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

Paul Brian Jacklin,[1] Michael JA Maresh,[2] Chris C Patterson,[3] Katharine P Stanley,[4] Anne Dornhorst,[5] Shona Burman-Roy,[1] Rudy W Bilous[6]

To cite: Jacklin PB, Maresh MJA, Patterson CC, *et al*. A cost-effectiveness comparison of the NICE 2015 and WHO 2013 diagnostic criteria for women with gestational diabetes with and without risk factors. *BMJ Open* 2017;**7**:e016621. doi:10.1136/bmjopen-2017-016621

► Prepublication history and additional material for this paper are available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2017-016621).

Received 3 March 2017
Revised 9 June 2017
Accepted 22 June 2017



[1]Royal College of Obstetricians and Gynaecologists, London, UK
[2]St. Mary's Hospital, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK
[3]Centre for Public Health, Queen's University Belfast, Belfast, UK
[4]Department of Obstetrics and Gynaecology, Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich, UK
[5]Department of Investigative Medicine, Hammersmith Hospital, Imperial College London, London, UK
[6]Newcastle University Medicine Malaysia, Johor, Malaysia

Correspondence to
Paul Brian Jacklin;
pjacklin@rcog.org.uk

ABSTRACT

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

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

Participants 6221 patients from four of the Hyperglycaemia and Adverse Pregnancy Outcomes (HAPO) study centres (two UK, two Australian), 6308 patients from the Atlantic Diabetes in Pregnancy 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 CE thresholds of £20 000 and £30 000 per QALY.

Results In a population of pregnant women from the four HAPO study centres and using NICE-defined risk factors for GDM, diagnosing GDM using NICE 2015 criteria had an NMB of £239 902 (relative to no treatment) at a CE threshold of £30 000 per QALY compared with WHO 2013 criteria, which had an NMB of £186 675. NICE 2015 criteria had a 51.5% probability of being cost-effective compared with 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 NMBs 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 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.

Strengths and limitations of this study

- This economic evaluation addresses an important clinical and policy issue. The existing economic evidence is limited and WHO has 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 Hyperglycaemia and Adverse Pregnancy Outcomes study for an economic analysis that 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 National Institute for Health and Care Excellence diagnostic criteria for gestational diabetes mellitus 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 gestational diabetes mellitus (GDM) remain the subject of considerable debate. The original definition was based on 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 Hyperglycaemia and Adverse Pregnancy Outcomes (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[3] based on the HAPO study levels of plasma glucose when fasting and at 1 and 2 hours after an oral 75 g glucose load that were

BMJ Open: first published as 10.1136/bmjopen-2017-016621 on 11 August 2017. Downloaded from http://bmjopen.bmj.com/ on March 28, 2024 by guest. Protected by copyright.

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

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

NICE, National Institute for Health and Care Excellence.

associated with covariate adjusted OR 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 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 light of new studies reporting their cost-effectiveness.[4]

In 2015 National Institute for Health and Care Excellence (NICE) published an updated guidance on diabetes in pregnancy,[6] which included recommendations on diagnostic thresholds for GDM that 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 on 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 analyses.

Population

The model population comprised women of gestational age 24–28 weeks without pre-existing diabetes. The analysis used 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 (online supplementary 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 (RFs) for GDM within a dataset.

1. 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).
2. 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 RFs for GDM. The results were obtained from laboratory records with no identifiers. RFs in addition to those recommended by NICE were used (eg, women with polycystic ovary syndrome, previous stillbirth or recurrent glycosuria).

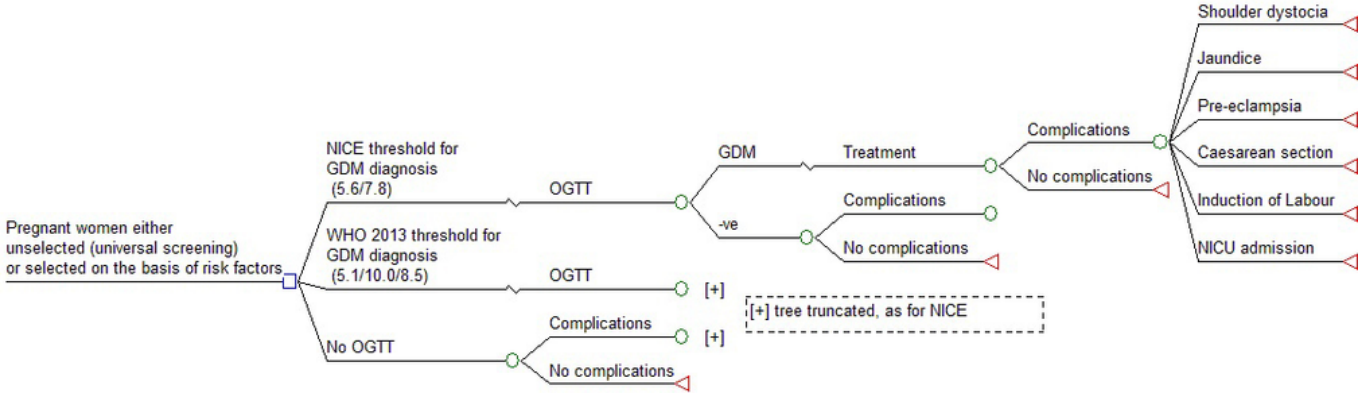


Figure 1 Model schematic. GDM, gestational diabetes mellitus; NICE, National Institute for Health and Care Excellence; OGTT, oral glucose tolerance test.

BMJ Open: first published as 10.1136/bmjopen-2017-016621 on 11 August 2017. Downloaded from http://bmjopen.bmj.com/ on March 28, 2024 by guest. Protected by copyright.

3. Atlantic Diabetes in Pregnancy (DiP): these data were collected between 2007 and 2013 as part of a research initiative in 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 RFs for GDM (body mass index above 30 kg/m^2, previous baby with birth weight ≥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 an RF approach.

The NICE RF 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 RFs. Our Norwich dataset only included the plasma glucose values from a 3-point (fasting, 1 and 2 hours) OGTT, and therefore, it was not possible to assess cost-effectiveness according to the presence of RFs in this group.

Permission was obtained from the relevant Caldecott Guardian to use anonymised patient OGTT data from the Norfolk and Norwich University Hospitals National Health Service (NHS) Foundation Trust for the analysis. The principal investigators from the Australian (Professor HD McIntyre) and British (Professor DR McCance) centres of the HAPO study and the principal 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:

1. Shoulder dystocia (SD): this was used to estimate serious perinatal complications (SPCs), a broader composite outcome (death, SD and birth trauma) used as a primary outcome in clinical trials. The estimation of SPC from SD has been described elsewhere.[6]
2. Caesarean section (CS)
3. Neonatal intensive care unit (NICU) admission
4. Jaundice requiring phototherapy
5. Pre-eclampsia (PE)
6. Induction of labour

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 (eg, 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, OGTT was blinded to the carers, unless there was overt diabetes, thus allowing direct comparison of OGTT with perinatal outcomes without intermediate treatment effects for those meeting the new diagnostic criteria for GDM.

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

1. 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.
2. 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 online supplementary report, tables x2–x7. The Cholesky decomposition of the variance covariance matrices from the regression analyses used in the base case probabilistic sensitivity analysis (PSA) is given in online supplementary report, tables x8–x13.

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, 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

BMJ Open: first published as 10.1136/bmjopen-2017-016621 on 11 August 2017. Downloaded from http://bmjopen.bmj.com/ on March 28, 2024 by guest. Protected by copyright.

Table 2 Relative treatment effects for model outcomes

Outcome	Relative risk (RR)	SE (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)

ACHOIS, Australian Carbohydrate Intolerance Study; NICU, neonatal intensive care unit.

nursery admissions were used. Similarly, the incidence of PE seemed high in ACHOIS in both arms, and again only Landon *et al* trial data were used. 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 PSA.

Costs

Costing was undertaken from the perspective of 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 online supplementary 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]

Quality-adjusted life years

Following previous studies,[6 14] a quality-adjusted life year (QALY) decrement of 2.2 was assigned to SPCs, defined as per the ACHOIS study as a composite outcome of SD, death and birth trauma.[11] More detail on the derivation of this QALY loss is provided in the online supplementary report (including online supplementary table x15). 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 £20 000–30 000 per QALY as a benchmark[15] for the opportunity cost of health foregone, and this paper assesses cost-effectiveness accordingly.

Sensitivity analysis

PSA, using Monte Carlo simulation (with 2000 iterations for each analysis), was undertaken in order to assess the impact of sampling uncertainty on model inputs. Parameters and distributions for the PSA are given in table 2 and online supplementary report, table x14. 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.

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

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

BMJ Open: first published as 10.1136/bmjopen-2017-016621 on 11 August 2017. Downloaded from http://bmjopen.bmj.com/ on March 28, 2024 by guest. Protected by copyright.

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

Threshold name	Norwich (n=12 754) (%)	HAPO all (n=6163) (%)	HAPO RF (n=3549) (%)	HAPO No RF (n=2614) (%)	DiP All (n=5290) (%)	DiP RF (n=1988) (%)	DiP No RF (n=3302) (%)
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

DiP, Diabetes in Pregnancy; GDM, gestational diabetes mellitus; HAPO, Hyperglycaemia and Adverse Pregnancy Outcomes; NICE, National Institute for Health and Care Excellence; RF, risk factor.

(4) and Atlantic DiP datasets, this is additionally broken down in the subgroups with and without NICE RFs.

Detailed deterministic and probabilistic results for HAPO (4) with RFs are shown in tables 5–7 and figure 2.

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 one case of SD.

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).

PSA reached a similar conclusion, with the NICE 2015 diagnostic threshold having the highest probability of being the most cost-effective treatment and the highest net monetary benefit (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 is achieved at an unacceptably high opportunity cost. The online supplementary report plots the incremental cost and QALY outcomes of 2000 simulations from the probabilistic analysis on the cost-effectiveness plane (see figure x1). While 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.

Summaries of results for all of the model populations and more detailed results are provided in the online supplementary report (tables x16–x27 and figures x2–x5).

Tables x28 and x29 in the online supplementary report show that, in both the HAPO (4) and Atlantic DiP populations with NICE RFs, 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 incremental cost-effectiveness ratios (ICERs) of less than £30 000 per QALY, and in the PSA, it has the highest NMB 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 with RF screening (as recommended in NICE guidelines), the additional women included in such an approach would be those without RFs, 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 online supplementary report).

It was not possible to stratify the Norwich dataset according to RFs, 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 with an alternative of no screening/no treatment. Second, the ICERs for the whole population were a weighted average of the populations with and without RFs. The ICER for the

Table 5 Clinical outcomes for HAPO (4) population with NICE risk factors (n=3549)

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	1004
WHO 2013	912	39	54	731	321	207	117	1016

CS, caesarean section; IOL, induction of labour; Jaund, jaundice requiring phototherapy; NICE, National Institute for Health and Care Excellence; NICU, neonatal intensive care unit; PE, pre-eclampsia; SD, shoulder dystocia.

BMJ Open: first published as 10.1136/bmjopen-2017-016621 on 11 August 2017. Downloaded from http://bmjopen.bmj.com/ on March 28, 2024 by guest. Protected by copyright.

Table 6 Deterministic analysis for the HAPO (four centres) population with NICE risk factors (n=3549)

Diagnostic threshold	Cost*	QALY*	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

*Costs and QALYs are measured relative to a baseline of no treatment.
HAPO, Hyperglycaemia and Adverse Pregnancy Outcomes; ICER, incremental cost-effectiveness ratio; NICE, National Institute for Health and Care Excellence; QALY, quality-adjusted life year.

population without RFs 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 rerun using the logistic regression models without backward elimination of glucose variables with non-significant coefficients, and these analyses are discussed in the online supplementary report with the results summarised in online supplementary tables x30.

DISCUSSION

In the NICE guideline analysis, 14 alternative diagnostic thresholds were compared and there was no single optimal diagnostic threshold that 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 on 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 RFs, 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 RFs, 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 with the original NICE guideline are due to a combination of using updated cost data and a modification of the statistical analysis using the Cholesky decomposition (see the Methods section).

One of the limitations of our analysis was that the 2-hour threshold was restricted to the historical WHO 1999 2-hour definition of 7.8 mmol/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

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	£0	54.1	21.0
NICE 2015	£239 902	43.3	51.5
WHO 2013	£186 675	2.7	27.6

*NMB is measured relative to a baseline of no treatment
CE, cost-effectiveness; HAPO, Hyperglycaemia and Adverse Pregnancy Outcomes; NICE, National Institute for Health and Care Excellence; NMB, net monetary benefit; QALY, quality-adjusted life year; WTP, willingness to pay.

BMJ Open: first published as 10.1136/bmjopen-2017-016621 on 11 August 2017. Downloaded from http://bmjopen.bmj.com/ on March 28, 2024 by guest. Protected by copyright.

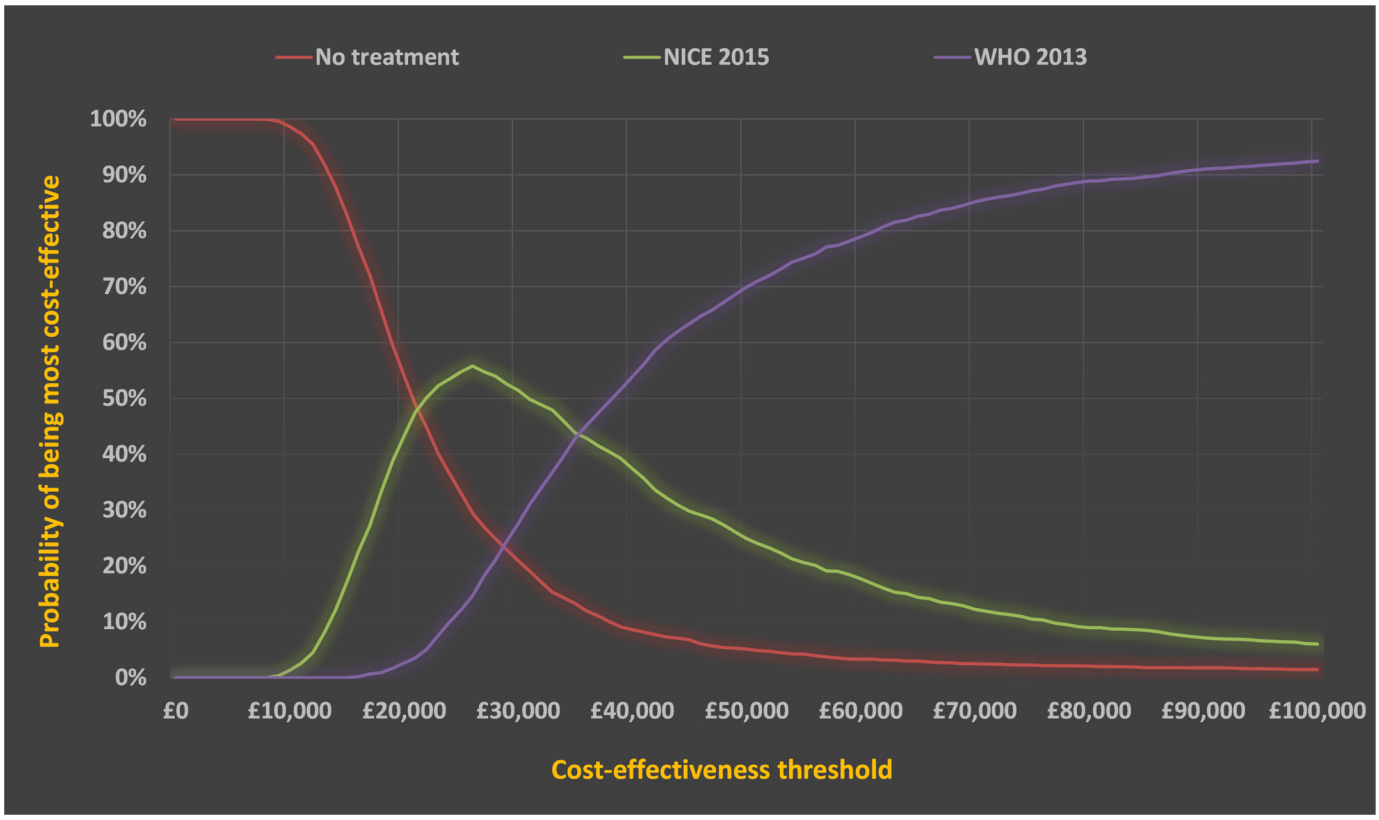


Figure 2 Cost-effectiveness acceptability curve indicating the probability of a diagnostic criteria or a no diagnosis/no treatment strategy being cost-effective at different cost-effectiveness thresholds for the HAPO (4) centres population with risk factors. NICE, National Institute for Health and Care Excellence.

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 that 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 online supplementary tables x2–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 CS and NICU admission where the prediction was based only on OGTT plasma glucose results (model 2 base case analysis regressions in online supplementary tables x2–x7). Nevertheless, as indicated in online supplementary tables x28 and x29, 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 on 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 on 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 healthcare 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 healthcare systems cannot be to maximise 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 healthcare system. If a population is divided into those with RFs and those without RFs, then the prevalence of GDM must be lower in the group without RFs (and the number needed

BMJ Open: first published as 10.1136/bmjopen-2017-016621 on 11 August 2017. Downloaded from http://bmjopen.bmj.com/ on March 28, 2024 by guest. Protected by copyright.

to screen higher) with concomitantly lower cost-effectiveness. However, the comparative cost-effectiveness of screening in those with and without RFs is affected by the respective prevalence in the two groups and differences in severity. In those diagnosed with GDM and who had RFs, there were, as anticipated, greater levels of hyperglycaemia than in those without RFs. As shown in table x31able x31 in the online supplementary report, 'true positives' or identified cases (RF present and GDM) had higher plasma glucose values than 'false negatives' or missed cases (RFs absent and GDM) when defining GDM positives according to WHO 2013 diagnostic thresholds.

We would therefore expect the women with RFs and GDM to be at greater risk of adverse outcomes than the women with GDM without RFs 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 RFs. Thus, the ICER would be greater in the population without RFs 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 RFs, was able to evaluate the cost-effectiveness of moving from RF screening to universal screening. While diagnosis in populations with RFs 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 RFs. Table 4 indicates the large differences that exist in prevalence between the populations with and without RFs. 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 RF 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 RF population were multiparous and would have had a large baby previously and/or a history of GDM. This approximation would overestimate slightly the benefits of universal screening, as the baseline risk in a group designated as being without NICE RFs 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 on estimates of reduction of CS rates of 50%, which we find implausible based on changes in diagnostic criteria alone, noting that ACHOIS and Landon *et al* found only a 4% and 21% reduction in CS, 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 that 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 HTA. Instrumental delivery was included as an outcome in 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 that failed to demonstrate a treatment effect on instrumental delivery rates. In addition, HTA reported smaller treatment effects for NICU admission and PE. Unlike our analysis, HTA did not assume 100% uptake of 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 RFs. HTA also assumed higher uptake of OGTT with RF screening compared with 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 reinforce that result. We investigated the impact of relaxing the assumption of 100% uptake in groups screened on the basis of RFs but found that it made a negligible difference to the results. For example, in a deterministic analysis of the HAPO (4) with NICE RFs, 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 HTA should not be overstated. 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 healthcare 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 SD 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 an SPC used in this analysis is likely to be overstated because of the relatively large weight given to death based on the intervention

BMJ Open: first published as 10.1136/bmjopen-2017-016621 on 11 August 2017. Downloaded from http://bmjopen.bmj.com/ on March 28, 2024 by guest. Protected by copyright.

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 SPCs. In this respect, the cost-effectiveness of diagnosing and treating GDM may be overstated. On the other hand, the model does not take account of any potential long-term effects on the offspring (eg, adiposity and the likelihood of subsequent pathology) as these effects are difficult to quantify but may underestimate 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 5 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 that our analysis represents a robust analysis of the cost-effectiveness of the NICE versus the WHO 2013 diagnostic thresholds for GDM based on 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 with 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 with 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 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 DiP.

Contributors PBJ 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. MJAM provided clinical input into the design of the health economic model; read, commented and edited various draft of the manuscripts. KPS supplied the Norwich dataset, provided clinical input into the design of the health economic model; read, commented and edited various draft of the manuscripts. AD provided clinical input into the design of the health economic model; read, commented and edited various draft of the manuscripts. CCP provided statistical advice, undertook statistical analysis of the HAPO dataset; read, commented and edited various drafts of the manuscript. SB-R reviewed the clinical literature, contributed to discussions of model design; read, commented and edited various drafts of the manuscript. RWB chaired the National Institute for Health and Care Excellence guideline, provided clinical input into the design of the health economic model; read, commented and edited various draft of the manuscripts.

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 1 April 2016), which received funding from the National Institute for Health and Care Excellence (NICE). PBJ and SBR are employees of the National Guideline Alliance (part of the Royal College of Obstetricians and Gynaecologists), which receives its funding from NICE(6).

Disclaimer The views expressed in this publication are those of the authors and not necessarily those of the institute.

Competing interests MJAM, KS, AD and RWB received travel expenses from the National Institute for Health and Care Excellence for attending clinical guideline development meetings.

Patient consent All patient-level data were anonymised and couldn't be used to identify individual patients.

Provenance and peer review Not commissioned; externally peer reviewed.

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

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

Author note The lead author, PBJ, 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.

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

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.
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:e98–82.
4. WHO Health organisation. Diagnostic criteria and classification of hyperglycaemia first detected in pregnancy WHO/NMH/MND/13.2. 2013 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). Diabetes in pregnancy: management of diabetes and its complications from preconception to the postnatal period. Clinical guideline NG3 (2015). 2015 www.nice.org.uk/guidance/ng3/resources/diabetes-in-pregnancy-management-of-diabetes-and-itscomplications-from-preconception-to-the-postnatal-period-51038446021 (accessed Feb 2016).

BMJ Open: first published as 10.1136/bmjopen-2017-016621 on 11 August 2017. Downloaded from http://bmjopen.bmj.com/ on March 28, 2024 by guest. Protected by copyright.

7. Meek CL, Lewis HB, Patient C, *et al*. Diagnosis of gestational diabetes mellitus: falling through the net. *Diabetologia* 2015;58:2003–12.
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;37:2442–50.
9. Herman WH. Insights offered by economic analyses. *Diabetes Care* 2014;37:2424–6.
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;20:1–348.
11. Crowther CA, Hiller JE, Moss JR, *et al*. 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* 2010;340:c1395.
14. Round JA, Jacklin P, Fraser RB, *et al*. Screening for gestational diabetes mellitus: cost-utility of different screening strategies based on a woman's individual risk of disease. *Diabetologia* 2011;54:256–63.
15. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. 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;36:3040–4.
18. Simmons D, Moses RG. Gestational diabetes mellitus: to screen or not to screen? Is this really still a question? *Diabetes Care* 2013;36:2877–8.
19. Armstrong R, Waters E, Doyle J. The Cochrane Collaboration. Reviews in health promotion and public health. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011, 2011. www.cochrane-handbook.org. (Accessed June 2015).
20. Werner EF, Pettker CM, Zuckerwise L, *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;35:529–35.
21. Donovan LE, Cundy T. Does exposure to hyperglycaemia in utero increase the risk of obesity and diabetes in the offspring? A critical reappraisal. *Diabet Med* 2015;32:295–304.
22. Landon MB, Rice MM, Varner MW, *et al*. Mild gestational diabetes mellitus and long-term child health. *Diabetes Care* 2015;38:445–52.

BMJ Open: first published as 10.1136/bmjopen-2017-016621 on 11 August 2017. Downloaded from http://bmjopen.bmj.com/ on March 28, 2024 by guest. Protected by copyright.