean Blood Pressure (mmHg)	-0.002	0.000	0.000	0.000	-0.001	-0.003	0.000	0.000	0.000	-0.005	-0.002	0.002					
Parity (1 v 0)	-0.016	0.010	0.008	-0.006	-0.011	-0.007	0.001	-0.001	0.003	-0.032	-0.007	-0.128	0.126				
(2+ v 0)	-0.008	0.002	0.014	-0.004	-0.039	-0.016	-0.003	0.005	0.006	-0.030	-0.009	-0.127	0.016	0.144			
(Unknown v 0)	-0.021	0.013	0.006	-0.004	0.012	0.003	0.005	0.002	0.002	-0.034	-0.008	-0.126	0.015	0.013	0.094		
Hospital admission before delivery	-0.006	-0.003	-0.008	-0.002	0.005	-0.016	-0.004	0.003	-0.001	-0.002	0.001	-0.040	-0.034	-0.028	-0.042	0.089	
1-hr blood glucose	0.007	-0.001	0.004	0.003	-0.008	-0.009	0.003	0.001	0.008	0.004	0.003	-0.004	-0.004	-0.009	-0.001	-0.008	0.051

Table x12. Cholesky decomposition of pre-clampsia variance covariance matrix (Model 1, base case)

	Constant	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Age at OGTT (yr)	BMI AT OGTT (kg/m2)	Smoker	Drinker	Family History DM	Gestational age at OGTT (wk)	Neonatal gender (F v M)	Parity (1 v 0)	Parity (2+ v 0)	Parity (Unknown v 0)	Family History HBP	Maternal UTI	2-hr blood glucose
Constant	1.842																
Centre (Manchester v Belfast)	-0.045	0.187															
(Brisbane v Belfast)	-0.059	0.083	0.173														
(Newcastle v Belfast)	-0.072	0.085	0.053	0.249													
Age at OGTT (yr)	-0.004	0.000	0.000	0.000	0.015												
BMI AT OGTT (kg/m2)	-0.002	-0.001	-0.001	-0.002	0.000	0.010											
Smoker	-0.010	0.019	0.002	0.001	0.026	0.001	0.243										
Drinker	0.002	0.007	-0.002	-0.004	-0.023	0.007	-0.009	0.192									
Family History DM	-0.017	0.013	-0.009	-0.006	-0.009	-0.001	-0.004	-0.004	0.124								
Gestational age at OGTT (wk)	-0.054	-0.003	0.000	0.002	-0.013	-0.012	-0.001	-0.001	-0.008	0.011							
Neonatal gender (F v M)	-0.023	-0.004	-0.006	-0.007	-0.003	0.009	0.001	-0.002	-0.001	-0.106	0.099						
Parity (1 v 0)	-0.020	0.013	0.016	-0.006	-0.014	-0.011	0.003	-0.001	0.001	-0.121	-0.123	0.163					
(2+ v 0)	-0.011	-0.001	0.020	-0.005	-0.047	-0.030	0.000	0.000	0.007	-0.117	-0.120	0.042	0.199				
(Unknown v 0)	-0.029	0.014	0.013	-0.004	0.017	0.002	0.006	0.001	0.000	-0.122	-0.125	0.036	0.025	0.127			
Family History HBP	-0.009	0.000	0.006	-0.005	-0.017	-0.014	0.001	-0.001	0.023	-0.034	-0.037	-0.060	-0.038	-0.048	0.108		
Maternal UTI	-0.004	-0.012	0.031	0.002	0.015	-0.001	-0.014	0.004	0.009	-0.021	-0.022	-0.030	-0.018	-0.028	-0.052	0.193	
2-hr blood glucose	0.006	-0.004	0.002	0.001	-0.012	-0.009	0.005	0.004	0.008	-0.006	-0.004	-0.006	-0.011	-0.002	-0.012	-0.009	0.061

78 **Table x13.** Cholesky decomposition of induction of labour variance covariance matrix (Model 1, base case)

	Constant	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Centre (Manchester v Belfast)	Age at OGTT (yr)	BMI AT OGTT (kg/m2)	Smoker	Drinker	Family History DM	Gestational age at OGTT	Neonatal gender (F v M)	Mean Blood Pressure	Parity (1 v 0)	Parity (2+ v 0)	Parity (Unknown v 0)	Hospital admission	1-hr blood glucose	2-hr blood glucose
Constant	0.794																	
Centre (Manchester v Belfast)	-0.018	0.074																
(Brisbane v Belfast)	-0.024	0.039	0.072															
(Newcastle v Belfast)	-0.024	0.039	0.020	0.097														
Age at OGTT (yr)	-0.001	0.000	0.000	0.000	0.006													
BMI AT OGTT (kg/m2)	0.000	0.000	0.000	0.000	0.000	0.006												
Smoker	-0.005	0.008	0.002	0.001	0.013	0.001	0.081											
Drinker	0.001	0.003	0.000	-0.002	-0.010	0.003	-0.006	0.071										
Family History DM	-0.006	0.005	-0.004	-0.002	-0.002	0.000	-0.002	-0.003	0.047									
Gestational age at OGTT (wk)	-0.021	-0.001	0.000	0.001	-0.004	-0.002	-0.001	0.000	-0.003	0.010								
Neonatal gender (F v M)	-0.007	-0.002	-0.002	-0.001	-0.001	0.002	0.001	-0.001	-0.001	-0.016	0.056							
Mean Blood Pressure (mmHg)	-0.001	0.000	0.000	0.000	-0.001	-0.001	0.000	0.000	0.000	-0.003	-0.001	0.001						
Parity (1 v 0)	-0.008	0.008	0.005	-0.004	-0.006	-0.004	0.000	-0.002	0.001	-0.020	-0.005	-0.075	0.061					
(2+ v 0)	-0.004	0.004	0.008	-0.002	-0.020	-0.010	-0.002	0.001	0.002	-0.019	-0.006	-0.075	0.005	0.067				
(Unknown v 0)	-0.012	0.007	0.003	-0.002	0.006	0.001	0.002	0.001	0.001	-0.021	-0.005	-0.075	0.005	0.004	0.050			
Hospital admission before delivery	-0.004	-0.004	-0.005	-0.001	0.002	-0.007	-0.002	0.003	0.000	-0.001	0.000	-0.017	-0.017	-0.016	-0.020	0.055		
1-hr blood glucose	0.004	0.000	0.004	0.002	-0.004	-0.003	-0.001	0.000	0.003	0.001	0.001	0.000	0.001	-0.001	0.001	0.000	0.040	
2-hr blood glucose	0.001	0.000	-0.003	-0.001	-0.001	-0.002	0.004	0.001	0.001	0.002	0.000	-0.001	0.000	-0.002	0.001	-0.001	-0.027	0.030

79

Table x14: Model unit costs

Category	Cost	Standard Error	Distribution[a]	Source
2 sample OGTT	£8.07	n/a	n/a	NICE 2015[b]
3 sample OGTT	£12.11	n/a	n/a	NICE 2015[b]
Rapilose OGTT solution	£3.48	n/a	n/a	BNF July 2016[c]
Health Care Assistant Band 3 (per hour)	£25	n/a	n/a	Unit Costs of Health and Social Care 2015[d]
Nurse Band 7 (per hour of patient contact)	£147	n/a	n/a	Unit Costs of Health and Social Care 2015[d]
Dietician	£38	n/a	n/a	Unit Costs of Health and Social Care 2015[d]
Antenatal appointment	£96	£9.07	Normal	NHS Reference Costs 2014-15[e]
Ultrasound scan	£112	£7.65	Normal	NHS Reference Costs 2014-15[e]
Rapid acting insulin	£0.02	n/a	n/a	BNF June 2016[c]
Regular insulin	£0.02	n/a	n/a	BNF June 2016[c]
Needles	£0.10	n/a	n/a	NHS Drugs Tariff June 2016[f]
Lancets	£0.03	n/a	n/a	NHS Drugs Tariff June 2016[f]
Strips	£0.18	n/a	n/a	NHS Drugs Tariff June 2016[f]
Treatment of GDM	£987	n/a	n/a	Calculated
Severe hypoglycaemia	£650	n/a	n/a	NHS Reference Costs 2014-15[e]
Admission to NICU	£1,176	£38	Normal	NHS Reference Costs 2014-15[e]
Caesarean section	£982	£80	Normal	NHS Reference Costs 2014-15[e]
Neonatal death	£777	£39	Normal	NHS Reference Costs 2014-15[e]
Shoulder dystocia	£1,394	£79	Normal	NHS Reference Costs 2014-15[e]
Birth trauma	£1,394	£79	Normal	NHS Reference Costs 2014-15[e]
Serious perinatal complication (death, shoulder dystocia, birth trauma)	£1,347	n/a	n/a	Calculated
Phototherapy	£788	£72	Normal	NHS Reference Costs 2014-15[e]
Pre-eclampsia	£4,750	n/a	n/a	NICE 2015[b]

(a) The method used to obtain standard errors and the choice of a normal distribution for probabilistic sampling is described in detail in the NICE 2015 guideline[6]

(b) National Institute for Health and Care Excellence (NICE) (2015) Diabetes in pregnancy: management of diabetes and its complications from preconception to the postnatal period. Clinical guideline NG3 (2015).
(c) British National Formulary. July 2016. https://www.medicinescomplete.com/mc/bnf/current/ (accessed 4 Aug 2016).
(d) Unit Costs of Health and Social Care 2015. Personal Social Services Research Unit, The University of Kent, 2015.
(e) Department of Health. NHS reference costs: financial year 2014–2015. https://www.gov.uk/government/publications/nhs-reference-costs-2014-to-2015 , Department of Health, 2015.
(f) NHS Electronic Drug Tariff, August 2016. http://www.drugtariff.nhsbsa.nhs.uk/#/00336026-DD_1/DD00336022/Home (accessed 4 Aug 2016).

QALYs

A QALY loss was estimated for each individual component (shoulder dystocia, death and birth trauma) of the composite serious perinatal outcome, which was used in the ACHOIS study.[11] A weighting for each individual component was derived according to their relative frequency in the selected studies to assess treatment effectiveness.[11, 12] These were then used in order to derive a weighted average for a serious perinatal complication as shown in Table x15. QALY losses from a serious perinatal complication could be experienced over a lifetime and therefore an annual discount rate of 3.5% was applied in line with NICE methods.[19] For each patient, an expected QALY decrement is calculated based on their risk of serious perinatal complications. These individual patient QALY decrements are then summed across all patients to give the total QALY decrement for the patient dataset for each different diagnostic threshold.

Table x15: QALY losses and weights from individual components of the composite outcome of serious perinatal complications

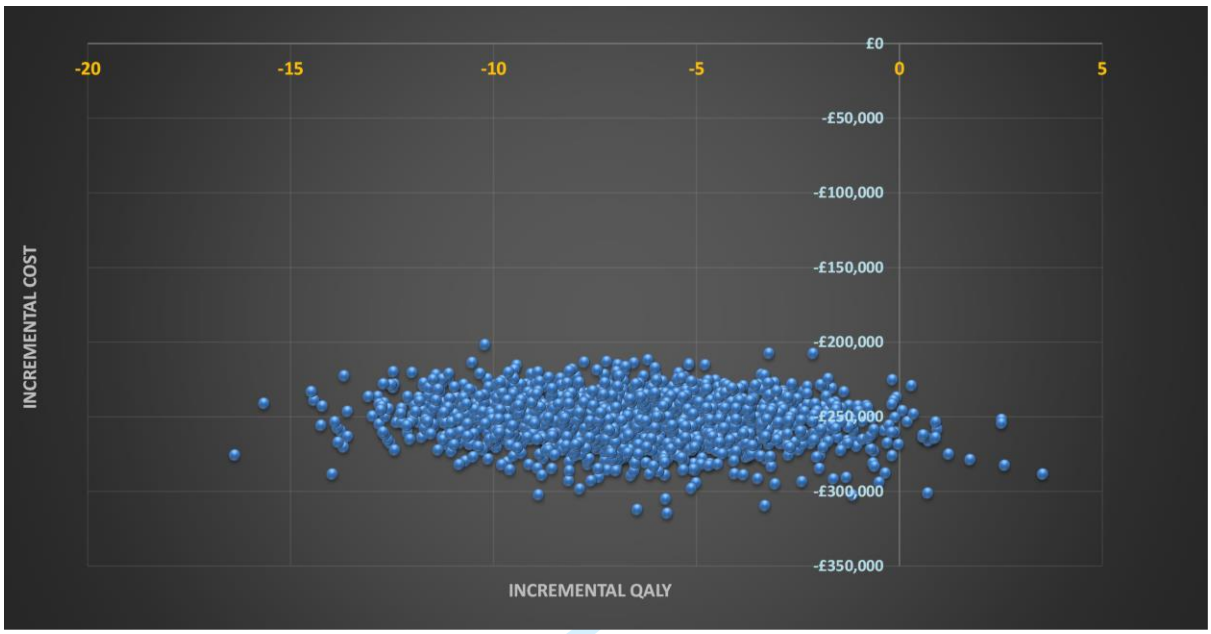
Complication	Weight	QALY	Weighted QALY
Death	0.08	25	2.00
Shoulder dystocia	0.73	0.2	0.15
Birth trauma	0.20	0.2	0.04

The analyses presented in this paper include a maternal health state utility which was estimated from quality of life data collected as part of the ACHOIS study. Whilst treatment conferred a small benefit in maternal health state utility, this was small in comparison to QALYs derived from infant outcomes. The

value of the maternal health state utility with and without treatment is the same as has been used previously.[6]

Results for the HAPO (4) population with risk factors

Figure x1: Cost-effectiveness plane for NICE 2015 compared with WHO 2013 for HAPO (4) with risk factors



Summary of results for each model population

Table x16: Summary of deterministic ICERs for each population with backward elimination of plasma glucose variables with non-significant coefficients

	All covariates		Plasma glucose covariates				
Diagnostic threshold	**HAPO Risk factor (n=3,549)**	**HAPO No Risk factor (n=2,614)**	**HAPO Risk factor (n=3,549)**	**HAPO No Risk factor (n=2,614)**	**Atlantic DiP Risk factor (n=1,988)**	**Atlantic DiP No Risk factor (n=3,302)**	**Norwich (n=12,754)**
No Treatment	-	-	-	-	-	-	-
NICE 2015	£20,400	£36,878	£22,281	£30,449	£20,830	£31,136	£28,893
WHO 2013	£33,596	£141,812	£36,473	£88,661	£35,941	£40,526	£37,918

Table x17: Probability that a threshold is cost-effective at a threshold of £30,000 per QALY and the net monetary benefit in each population using regression models with backward elimination of plasma glucose variables with non-significant coefficients

	All covariates		Plasma glucose covariates				
Diagnostic threshold	**HAPO Risk factor (NMB)**	**HAPO No Risk factor (NMB)**	**HAPO Risk factor (NMB)**	**HAPO No Risk factor (NMB)**	**Atlantic DiP Risk factor (NMB)**	**Atlantic DiP No Risk factor (NMB)**	**Norwich (NMB)**
No Treatment	21.0% (£0)	78.1% (£0)	33.7% (£0)	69.3% (£0)	30.6% (£0)	70.0% (£0)	61.2% (£0)
NICE 2015	51.5% (£239,902)	21.9% (-£57,790)	53.2% (£104,075)	30.7% (£36,652)	54.6% (£113,042)	23.5% (-£37,716)	29.3% (-£96,248)
WHO 2013	27.6% (£186,675)	0.1% (-£111,179)	13.2% (£13,836)	0.1% (£79,581)	14.9% (£36,377)	6.6% (-£109,809)	9.6% (-£414,428)

Results for the HAPO (4) population without risk factors

Table x18: Clinical outcomes for HAPO (4) population without NICE risk factors (n=2,614)

Diagnostic threshold	Diagnosed	SD	SPC	CS	NICU	Jaund	PE	IOL
No Treatment	0	24	34	466	188	126	55	647
NICE 2015	208	23	31	460	184	124	51	655
WHO 2013	253	23	31	459	184	123	51	657

Table x19: Deterministic analysis for HAPO (4) population without NICE risk factors (n=2,614)

Diagnostic threshold	Cost[a]	QALY[a]	Incremental cost	Incremental QALY	ICER
No Treatment	£0	0.00	n/a	n/a	n/a
NICE 2015	£238,074	6.46	£238,074	6.46	£36,878
WHO 2013	£297,364	6.87	£59,290	0.41	£141,812

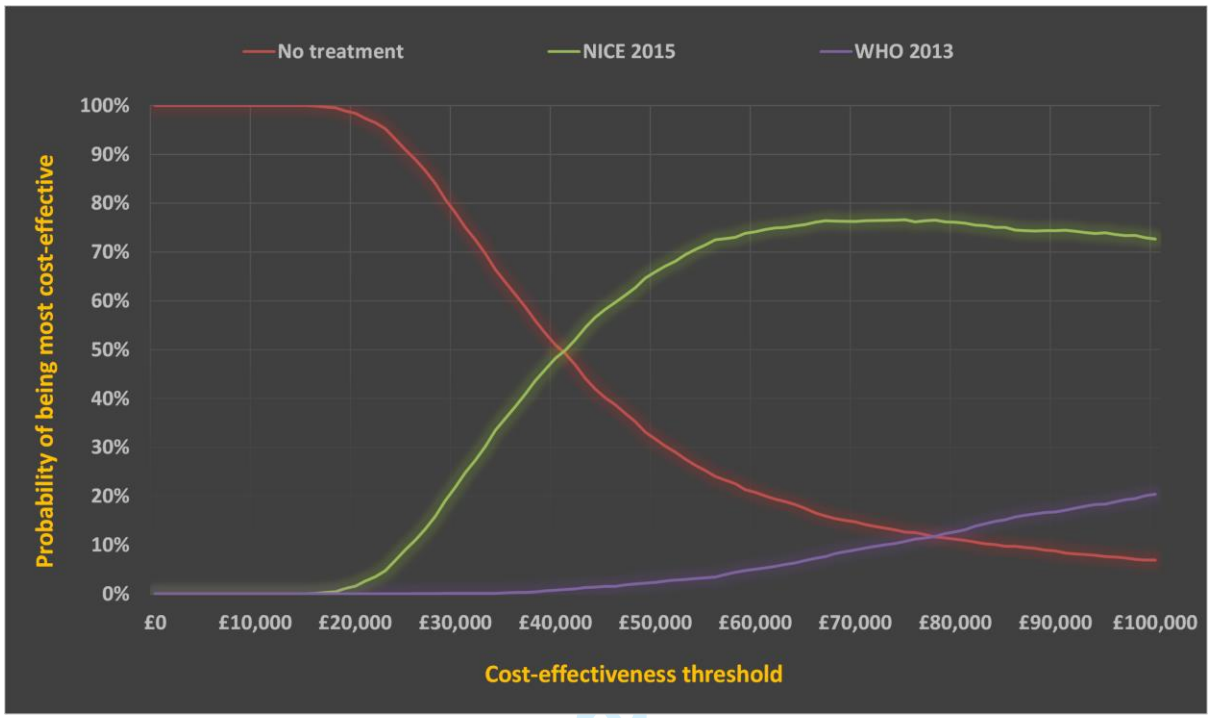
a) *Costs and QALYs are measured relative to a baseline of No Treatment*

Table x20: Probabilistic sensitivity analysis for HAPO (4) in a population without NICE risk factors

Diagnostic threshold	NMB[a] CE threshold £30,000 per QALY	Probability cost-effective CE threshold £30,000 per QALY
No Treatment	£0	78.1%
NICE 2015	-£57,790	21.9%
WHO 2013	-£111,179	0.1%

a) NMB is measured relative to a baseline of no treatment

Figure x2: Cost-effectiveness acceptability curve indicating the probability of a threshold or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for HAPO (4) population without risk factors



Results for the Atlantic DiP population with risk factors

Table 21: Clinical outcomes for Atlantic DiP population with NICE risk factors (n=1,988)

Diagnostic threshold	Diagnosed	SD	SPC	CS	NICU	Jaund	PE	IOL
No Treatment	0	25	34	408	177	122	73	522
NICE 2015	497	19	26	391	163	116	56	545
WHO 2013	749	17	24	385	158	112	51	555

Table x22: Deterministic analysis for the Atlantic DiP population with NICE risk factors (n=1,988)

Diagnostic threshold	Cost[a]	QALY[a]	Incremental cost	Incremental QALY	ICER
No Treatment	£0	0.00	n/a	n/a	n/a
NICE 2015	£414,714	19.91	£414,714	17.46	£20,830
WHO 2013	£638,590	26.14	£223,876	6.23	£35,941

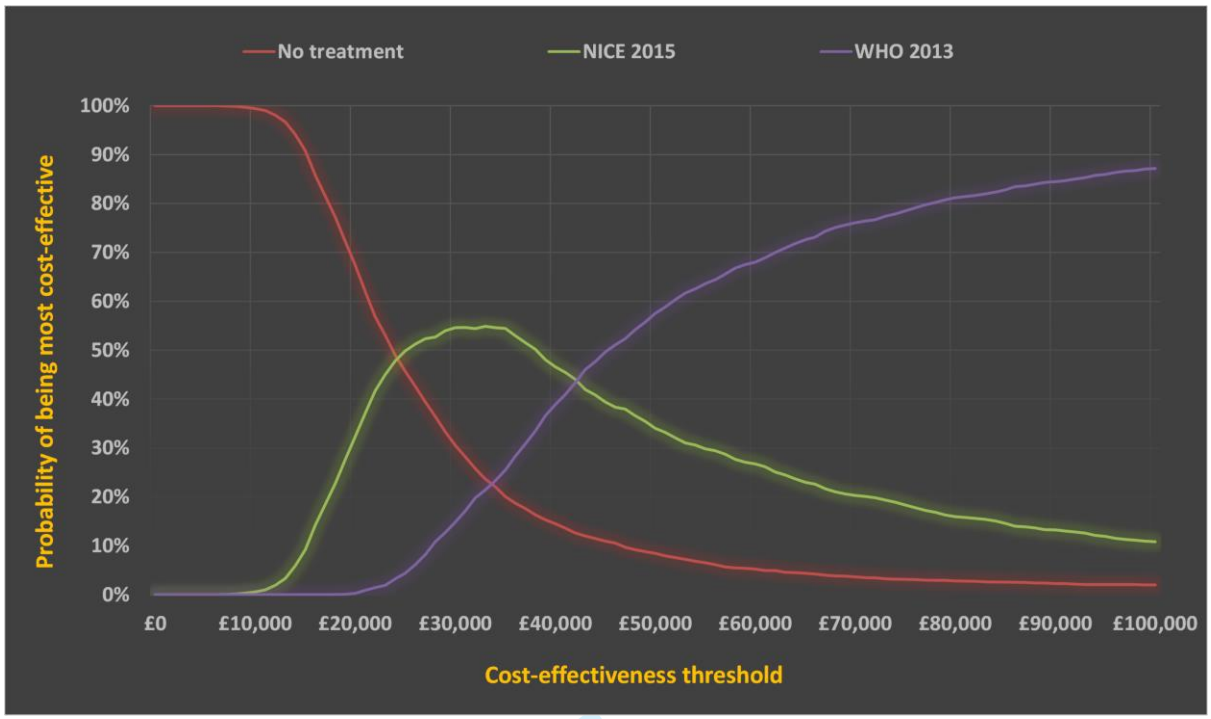
a) *Costs and QALYs are measured relative to a baseline of No Treatment*

Table x23: Probabilistic sensitivity analysis for Atlantic in a population with NICE risk factors

Diagnostic threshold	NMB[a] CE threshold £30,000 per QALY	Probability cost-effective CE threshold £30,000 per QALY
No Treatment	£0	30.6%
NICE 2015	£113,042	54.3%
WHO 2013	£36,377	14.9%

a) NMB is measured relative to a baseline of no treatment

Figure x3: Cost-effectiveness acceptability curve indicating the probability of a threshold or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for the Atlantic DiP centres population with risk factors



Results for the Atlantic DiP population without risk factors

Table x24: Clinical outcomes for Atlantic DiP population without NICE risk factors (n=3,302)

Diagnostic threshold	Diagnosed	SD	SPC	CS	NICU	Jaund	PE	IOL
No Treatment	0	33	45	575	254	168	84	828
NICE 2015	194	31	42	569	248	166	79	837
WHO 2013	371	30	41	564	245	163	76	844

Table x25: Deterministic analysis for the Atlantic DiP population without NICE risk factors (n=3,302)

Diagnostic threshold	Cost[a]	QALY[a]	Incremental cost	Incremental QALY	ICER
No Treatment	£0	0.00	n/a	n/a	n/a
NICE 2015	£231,633	7.44	£231,633	7.44	£31,136
WHO 2013	£402,014	11.64	£170,381	4.20	£40,526

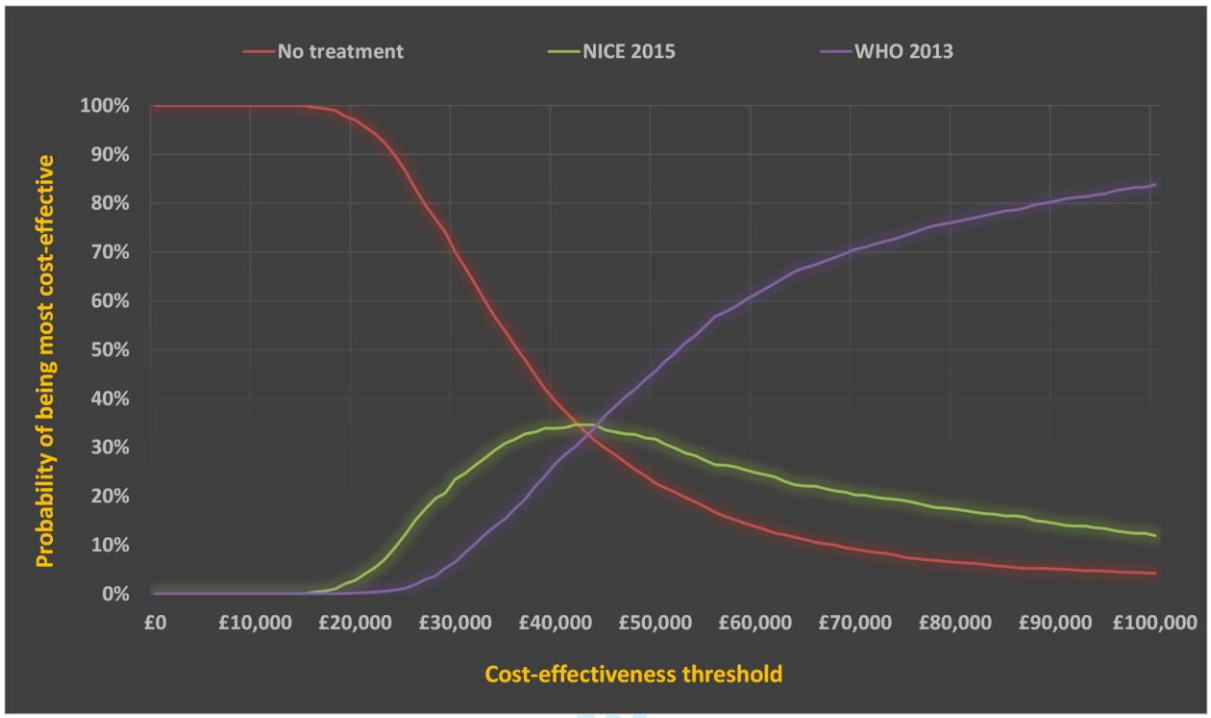
a) *Costs and QALYs are measured relative to a baseline of No Treatment*

Table x26: Probabilistic sensitivity analysis for the Atlantic DiP population without NICE risk factors

Diagnostic threshold	NMB[a] CE threshold £30,000 per QALY	Probability cost-effective CE threshold £30,000 per QALY
No Treatment	£0	70.0%
NICE 2015	-£37,716	23.5%
WHO 2013	-£109,809	6.6%

a) NMB is measured relative to a baseline of no treatment

Figure x4: Cost-effectiveness acceptability curve indicating the probability of a threshold or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for the Atlantic DiP centres population without risk factors



Results for the Norwich population

Table x27: Clinical outcomes for Norwich population (n=12,754)

Diagnostic threshold	Diagnosed	SD	SPC	CS	NICU	Jaund	PE	IOL
No Treatment	0	132	182	2,333	1,005	699	346	3,173
NICE 2015	888	122	168	2,305	981	687	318	3,214
WHO 2013	1,771	117	161	2,283	965	676	301	3,248

Table x28: Deterministic analysis for the Norwich population (n=12,754)

Diagnostic threshold	Cost[a]	QALY[a]	Incremental cost	Incremental QALY	ICER
No Treatment	£0	0.00	n/a	n/a	n/a
NICE 2015	£979,903	33.91	£979,903	33.91	£28,893
WHO 2013	£1,803,196	55.63	£823,293	21.72	£37,918

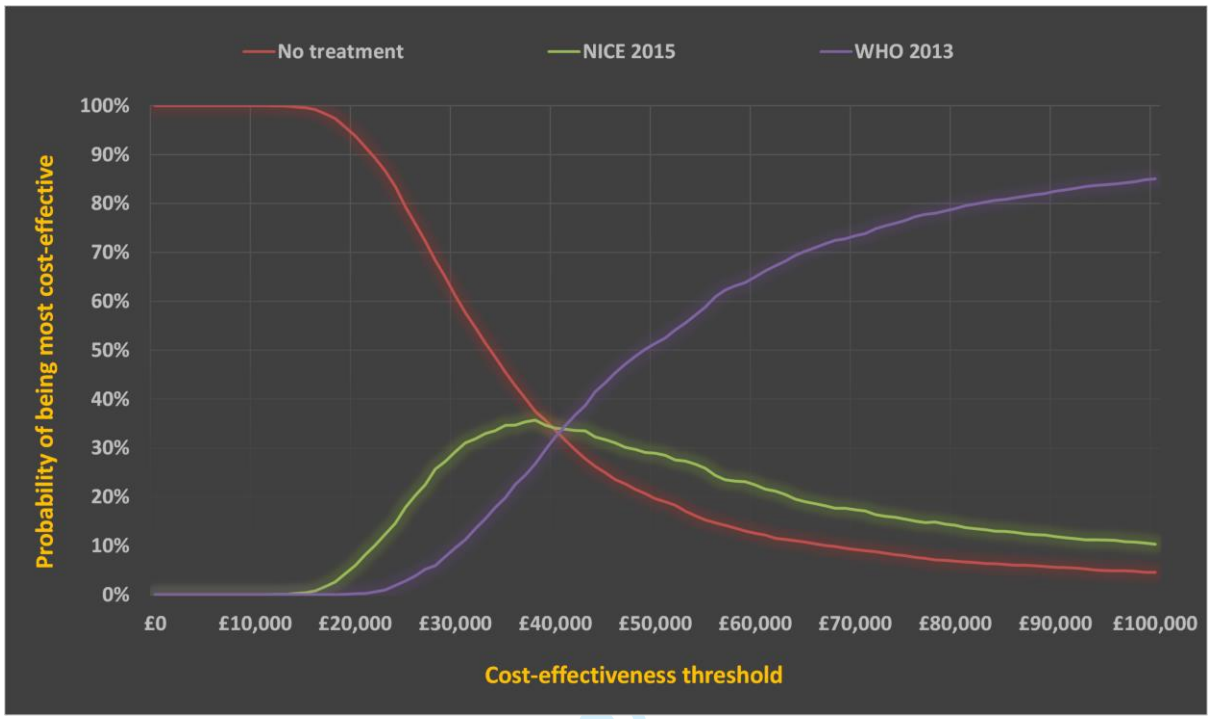
a) *Costs and QALYs are measured relative to a baseline of No Treatment*

Table x29: Probabilistic sensitivity analysis for the Norwich population

Diagnostic threshold	NMB[a] CE threshold £30,000 per QALY	Probability cost-effective CE threshold £30,000 per QALY
No Treatment	£0	61.2%
NICE 2015	-£96,248	29.3%
WHO 2013	-£414,428	9.6%

a) NMB is measured relative to a baseline of no treatment

Figure x5: Cost-effectiveness acceptability curve indicating the probability of a threshold or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for the Norwich population



Deterministic sensitivity analysis

The cost-effectiveness of universal screening was not generally affected when the model was re-run using the regression models without backward elimination of non-significant variables with no screening/no treatment continuing to be the cost-effective option in populations not selected on the basis of NICE risk factors (see Table x30). In the Norwich population, universal screening was borderline cost-effective compared to no screening/no treatment at £30,000 per QALY but the same point remains that a risk factor subset in this population would have a lower ICER than that reported, and that a subset without risk factors, (i.e. those additionally incorporated as a result of universal screening compared to risk factor screening), would have a higher ICER. In populations with NICE risk factors the NICE 2015 diagnostic thresholds were still found to be cost-effective at a threshold of £30,000 per QALY, with broadly similar ICERs as previously. Similarly, the WHO 2013 diagnostic threshold was never found to be cost effective even in a population with risk factors.

Table x30: Summary of deterministic ICERs for each population without backward elimination of non-significant coefficients

	All covariates		Plasma glucose covariates				
Diagnostic threshold	**HAPO Risk factor (n=3,549)**	**HAPO No Risk factor (n=2,614)**	**HAPO Risk factor (n=3,549)**	**HAPO No Risk factor (n=2,614)**	**Atlantic DiP Risk factor (n=1,988)**	**Atlantic DiP No Risk factor (n=3,302)**	**Norwich (n=12,754)**
No Treatment	-	-	-	-	-	-	-
NICE 2015	£20,162	£38,869	£21,786	£33,473	£19,557	£32,762	£27,354
WHO 2013	£30,734	£94,585	£32,267	£58,604	£35,285	£39,076	£38,402

Mean plasma glucose values according to risk factor status

Table x311: Mean plasma glucose values in HAPO (4) and Atlantic DiP population according to their risk factor status

	HAPO (4)			Atlantic DiP		
	Fasting	1-hour	2-hour	Fasting	1-hour	2-hour
True Positives	5.24	9.90	7.89	5.21	10.21	7.61
False Positives	4.50	7.20	5.95	4.33	6.75	5.33
True Negatives	4.44	6.95	5.78	3.92	5.99	4.76
False Negatives	4.89	9.52	7.41	4.90	9.51	7.12

CHEERS Statement

CHEERS checklist—Items to include when reporting economic evaluations of health interventions

Section/item	Item No	Recommendation	Reported on page No/ line No
Title and abstract			
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	Yes Page 1 Line 2
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.	Yes Page 2 Lines 30-55
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study.	Yes Page 5-6 Lines 92-104
		Present the study question and its relevance for health policy or practice decisions.	Yes Page 5 Lines 74-90
Methods			
Target population and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	Yes Page 6-8 Lines 116-149
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	Yes Page6-7 Lines 116-122
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	Yes Page 12 Line 219
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	Yes Page 6 Line 100-104; 108-111
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	Yes Page 12 Line 230 Supp. Report

Section/item	Item No	Recommendation	Reported on page No/ line No
			Page 15 Line 90
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	Yes Page 12 Line 230 Supp. Report Page 15 Line 90
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	Yes Pages 8-9 Lines 152-172
Measurement of effectiveness	11a	*Single study-based estimates:* Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	Yes Pages 10 Lines 202-205
	11b	*Synthesis-based estimates:* Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	Yes Pages 10 Lines 199-201
Measurement and valuation of preference based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	
Estimating resources and costs	13a	*Single study-based economic evaluation:* Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	
	13b	*Model-based economic evaluation:* Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	Yes Page 12-13 Lines 218-243 Supp. Report Page 14 Line 80
Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for	Yes Page 12 Lines 230

Section/item	Item No	Recommendation	Reported on page No/ line No
		converting costs into a common currency base and the exchange rate.	
Choice of model	15	Describe and give reasons for the specific type of decision-analytical model used. Providing a figure to show model structure is strongly recommended.	Yes Page 6 Lines 108-111; 113-114
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	Yes Page 8-9 Lines 151-172 Yes Page 13 Lines 245-248 Supp. Report Page 2-7 Supp. Report Page 15 Lines 84-102 +References to other sources
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	Yes Page 9-10 Lines 174-191 Yes Page 13-14 Lines 259-266 Supp. Report Page 2-7
Results			
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	Yes Page 11 Lines 215-216 Page 13

Section/item	Item No	Recommendation	Reported on page No/ line No
			Lines 243-244 Supp. Report Page 2-15
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	Yes Page 15 Lines 279-281 Supp. Report Page 17-25
Characterising uncertainty	20a	*Single study-based economic evaluation:* Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	
	20b	*Model-based economic evaluation:* Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	Yes Supp. Report Page 17 Lines 116-119 Supp. Report Page 26 Lines 225-227
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	Yes Supp. Report Page 17 Lines 111-114
Discussion			
Study findings, limitations, generalisability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the	Yes Pages 18-25

Section/item	Item No	Recommendation	Reported on page No/ line No
		findings and how the findings fit with current knowledge.	
Other			
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	Yes Page 26 Lines 547-552
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	Yes Page 26 Lines 557-560