Relph Sophie (Orcid ID: 0000-0002-2007-9647)
Khalil Asma (Orcid ID: 0000-0003-2802-7670)

Improving antenatal detection of small-for-gestational-age fetus: economic evaluation of Growth Assessment Protocol

S. Relph[1], M. C. Vieira[1,2], A. Copas[3], K. Coxon[4], A. Alagna[5], A. Briley[1,6], M. Johnson[7], L. Page[8], D. Peebles[9], A. Shennan[1], B. Thilaganathan[10,11], N. Marlow[9], C. Lees[7], D. A. Lawlor[12,13,] A. Khalil[10,11], J. Sandall[1], D. Pasupathy[1,14] and A. Healey[15] on behalf of the DESIGN Trial Team

1. Department of Women and Children's Health, School of Life Course Sciences, Faculty of Life Sciences and Medicine, King's College London, London, UK.
2. Department of Obstetrics and Gynaecology, University of Campinas (UNICAMP), School of Medical Sciences, Sao Paulo, Brazil
3. Centre for Pragmatic Global Health Trials, Institute for Global Health, University College London, London, UK
4. Faculty of Health, Social Care and Education, Kingston and St. George's University, London, UK
5. The Guy's & St Thomas' Charity, London, UK
6. Caring Futures Institute, College of Nursing and Health Sciences, Flinders University, Adelaide, Australia
7. Department of Surgery and Cancer, Imperial College London, London, UK
8. West Middlesex University Hospital, Chelsea & Westminster Hospital NHS Foundation Trust, London, UK
9. UCL Institute for Women's Health, University College London, London, UK
10. Fetal Medicine Unit, St George's University Hospitals NHS Foundation Trust, London, UK
11. Molecular & Clinical Sciences Research Institute, St George's, University of London, London, UK
12. Population Health Science, Bristol Medical School, University of Bristol, Bristol, UK
13. Bristol NIHR Biomedical Research Centre, Bristol, UK

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/uog.26022

14. Reproduction and Perinatal Centre, Faculty of Medicine and Health, University of Sydney, Sydney, Australia
15. Department of Health Service and Population Research, David Goldberg Centre, King's College London, London, UK

Corresponding Author:
Sophie Relph, sophie.relph@kcl.ac.uk
Department of Women and Children's Health, School of Life Course Sciences,
Faculty of Life Sciences and Medicine,
King's College London,
10th Floor North Wing, St. Thomas' Hospital,
Westminster Bridge Road,
London, SE1 7EH

Keywords

Growth assessment protocol, small for gestational age, antenatal screening, cost-effectiveness, economic evaluation

Running Head: Cost-effectiveness: the Growth Assessment Protocol

Contribution

What are the novel findings of this work?

The Growth Assessment Protocol (GAP) is expected to cost an additional £34,559 per 1,000 pregnancies. In terms of cost-effectiveness, it is most likely that GAP was more clinically effective and more costly (44% chance), with a low chance (11%) of it being both more clinically effective and less costly.

What are the clinical implications of this work?

When implemented as seen in sites recruited to the DESiGN trial, the economic case for replacing standard care with GAP, for the improvement in antenatal detection of SGA and stillbirth rates, is weak.

Abstract

Objectives

To determine whether the Growth Assessment Protocol (GAP), as implemented in the DESIGN trial, is cost-effective in terms of antenatal detection of small for gestational age (SGA) neonates, when compared to standard care.

Methods

Design: An incremental cost-effectiveness analysis undertaken from a UK National Health Service hospital provider perspective. Setting: Thirteen maternity units from England, UK, were recruited to the DESiGN trial, a randomised cluster control trial. Population: Singleton, non-anomalous pregnancies in which the baby was born after 24^{+0} gestational weeks between 05 November 2015 and 28 February 2019. Analysis: Probabilistic decision modelling using clinical trial data. Main outcome measures: The expected incremental cost of GAP and additional number of SGA neonates identified antenatally, the incremental cost-effectiveness ratio (ICER) of GAP (cost per additional SGA neonate identified). Secondary analysis: the ICER estimated as the incremental cost per infant quality adjusted life year (QALY) gained.

Results

The expected incremental cost of GAP over standard care was an additional £34,559 per 1000 pregnancies with a 68% probability that GAP would increase costs (including hospital care and implementation costs) to sustain programme delivery. GAP identified an additional 1.77 SGA neonates per 1000 pregnancies (55% probability of being more clinically effective). The ICER for GAP was £19,525 per additional SGA neonate identified (44% probability that GAP would jointly increase cost *and* identify more SGA neonates than standard care). The probability of GAP being the dominant clinical strategy was low (11%). The expected incremental additional cost per infant QALY gained ranged from £68,242 to £545,940 depending on assumptions regarding the QALY value of SGA detection.

Conclusion

The economic case for replacing standard care with GAP is weak based on the analysis reported here. This conclusion should be viewed in the context that cost-effectiveness analyses are always limited by the assumptions made, and our study is no different.

INTRODUCTION

Reducing the prevalence of stillbirth is a global priority.[1] In high-income countries, approximately four of every ten stillborn babies are growth restricted.[2] Stillbirth prevention strategies therefore target risk assessment, antenatal diagnosis, surveillance and timely birth of small for gestational age (SGA, fetal/birthweight below the 10th centile for gestational age) babies.[3,4] The Growth Assessment Protocol (GAP) is a complex intervention that aims to prevent stillbirth through improving antenatal care processes and antenatal detection of SGA.

The DESiGN trial (DEtection of the Small for GestatioNal age Fetus) was the first pragmatic randomised cluster control trial that compared the effect of GAP and standard care in the UK,[5] finding no statistically significant difference in the rate of ultrasound detection of SGA (primary outcome, 25·9% in the intervention and 27·7% in the standard care arm, adjusted difference 2·4%, 95% CI -6·1% to 10·8%; p=0·58) when implemented in this setting.[6]

The inclusion of economic evaluations in healthcare research is recommended to assist decision making about the adoption or spread of implementation.[7] An economic evaluation studying the cost-effectiveness of GAP has not previously been published. The objective of this study was to determine whether GAP was a cost-effective approach to improving antenatal detection of SGA and prevention of stillbirth within hospitals implementing the programme, when compared to hospitals retaining standard practice.

METHODS

This report has been written with reference to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS), for which the entire checklist was adhered to.[8] The trial was registered with the ISRCTN registry (ISRCTN67698474). An analysis plan for developed in 2019 and approved by the joint Steering and Data Monitoring Committee (available on request). See Supplementary File 1 for further details of the methods.

Study design

A probabilistic cost-effectiveness analysis was undertaken using decision-analytic methods applied to clinical data from the DESiGN trial. Costs were estimated from an NHS provider perspective. Costs and clinical outcomes in hospitals randomised to implement GAP were compared to those in hospitals randomised to continue standard care.

Trial design, population and setting

Thirteen cluster (English maternity unit/hospital, predominantly in London) sites were randomly allocated to the GAP intervention (n=7), or to standard care (n=6) between November 2015 and July 2017. Two cluster sites allocated to GAP withdrew before commencing implementation, citing concerns over its expected financial impact. Five remaining cluster sites actually implemented GAP.[6]

All women giving birth in a cluster site were included in the trial database. Women with multiple births, babies with congenital fetal anomalies and those born before 24^{+1} gestational weeks were excluded from the analysis. Data were collected for births during the trial outcome comparison phase (01 September 2018 – 28 February 2019 for most sites) and from the pre-randomisation phase (one year prior to cluster randomisation) for baseline adjustments.

Intervention and comparator

The GAP intervention was designed by a team at The Perinatal Institute, Birmingham, UK. It includes additional staff training, stratification of pregnant women according to risk of SGA, SGA screening protocols that differ by risk strata, defining SGA using fetal or birthweight centiles customised to the woman's (height, weight, parity, ethnicity) and baby's (sex and gestational age)

characteristics, and an audit of missed cases of SGA.[9] Standard care described the screening strategies already implemented within allocated clusters; influenced by the Royal College of Obstetrics and Gynaecology guideline on the 'Detection and Management of the SGA Fetus'.[10] The trial protocol specified that these sites should not implement GAP, nor assess fetal or neonatal size using customised growth charts.

Time horizon

Fetal surveillance using the intervention commences at 24 weeks' gestation. Costs incurred before this gestational age were not expected to vary by intervention and were therefore not included. After this threshold, we included all major antenatal, intrapartum, neonatal and postnatal costs until the date at which the mother or infant were discharged from their care episode that included birth. Costs were not discounted because all were expected to occur within a single year.

Trial data for economic evaluation

All data required for analysis of clinical outcomes and costs were collected from routinely-collected electronic patient records (EPRs). The data collection and management methods, including detailed description of data quality checks made, have previously been published.[11]

Clinical outcome

The primary outcome of the trial was the antenatal detection of SGA (ultrasound screen positive) in a fetus confirmed to be SGA at birth.[5] In this economic analysis, screening outcomes for both SGA (true positive/false negative) and non-SGA (false positive/true negative) babies were studied. For sensitivity analysis we also present results based on a secondary definition of SGA used in the clinical trial and more likely to be used in routine practice (see below). One site was excluded from the analysis of false positive and true negative cases because it provided no data to define these.

Resource utilisation

Data were collected from EPRs on all significant antenatal, intrapartum, postnatal and neonatal activities (Supplementary File 2). Costs were calculated by multiplying units of activity by the

appropriate unit cost. These were then summed to obtain antenatal, intrapartum, postnatal, and neonatal subtotals and a total cost for each birth.

For estimating resource use during the antenatal period, data were widely available on ultrasound scans and antenatal inpatient admissions. Unfortunately, data on antenatal appointments and unscheduled outpatient attendances were either completely or systematically missing at five of the six standard care sites. To maximise the number of clusters available for antenatal cost analysis we excluded these resource items in the base case analysis. This pragmatic decision was guided by the hypothesis that the main cost impact of GAP would most likely arise from an increased number of scans, and the effect of excluding this data was further explored by sensitivity analysis (see below). The primary economic analysis was subsequently carried out using nine cluster sites for antenatal costs (control n=5, intervention n=4). Unlike antenatal care, costs later in the pathway were expected to vary by screening outcome but not by treatment arm (and were therefore included together without differentiation by treatment arm); all 13 sites were therefore available for calculations and contributed data for intrapartum costs, but because of data availability, only nine sites were used for postnatal costs (control n=3, intervention n=6) and 11 sites were used for neonatal costs (control n=4, intervention n=7).

Data were also collected on activities relevant to GAP implementation. The number and type of staff employed was collected from site clinical leads. The numbers of staff members from each professional group (doctors, midwives and sonographers) who attended the site-wide training launches were obtained from the intervention provider. The time taken to complete each training type was estimated as the median reported by participants of semi-structured interviews conducted during the trial process evaluation.[5] The fee charged by the GAP provider, a one-off set-up cost of £500, plus an annual cost titrated by the annual birth rate at the purchasing trust (Supplementary File 3), was also included. We found no evidence during interviews that the generation or use of the GAP fetal growth charts had changed the expected antenatal clinic appointment duration (midwives or

sonographers were still expected to see the same number of women during a session, even if this incurred a loss of rest breaks), and therefore have not included costs for these activities.

Unit costs

Unit costs for each maternity or neonatal care activity were estimated following a systematic review of maternity costs published within economic evaluations conducted in the UK,[12] and review of the available costs published by the Department of Health as part of the national maternity tariff from 2015-16 and 2017-18 (Supplementary File 3).[13,14] Costs were then inflated where appropriate to 2018-19 prices.[15,16] Hourly costs were estimated for each staff group using Unit Costs of Health and Social Care 2018 data published by the Personal Social Services Research Unit (Supplementary File 3).[17]

Modelling approach

The cost-effectiveness model directly linked the costs of care to the four mutually exclusive screening outcomes (Figure 1). The two main modelling outputs were: the total cost of hospital care per 1,000 births under GAP or standard care (sum of antenatal, intrapartum, postnatal, and neonatal care costs with/without GAP implementation costs), and the number of true positive screening outcomes per 1,000 births expected under GAP or standard care. These outputs were then used to evaluate which of four possible conclusions data were consistent with: (1) GAP is associated with a lower cost of care and more true positive births (GAP is the "dominant" clinical strategy); (2) GAP is associated with higher cost and more of true positives (a trade-off); (3) GAP is associated with higher cost and fewer true positives (standard care is "dominant"); or (4) GAP is associated with lower cost and fewer true positives (a trade-off). If conclusion 2 held, we planned to estimate the incremental cost-effectiveness ratio (ICER) for GAP: the expected incremental cost per additional true positive SGA neonate identified.

To reflect uncertainty in the input parameters, expected principally from trial sampling error, the cost-effectiveness analysis was conducted probabilistically. Uncertainty around intervention cost-

effectiveness is therefore presented as (i) the probability of observing alternative cost-effectiveness outcomes and (ii) visually by plotting a 95% confidence ellipse on to the cost-effectiveness plane.[18]

Estimation of input parameters

As for the main trial, a cluster-summary approach was used to extract statistical information required for probabilistic economic modelling.[19] Multivariate analysis of individual screening outcomes and costs, adjusted using maternal age, parity, ethnicity and (for cost outcomes only) body mass index, was conducted to obtain cluster-level predicted values for the proportion of births and the mean cost per birth (for each subtotal of hospital care) associated with each screening outcome during the trial outcome comparison phase. Cluster summary values for the proportion of births expected to be SGA or not-SGA were based on the unadjusted mean value for site clusters.

For input parameters that are subject to a potential treatment effect of GAP (proportion of SGA or not-SGA births in which SGA was detected antenatally, and antenatal costs by screening outcome), the two sites allocated to GAP that did not attempt implementation (as per main trial analysis[5]) were excluded. Linear regression models were fitted to cluster summary values and used to generate probability distributions for each treatment allocation; for screening outcomes expressed as cluster-level proportions a linear regression model was fitted to the logit transformation of the observed outcome with a re-transformation to obtain the predicted proportion to avoid deriving predictions with implausible values. Linear predictions by treatment allocation were also adjusted for trial baseline outcomes (derived from births during the pre-implementation phase) for each site cluster and a trial stratification variable. Monte Carlo simulation was used to obtain 10,000 random draws from a multivariate normal distribution of linear predictions by treatment allocation with adjustments made to predictions using the Cholesky decomposition method to account for correlation between regression parameters.[20]

For the remaining input parameters (proportion of SGA or non-SGA babies born and intrapartum, postnatal and neonatal costs by screening outcome) probability distributions were again generated using Monte Carlo simulation from a pre-specified probability distribution. Probability

distributions were parameterised using the relevant cluster summary data. In all cases, the selection of an appropriate distribution was guided by the need to generate a plausible range of parameter values (i.e., cost per birth constrained to be ≥0 and SGA/non-SGA proportions bounded by the values 0 and 1).

Sensitivity analysis

To assess whether exclusion of unscheduled and scheduled outpatient attendances/appointments from antenatal costs may have biased comparisons between GAP and standard care we repeated our probabilistic analysis of total costs of hospital care under two alternative scenarios: the base case plus an uplift for unscheduled attendance costs; and the base case plus an uplift for scheduled clinic appointment costs (uplifts calculated from utilisation of the resource type in clusters that provided quality data on these). We also tested the sensitivity of our main conclusions to the use of an alternative definition of SGA status used for modelling screening outcomes: SGA defined as a birthweight of < 10th weight centile by population references for standard care and by the customised standard (Gestation-Related Optimal Weight charts) for GAP.[21,22]

Secondary analysis

To aid interpretation of our findings we performed a secondary analysis that re-calibrated the antenatal detection of SGA births into neonate quality-adjusted life expectancy (QALY) gains arising from stillbirth prevention. QALYs are the accepted outcome metric for establishing whether new health care technologies are a cost-effective use of NHS resource.[23] We combined these values with our existing modelling to determine whether the incremental cost and QALY implications associated with GAP satisfy cost-effectiveness criteria currently used by the UK National Institute for Health and Care Excellence (NICE) to guide NHS resource allocation. NICE criteria stipulate that new health technologies should not exceed a cost of £20,000 to £30,000 for every QALY gained.[23]

We used the clinical trial data to establish a baseline incidence of stillbirth among SGA births in which SGA was not detected antenatally. Previous studies suggest that 50% of stillbirths occurring among babies with undetected SGA could be avoided if SGA were detected,[24-26] and NICE estimate

that avoidance of stillbirth would be expected to gain 23.73 (range 15-30) discounted QALYs (applying a discount rate of 3.5%). Combining this evidence, three estimates of the QALY benefit per SGA birth detected antenatally were derived and applied in our secondary analysis: a "central" estimate (assuming 50% of stillbirths linked to undetected SGA are prevented with 23.73 QALYs gained per stillbirth avoided); a "high" (and highly optimistic) estimate (all stillbirths prevented, 30 QALYs gained); and a "low" estimate (25% of stillbirths prevented, 15 QALYs gained).

In an extension to our secondary analysis, we also performed a conditional incremental net benefit (INB) analysis.[27] This was used to assess whether a cost-effective rate of SGA antenatal detection under GAP is likely to be achievable given a plausible distribution of values for this parameter extracted from the trial data. The analysis is repeated under the varying assumptions regarding the QALY value of detecting an SGA birth antenatally, as described above. To implement the conditional INB analysis we firstly monetise the QALY benefit of early detection using the NICE cost-effectiveness threshold. Subtracting incremental monetised benefits from the incremental cost gives the incremental net benefit of GAP. If INB>0 then GAP is considered cost-effective at the chosen threshold level (we adopt the lower value preferred by NICE of £20,000 per QALY gained). The INB is estimated at varying levels of the SGA detection rate corresponding to the deciles within the distribution for this parameter. All other parameters were varied probabilistically (as described earlier).

RESULTS

Data were collected on 209,314 pregnancies, of whom 24,906 women and their babies were included in analyses for the outcome comparison phase (n=13,810 in the standard care arm and n=8,882 in the intervention arm), using adjustments from 55,950 women and their babies included in the pre-randomisation phase (n=29,404 in the standard care arm and n=21,596 in the intervention arm). The consort diagram, characteristics of women included during the pre-randomisation and outcome comparison phases, and results of the analysis for primary and secondary clinical outcomes have been published previously.[6]

Model parameter values estimated from the trial data based on the primary outcome definition of small babies identified at birth are described in Table 1; parentheses contain corresponding values based on the secondary definition of SGA. Cost-effectiveness results relating to the primary definition of SGA status are presented in Table 2 (results pertaining to the secondary definition in parentheses). The expected total cost of all hospital care per 1,000 births was estimated to be £23,763 higher under GAP than under standard care, with a 62% probability that GAP would increase hospital care costs. The cost of implementing GAP (staff training costs and license fees) was estimated to be an additional £10,796 per 1,000 births. The total expected incremental cost of GAP compared to standard care was £34,559 more per 1,000 births, with a 65% probability that GAP would be more costly than standard care.

The expected clinical benefit of GAP observed in the DESiGN trial, in terms of antenatal detection of small babies, was marginal: an additional 1.77 SGA babies detected per 1,000 births (55% probability that GAP would increase antenatal detection compared to standard care). The incremental cost per additional SGA baby identified antenatally was £19,525, with a 44% probability that GAP would be both cost increasing and clinically beneficial compared to standard care. There was only an 11% probability that GAP would dominate standard practice in terms of cost-effectiveness (Figure 2).

Sensitivity analysis

Use of the secondary definition of SGA (SGA defined by customised centiles in GAP implementing clusters and by population centiles in standard care clusters) led to a reduction in the incremental total cost of GAP to £30,861 more per 1,000 births and a small increase in the expected number of additional SGA babies identified antenatally to 2.52 per 1,000 births. The ICER for GAP using the secondary SGA definition was £12,246 per additional case identified. Probability values were close to those observed in the primary analysis.

An uplift applied separately for antenatal appointments and unscheduled attendance costs had little impact on the incremental total cost of GAP (Supplementary File 4).

Secondary analysis

The incremental cost of GAP per additional infant QALY gained was estimated to exceed the NICE cost per QALY threshold range: QALY-based ICERs ranged from £68,242 to £542,940 per QALY gained, depending on assumptions adopted. Using the primary definition of SGA, GAP only achieves a cost-effective rate of antenatal detection when adopting the "high" level assumptions regarding the QALY value of antenatal detection, and only when the rate of detection exceeds 41.5% under GAP (a detection value that is at the 80th decile value within the distribution for this parameter, Figure 3). Similar findings are observed for the analysis based on the secondary definition of SGA.

DISCUSSION

Main Findings

Evidence from this trial-based cost-effectiveness analysis, based on units recruited to the DESiGN trial, suggests that the adoption of GAP in place of standard care will (on average) increase costs to NHS providers while offering only marginal clinical benefit. After full consideration of the margins of uncertainty around economic and clinical parameters of relevance, the expected incremental cost of GAP was £34,559 per 1,000 births, though there remains some uncertainty regarding the magnitude of this effect (a 68% probability that the incremental cost of GAP is positive and a 32% probability that GAP would be cost reducing). Overall, 31% of the expected incremental cost of GAP was attributable to programme implementation. Compared to other estimated resource effects, these costs are also least affected by sampling uncertainty inherent to the clinical trial. There was no convincing evidence that GAP was the dominant clinical strategy in cost-effectiveness terms (11% probability that it reduced costs while also improving the rate of SGA detection). After taking full account of sampling error in the trial data, the most likely outcome observed was that GAP only delivered marginal expected clinical gains and at an additional cost, to those NHS providers who participated in the trial. In secondary analysis, we found no convincing evidence that GAP provided a cost-effective alternative to standard care within participating clusters when applying a routinely adopted NHS cost-effectiveness threshold to our findings.

Strengths and limitations

The main strength is that the evaluation was conducted using data on resource use recorded routinely during clinical practice; it is therefore expected to offer a reliable assessment of the hospital-based resource impact of GAP implementation within the study sites randomised to the programme. GAP was compared to standard contemporaneous practice, rather than to no care, and so the findings reflect the expected cost of implementing GAP over and above those of current practice (analysis method recommended by NICE).[28]

One limitation of this analysis is the adopted time horizon. Estimates of cost were restricted to those incurred by the NHS provider until the end of the care episode including birth. We have accounted neither for infant or adult healthcare costs incurred from morbidity associated with preterm or early term birth that may follow SGA detection, nor intermediate and longer-term health or societal costs associated with stillbirth avoidance including any costs of litigation avoided (estimates have been published previously).[29]

We were limited by the availability and quality of data collected from some clusters. One implementer site was excluded from some analyses because it could not provide data on true negative and false positive SGA diagnoses. Hospital administrative data were entirely missing or not usable for two sites allocated to standard care, and systematically missing for some resource items at three of the remaining standard care clusters. Exclusion of scheduled and unscheduled antenatal hospital appointments or day attendances because of this missing data had only a small effect on ICER, as shown through sensitivity analysis. In all clusters, we were unable to distinguish between women who had absence of an activity recorded because it had occurred elsewhere, because the woman had not received the care anywhere, or because it had occurred but had not been recorded; we introduced assumptions in which plausible limits were applied to deal with this.

Another limitation is the choice of primary SGA definition (SGA as defined by both population and customised weight charts). Using this definition, false positive screening outcomes were defined in approximately 5% of babies who met criteria for SGA definition, but by only one and not both chart types. A sensitivity analysis that adopted an SGA definition that would be more likely to be applied in routine clinical settings (SGA defined by customised centiles in GAP implementing clusters and by population centiles in standard care clusters) produced a lower ICER (£12,246 per SGA baby correctly detected) with comparable estimates of uncertainty around observing alternative cost-effectiveness outcomes. As for the primary SGA definition, application of this alternative definition in the analysis demonstrated that GAP would have only delivered a cost-effective outcome when judged against a

routinely applied cost-effectiveness threshold used in NHS decision making if an unlikely combination of assumptions were built in to the analysis.

Finally, the findings and implications of this economic evaluation are applicable only to healthcare systems that have similar resource availability and national protocols as the clusters included within the DESiGN trial.

Interpretation

The cost-effectiveness of GAP has not previously been studied. The GAP provider has conducted a cost-benefit analysis in which the effect of increasing the frequency of fetal growth scans for women at high risk of SGA was studied.[30] Based on estimates regarding the relationship between SGA detection and infant outcomes (1 fewer stillbirth per 1000 births, £20,000 saved per 1,000 births for reduced neonatal admissions, £25,000/1000 births saved for reductions in cerebral palsy, and £70,000/1000 births saved by reduced litigation), this estimated a cost-saving of £120,000 per 1,000 pregnancies which was attributed to fewer neonatal admissions, lower perinatal morbidity, mortality, cerebral palsy and litigation. Our analysis differed in scope (we did not consider costs of litigation or long-term outcomes) and drew on data generated from a "gold standard" research design linked directly to the implementation of GAP within NHS maternity settings. These reasons alone are likely to explain differences in findings. Whilst we acknowledge that improved SGA detection is expected to reduce both stillbirth and long-term disability related to fetal brain injury, the DESiGN trial found only marginal differences in rates of SGA detection that were not statistically significant.

Our economic evaluation is not supportive of GAP providing a cost-effective improvement to care processes aimed at stillbirth prevention. The expectation based on evidence from this evaluation is that it will increase the costs of hospital care and require an ongoing resource commitment in terms of staff training and software licensing. These additional costs need to be balanced against the small expected incremental clinical benefit that GAP might offer above standard care. We estimated that, even with highly optimistic (and arguably unrealistic) assumptions regarding preventable numbers of

stillborn infants arising from early detection, the QALY value of stillbirth avoidance linked to these small clinical gains will be of insufficient magnitude to justify costs when judged against cost-effectiveness thresholds used in NHS decision making. This conclusion is only likely to have been strengthened had our analysis included longer-term NHS costs arising from stillbirth avoidance and any iatrogenic cost and QALY impacts associated with antenatal detection. Other longer-term cost- and QALY- related benefits claimed to be linked to the early detection of SGA births (e.g., avoidance of litigation costs, perinatal morbidity, and avoidance of long-term developmental disorders) would need to be substantial to offset our core findings. This seems unlikely given the additional rates of antenatal detection observed under GAP in this study.

Conclusion

The economic case for replacing standard care with GAP is weak based on the analysis and evidence reported here. This conclusion should be viewed alongside the context that cost-effectiveness analyses are always limited by the assumptions made, and our study is no different.

ACKNOWLEDGEMENTS

The DESiGN trial team is also comprised of Bolaji Coker, Maria Elstad, Walter Muruet-Gutierrez, Natalie Moitt and Louisa Delaney (all of King's College London), and Professor Lesley McCowan of the University of Auckland. We would like to thank the members of the DESiGN Collaborative Group for their contribution to this trial: Spyros Bakalis, Claire Rozette and Marcelo Canda (from Guy's and St Thomas' Hospital NHS Foundation Trust), Simona Cicero, Olayinka Akinfenwa, Philippa Cox and Lisa Giacometti (from Homerton University Hospital NHS Foundation Trust), Elisabeth Peregrine, Lyndsey Smith and Sam Page (from Kingston Hospital NHS Foundation Trust), Deepa Janga and Sandra Essien (from North Middlesex University Hospital NHS Trust), Renata Hutt (from Royal Surrey County Hospital NHS Foundation Trust), Yaa Acheampong, Bonnie Trinder and Louise Rimell (from St George's University Hospitals NHS Foundation Trust), Janet Cresswell and Sarah Petty (from Chesterfield Royal Hospital NHS Foundation Trust), Bini Ajay, Hannah O'Donnell and Emma Wayman (from Croydon Health Services NHS Trust), Mandish Dhanjal, Muna Noori, and Elisa Iaschi (from Imperial College Healthcare NHS Trust), Raffaele Napolitano, Iris Tsikimi and Rachel Das (from University College London Hospitals NHS Foundation Trust), Fiona Ghalustians and Francesca Hanks (from Chelsea and Westminster Hospital NHS Foundation Trust), Laura Camarasa (from Hillingdon Hospitals NHS Foundation Trust), Hiran Samarage and Stephen Hiles (from London North West Healthcare NHS Trust). We would also like to thank the DESiGN Trial Steering Committee/Data Monitoring Committee members: Anna David (from University College London), David Howe (from University Hospital Southampton), Nadine Seward (from King's College London), Elizabeth Allen (from the London School of Hygiene and Tropical Medicine), and Jillian Francis (from The University of Melbourne). At last, we wish to thank the Stillbirth Clinical Study Group and the Royal College of Obstetricians and Gynaecologists for reviewing the study protocol during development of the study.

COMPETING INTERESTS

NM reports personal fees from Takeda, personal fees from RSM Consulting, personal fees from Novartis, outside the submitted work. BT is the Clinical Director of the Tommy's National Centre for Maternity Improvement based at the Royal College of Obstetrics and Gynaecology; the Centre's objective is to translate the latest evidence into clinical practice in the UK. DAL has received support from Medtronic Ltd and Roche Diagnostics for research unrelated to that presented here.

CONTRIBUTION TO AUTHORSHIP

DP is the Chief Investigator of the DESiGN trial. MCV, AH, KC, AA, DPe, NM, CCL, DAL, AK, JS, AC and DP designed this study. SR and the DESiGN trial team obtained the data locally and conducted the data management procedures. SR and AH undertook the economic evaluation. SR, AH, DP, MCV, AK and JS reviewed and interpreted the results. SR, AH, MCV and DP have drafted and edited the manuscript. All authors have reviewed the draft manuscript, read and approved the final version of the manuscript.

ETHICAL APPROVAL

Ethical approval for this trial has been obtained through the Health Research Authority (HRA) Integrated Research Applications System (IRAS) from the London Bloomsbury Research Ethics Committee (Ref. 15/LO/1632) and the Confidentiality Advisory Group (Ref. 15/CAG/0195). King's College London is the sponsor for this trial.

FUNDING

This study was funded by Guy's and St Thomas' Charity (MAJ150704), Stillbirth and Neonatal Death Charity - SANDS (RG1011/16) and Tommy's Charity. MCV was supported by CAPES (BEX 9571/13–2). SR, KC and AH were supported by the National Institute for Health Research (NIHR)

Applied Research Collaboration (ARC) South London at King's College Hospital NHS Foundation Trust. AH is also a member of King's Improvement Science, which offers co-funding to the NIHR ARC South London and comprises a specialist team of improvement scientists and senior researchers based at King's College London. Its work is funded by King's Health Partners (Guy's and St Thomas' NHS Foundation Trust, King's College Hospital NHS Foundation Trust, King's College London and South London and Maudsley NHS Foundation Trust), Guy's and St Thomas' Charity and the Maudsley Charity. NM receives a proportion of funding from the Department of Health's NIHR Biomedical Research Centres funding scheme at UCLH/UCL. DAL's contributions were supported by the Bristol NIHR Biomedical Research Centre and her NIHR Senior Investigator Award (NF-0616-10102). JS is supported by an NIHR Senior Investigator Award and the NIHR Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust. DP was funded by Tommy's Charity during the period of the study. The views expressed are those of the author[s] and not necessarily those of the NIHR, the Department of Health and Social Care or any of the other listed funders. None of the funders influenced the design, analyses or interpretation of results.

REFERENCES

1. Every Woman Every Child: The Global Strategy for Women's, Children's and Adolescents' Health (2016-2030). https://www.who.int/life-course/partners/global-strategy/ewec-globalstrategyreport-200915.pdf?ua=1 [Accessed 25 February 2021].
2. Gardosi J, Kady SM, McGeown P, Francis A, Tonks A. Classification of stillbirth by relevant condition at death (ReCoDe): population based cohort study. *BMJ* 2005; **331**: 1113-1117.
3. O'Conner D. Saving Babies' Lives: A care bundle for reducing stillbirth.NHS England, 2016.
4. McCowan LM, Figueras F, Anderson NH. Evidence-based national guidelines for the management of suspected fetal growth restriction: comparison, consensus, and controversy. *Am J Obstet Gynecol* 2018; **218**: S855-S868.
5. Vieira MC, Relph S, Copas A, Healey A, Coxon K, Alagna A, Briley A, Johnson M, Lawlor DA, Lees C, Marlow N, McCowan L, Page L, Peebles D, Shennan A, Thilaganathan B, Khalil A, Sandall J, Pasupathy D, Group DEC. The DESiGN trial (DEtection of Small for Gestational age Neonate), evaluating the effect of the Growth Assessment Protocol (GAP): study protocol for a randomised controlled trial. *Trials* 2019; **20**: 154.
6. Vieira M, Relph S, Muruet-Gutierrez W, Elstad M, Coker B, Moitt N, Delaney L, Winsloe C, Healey A, Coxon K, Alagna A, Briley A, Johnson M, Page L, Peebles D, Shennan A, Thilaganathan B, Marlow N, McCowan L, Lees C, Lawlor D, Khalil A, Sandall J, Copas A, Pasupathy D, DESiGN Collaborative Group. Evaluation of the Growth Assessment Protocol (GAP) for antenatal detection of small for gestational age: the DESiGN cluster randomised trial. *PLOS Medicine* 2022.
7. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council G. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008; **337**: a1655.
8. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, Augustovski F, Briggs AH, Mauskopf J, Loder E, Force IHEEPG-CGRPT. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value Health* 2013; **16**: 231-250.
9. Clifford S, Giddings S, South M, Williams M, Gardosi J. The Growth Assessment Protocol: a national programme to improve patient safety in maternity care. *MIDIRS Midwifery Digest* 2013; **23**: 516-523.
10. Royal College of Obstetricians & Gynaecologists. Small for Gestational Age Fetus: Investigation & Management. Green-top Guideline No. 31. https://www.rcog.org.uk/guidance/browse-all-guidance/green-top-guidelines/small-for-gestational-age-fetus-investigation-and-management-green-top-guideline-no-31/.].
11. Relph S, Elstad M, Coker B, Vieira MC, Moitt N, Gutierrez WM, Khalil A, Sandall J, Copas A, Lawlor DA, Pasupathy D, team DT. Using electronic patient records to assess the effect of a complex antenatal intervention in a cluster randomised controlled trial-data management experience from the DESiGN Trial team. *Trials* 2021; **22**: 195.
12. Relph S, Delaney L, Melaugh A, Vieira MC, Sandall J, Khalil A, Pasupathy D, Healey A, Team DET. Costing the impact of interventions during pregnancy in the UK: a systematic review of economic evaluations. *Bmj Open* 2020; **10**: e040022.
13. Department of Health. Reference costs 2015-16. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/577083/Reference_Costs_2015-16.pdf [Accessed 20 December 2018].
14. NHS Improvement. National tariff payment system 2017/18 and 2018/19. https://improvement.nhs.uk/resources/national-tariff-1719/ [Accessed 05 December 2018.].
15. Department of Health. Pay & Price Series 2015/16. www.info.doh.gov.uk [Accessed 11 December 2018].

16. NHS Improvement. Economic Assumptions for 2016/17 to 2020/21. https://www.gov.uk/government/publications/economic-assumptions-201617-to-202021/economic-assumptions-201617-to-202021 [Accessed 11 December 2018].
17. Personal and Social Services Research Unit. Unit Costs of Health and Social Care 2018, Appendix V. https://www.pssru.ac.uk/pub/uc/uc2018/sources-of-information.pdf [Accessed 10 August 2020].
18. Glick H, Doshi J, Sonnad S, Polsky d. *Economic Evaluation in Clinical Trials*. Oxford University Press: Oxford, UK, 2015.
19. Hayes R, Moulton L. *Cluster Randomized Trials*. Taylor & Francis: Abingdon, UK, 2009.
20. Briggs A, Sculpher M, Claxton K. *Decision Modelling for Health Economic Evaluation*. Oxford University Press: Oxford, UK, 2006.
21. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. *Radiology* 1991; **181**: 129-133.
22. Perinatal Institute. GROW-App UK. https://app.growservice.org/uk/ [Accessed 14 June 2019].
23. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781 [Accessed 28 November 2020].
24. Gardosi J, Madurasinghe V, Williams M, Malik A, Francis A. Maternal and fetal risk factors for stillbirth: population based study. *BMJ* 2013; **346**: f108.
25. Ego A, Monier I, Skaare K, Zeitlin J. Antenatal detection of fetal growth restriction and risk of stillbirth: population-based case-control study. *Ultrasound Obstet Gynecol* 2020; **55**: 613-620.
26. Smith G. Should we implement universal screening with late pregnancy ultrasound to prevent stillbirth? *BJOG* 2018; **125**: 101-103.
27. McCabe C, Paulden M, Awotwe I, Sutton A, Hall P. One-Way Sensitivity Analysis for Probabilistic Cost-Effectiveness Analysis: Conditional Expected Incremental Net Benefit. *Pharmacoeconomics* 2020; **38**: 135-141.
28. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). *Lancet* 2002; **360**: 711-715.
29. Campbell HE, Kurinczuk JJ, Heazell A, Leal J, Rivero-Arias O. Healthcare and wider societal implications of stillbirth: a population-based cost-of-illness study. *BJOG* 2018; **125**: 108-117.
30. Williams M, Turner S, Butler E, Gardosi J. Fetal growth surveillance - Current guidelines, practices and challenges. *Ultrasound* 2018; **26**: 69-79.

FIGURE TITLES

Figure 1 Decision analytical model

Figure 2 Cost-effectiveness plane - The proportion of the 10,000 incremental paired cost and clinical effect differences in each quadrant determine the probability of observing each of the four possible outcomes.

Figure 3 The expected incremental net benefit of GAP compared to the NICE cost-effectiveness threshold, conditional on a 'low', 'central' and 'high' estimates of QALYs gained, presented per decile of the expected distribution of SGA detection rate. The figure plots the "incremental net benefit" (INB) of GAP (y-axis) for each % of SGA detection achieved by GAP (x-axis) that would arise under a range of scenarios. INB is a measure of the cost-effectiveness of GAP: INB > 0 implies that the value the NHS places on the gains in Quality Adjusted Life Years (QALY) from improved SGA detection outweighs the additional cost of GAP implementing GAP (i.e. GAP is "cost-effective" compared to standard care); INB <0 implies that GAP offers comparatively poor value for money. The levels of SGA detection shown on the horizontal axis of the figure correspond to decile values from a simulated probability distribution, generated using statistical information generated from trial data, of % of SGA births detected under GAP. A value corresponding to the 20th percentile of the distribution indicates that there is a relative low probability (20%) of observing clinical values at or below that value. The figure is intended to communicate the rate of SGA detection that GAP would need to achieve for it to be considered a cost-effective alternative to standard care, and whether these rates of detection are statistically likely to occur. The analysis of INB is repeated using alternative assumptions regarding the QALY benefits associated with early detection of an SGA neonate. Overall the figure suggests that the rate of SGA detection required to generated a cost-effective outcome for GAP (INB>0) would be unlikely to occur.

Table 1: Model input parameters[1] (N=10,000 model simulations)

		GAP		Standard care		
Screening outcome		**Mean of parameter distribution**	**Interquartile range**	**Mean of parameter distribution**	**Interquartile range**	**Probability distribution from which parameter values are sampled**
% SGA neonates true positive		30.9% (30.8%)	17.8% (17.1%)	28.5% (28.3%)	15.8% (15.5%)	Multivariate normal
% non-SGA neonates false positive		2.3% (2.4%)	1.6% (1.6%)	1.6% (1.6%)	1.0% (1.0%)	
		Mean of parameter distribution		**IQR**		
Antenatal cost under standard care (per birth)	True positive	£1,276 (£1,263)		£339 (£329)		Gamma: α=25.5 (28.0), β=50.1 (45.1)
	False negative	£848 (£829)		£156 (£151)		Gamma: α=53.3 (55.2), β=15.9 (15.0)
	True negative	£670 (£690)		£118 (£141)		Gamma: α=42.0 (42.2), β=16.4 (16.3)
	False positive	£1,074 (£1,075)		£324 (£349)		Gamma: α=19.4 (16.8), β=55.2 (64.1)
		Mean of parameter distribution	**IQR**	**95% Confidence interval[2]**		
Incremental effect of GAP on antenatal cost (per birth)	True positive	£232 (£164)	£438 (£380)	-£406 to £861 (-£387 to £718)		Multivariate normal
	False negative	£45 (-£4)	£25 (£58)	£8 to £82 (-£89 to £79)		
	True negative	-£1 (-£2)	£73 (£74)	-£109 to £107 (-£107 to £108)		
	False positive	-£233 (-£336)	£468 (£311)	-£711 to £248 (-£780 to £122)		
		Mean of parameter distribution		**IQR**		
Antenatal cost under GAP (per birth)	True positive	£1,508 (£1,428)		£550 (£501)		NA[3]
	False negative	£894 (£826)		£159 (£162)		
	True negative	£689 (£688)		£160 (£158)		
	False positive	£846 (£750)		£461 (£455)		
Cost of intrapartum care (per birth)	True positive	£3,022 (£2,996)		£115 (£115)		Gamma: α=1237.2 (1202.6), β=2.4 (2.5)
	False negative	£2,724 (£2,699)		£104 (£106)		Gamma: α=1237.2 (1202.6), β=2.2 (2.2)
	True negative	£2,798 (£2,711)		£87 (£88)		Gamma: α=1744.6 (1707.6), β=1.6 (1.6)
	False positive	£2,801(£2,791)		£91 (£90)		Gamma: α=1744.5 (1707.6), β=1.6 (1.6)

Cost of postnatal care (per birth)	True positive	£729 (£693)	£323 (£288)	Gamma: α=8.7 (9.7), β=83.8 (70.9)
	False negative	£467 (£451)	£207 (£190)	Gamma: α=8.7 (9.8), β=53.4 (46.0)
	True negative	£364 (£357)	£133 (£133)	Gamma: α=13.3 (12.7), β=27.1 (28.2)
	False positive	£561 (£547)	£203 (£204)	Gamma: α=13.3 (12.7), β=42.0 (43.0)
Cost of neonatal care (per birth)	True positive	£2,803 (£2,767)	£1,395 (£1,537)	Gamma: α=4.7 (5.5), β=593.6 (509.8)
	False negative	£1,110 (£1,177)	£599 (£805)	Gamma: α=4.7 (3.5), β=212.1 (341.2)
	True negative	£416 (£405)	£203 (£215)	Gamma: α=7.4 (6.3), β=56.4 (64.6)
	False positive	£2,351 (£2,203)	£1,151 (£1,158)	Gamma: α=7.4 (6.3), β=320.3 (350.1)
GAP implementation cost (software license and recurrent staff training; per birth)		£10.80 (£10.80)	£0.84 (£0.84)	Gamma: α=305.9 (305.9), β=0.03 (0.03)
% births identified as SGA		7.4% (10.0%)	2.0% (3.4%)	Beta: α=22.0 (14.1), β=274.6 (125.2)
% undetected SGA neonates still born (secondary analysis)		0.98% (0.81%)	0.84% (0.69%)	Beta: α=2.0 (2.2), β=206.8 (272.3)

[1] Reported values based on clinical trial primary outcome definition of SGA status at birth (infants who weigh less than the 10th centile based on customised *and* population growth charts).Values in parentheses are the corresponding values based on the secondary definition of SGA status (infants who weigh less than the 10th centile based on customised or population charts if managed under GAP or standard care respectively).

[2] 95% confidence limits approximated by the 2.5th and 97.5th percentiles from each output distribution.

[3] NA=Not applicable. The antenatal cost of GAP per birth by screening outcome was derived indirectly by adding the incremental cost of GAP (its treatment effect) to the antenatal cost per birth under standard care.

Table 2: Cost-effectiveness analysis: results[1]

Screening outcome for modelled cohort of births	Expected number of births (per 1000 births)	
	GAP	**Standard care**
True positive	23 (31)	21 (29)
False negative	51 (70)	53 (72)
True negative	905 (878)	911 (885)
False positive	21 (21)	15 (14)
	Expected incremental cost of GAP (per 1000 births)	**Probability[2] (%) that GAP is cost increasing**
GAP implementation cost (annual software licence and recurrent staff training)	£10,796 (£10,796)	100% (100%)
Incremental hospital care costs		
Antenatal	£4754 (-£21)	54% (50%)
Labour	£1,122 (£1,308)	60% (59%)
Postnatal	£1721 (£1,966)	61% (61%)
Neonatal	£16,165 (£16,812)	65% (66%)
Total incremental hospital care costs	£23,763 (£20,065)	62% (60%)
Total incremental cost (implementation +hospital care)	£34,559 (£30,861)	68% (65%)
		Probability[2] (%) that GAP increases antenatal detection of SGA neonates
Additional number of true positive neonates under GAP (per 1000 births)	1.77 (2.52)	55% (55%)
Incremental cost per additional true positive neonate (ICER)	£19,525 (£12,246)	
Probability[3] (%) that GAP increases total cost of care and detects more SGA neonates antenatally	44% (45%)	
Probability GAP is dominant	11% (10%)	
Probability standard care is dominant	24% (20%)	
Probability GAP reduces total cost and detects fewer SGA neonates antenatally	21%(25%)	

Secondary analysis				**Probability[4] (%) that GAP is cost-effective**	
Infant QALY loss per undetected still birth	Expected number of still births avoided due to GAP (per 1000 births)	Infant QALY gains per 1000 births	Expected cost per QALY gained (ICER)	CET[5]=20K per QALY gained	CET[5]=£30K per QALY gained
High estimate	0.017 (0.019)	0.51 (0.58)	£68,242 (£52,993)	38% (40%)	41% (43%)

Central estimate	0.008 (0.010)	0.20 (0.23)	£172,547 (£133,991)	35% (36%)	35% (37%)
Low estimate	0.004 (0.005)	0.063 (0.073)	£545,940 (£423,948)	33% (35%)	33% (35%)

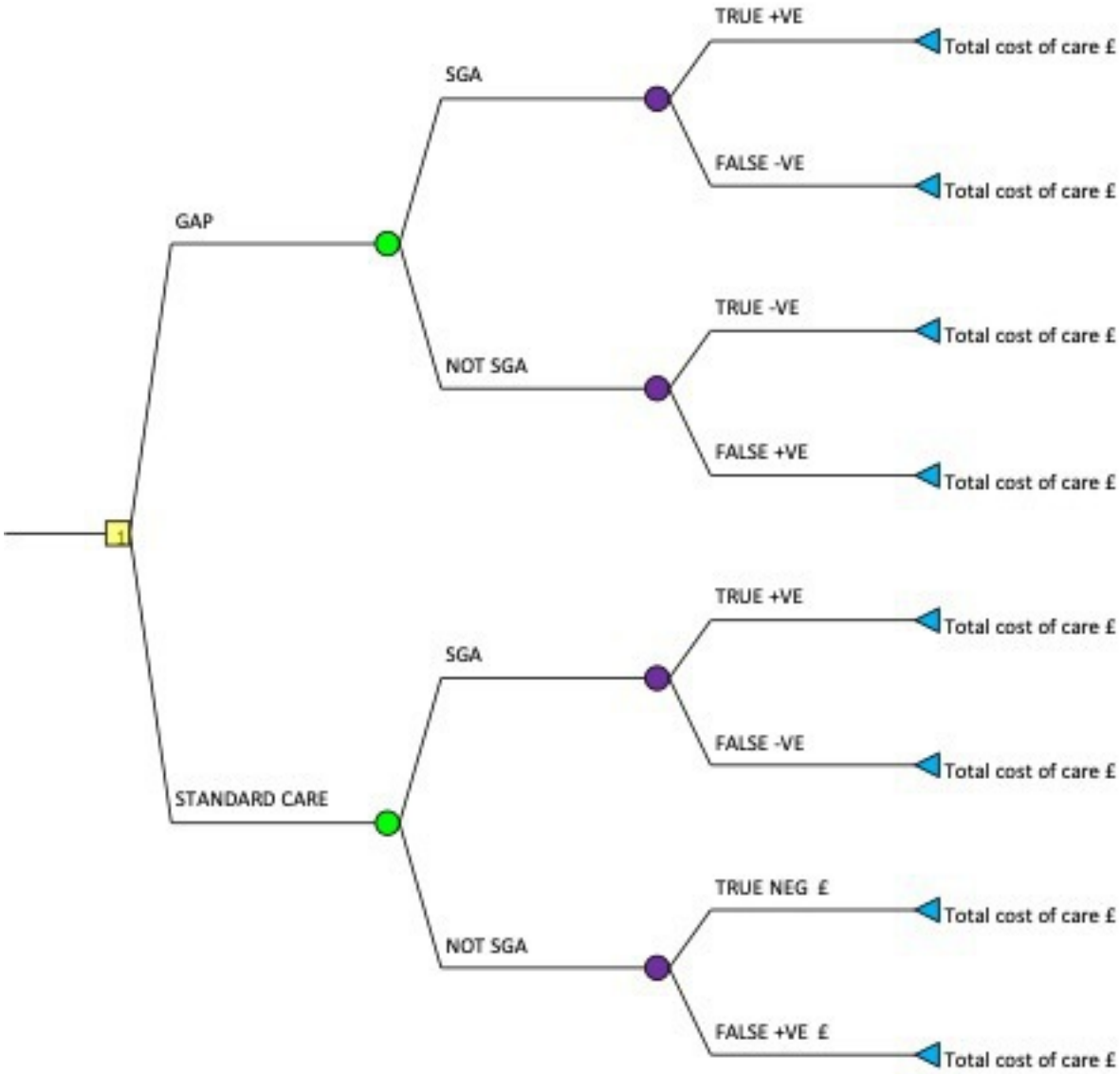
[1] Reported values based on clinical trial primary outcome definition of SGA status at birth (infants who weigh less than the 10th centile based on customised *and* population growth charts).Values in parentheses are the corresponding values based on the secondary definition of SGA status (infants who weigh less than the 10th centile based on customised or population charts if managed under GAP or standard care respectively).

[2] Probabilities derived from the % of model simulations resulting in positive incremental cost or clinical effect.

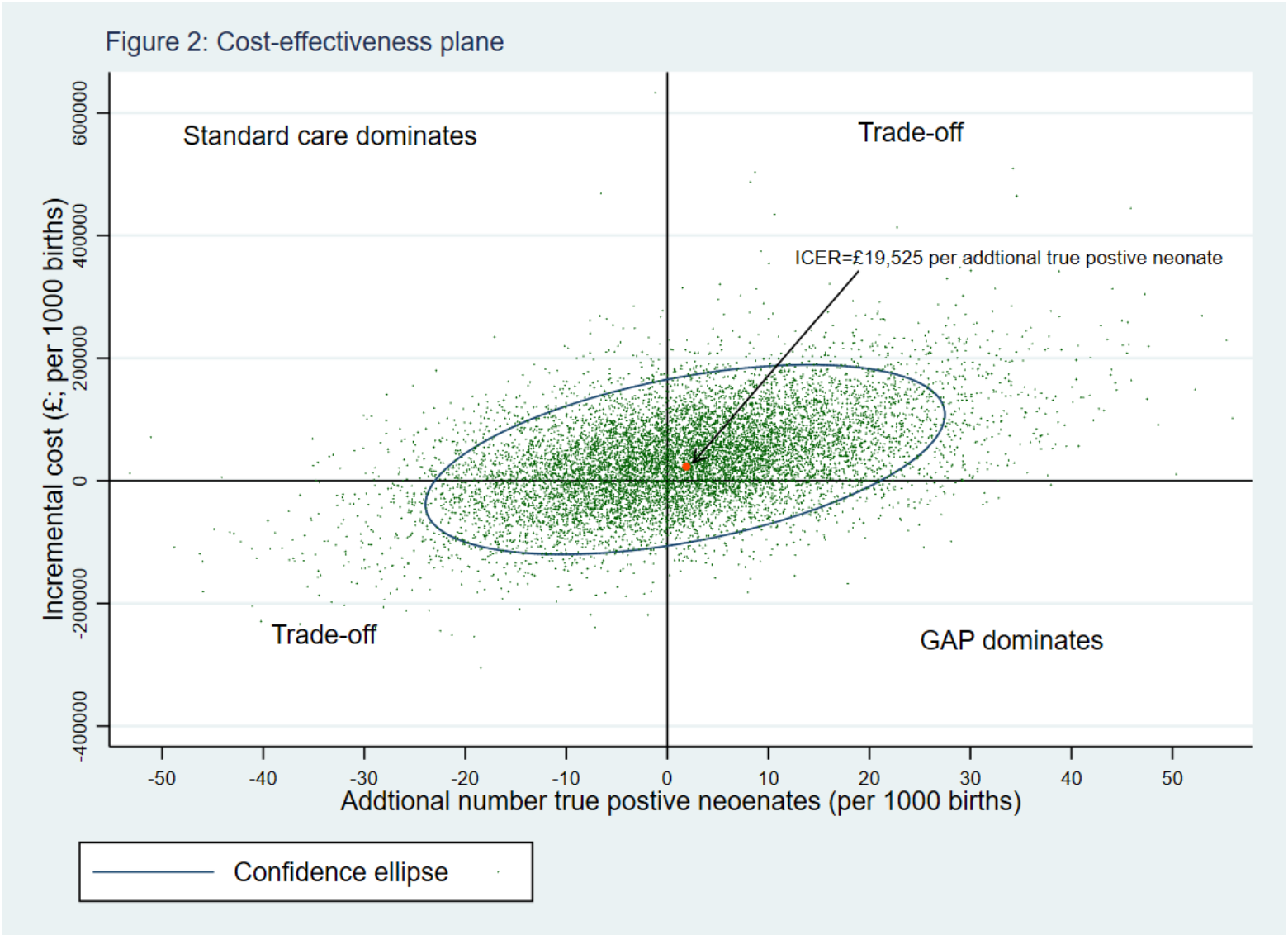
[3] Probabilities derived from the % of incremental cost and clinical effect pairing located within each of the four quadrants of the cost-effectiveness plane (see figure 2).

[4] Probabilities based on the % of incremental net benefit values across model simulations that are > 0 (i.e., where GAP is cost-effective).

[5] CET=Cost-effectiveness threshold.



UOG_26022_Figure 1.jpg



UOG_26022_Figure 2.tif

UOG_26022_Figure 3.tif