PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Repeat screening for syphilis in pregnancy as an alternative screening strategy in the United Kingdom: a cost-effectiveness analysis.
AUTHORS	untington, Susie; Weston, Georgie; Seedat, Farah; Marshall, john; Bailey, Heather; Tebruegge, Marc; Ahmed, Imtyaz; Turner, Katy; Adams, Elisabeth

VERSION 1 – REVIEW

REVIEWER	Catherine Albright University of Washington, Seattle, WA USA
REVIEW RETURNED	01-Apr-2020

GENERAL COMMENTS	Huntington et al have performed a cost-effectiveness analysis of repeat syphilis screening in pregnancy and found that in the UK, this is unlikely to be cost-effective. - While utility discounting of 3.5% is standard in the UK, I have found it very informative to vary the discounting from 0 to 5% in a sensitivity analysis. Might you consider doing and reporting this sensitivity analysis? Utility discounting has never made as much sense to me as cost discounting. - Please state the standard NICE willingness to pay threshold in your methods.

REVIEWER	Lucy Abel University of Oxford, UK
REVIEW RETURNED	25-Jun-2020

GENERAL COMMENTS	This is a well-written and interesting decision model that addresses an important question in antenatal care. The authors construct a decision tree to compare the costs and consequences of repeat syphilis screening in the third trimester. This has the potential to reduce the risk of congenital syphilis, which can result in substantial lifetime health losses and healthcare costs for the baby. However, as the paper clearly shows, the low incidence of syphilis in the UK means additional screening is unlikely to be cost-effective at this time. The introduction and abstract are clear, describe the condition and screening programme in detail, and provide a thorough rationale for the economic evaluation. The discussion is also well-written and places the work in context. The authors should also be applauded for their data transparency, which has made reviewing this paper much easier. I have three major concerns that need addressing before publication.

	1. The authors perform a cost-utility analysis using QALYs, but relegate this to a sensitivity analysis and instead make their primary outcome cost-per-CS case avoided. However, there is no standard cost-effectiveness threshold for cost-per CS case avoided, so all of their cost-effectiveness judgements are made on the basis of the cost-utility estimate. This is also how the majority of readers will approach the paper. As a result, the cost-per-QALY estimate should be presented as a co-primary outcome. The probabilistic sensitivity analysis should be repeated using QALYs, and a CE plane should be presented. The authors say in the discussion that the uncertainty surrounding long-term costs and QALYs is too great to present them as anything other than sensitivity analysis, but that is simply an argument for conducting proper PSA with sensible credible intervals that fully capture that uncertainty. The paper would be greatly improved by conducting this PSA and reporting this as a cost-effectiveness acceptability curve across a range of thresholds. This would also address concerns – briefly raised in the paper – surrounding how to value health gains and costs that occur far in the future. 2. Maternal outcomes are not included in the analysis. This exclusion is mentioned but not justified. There are clearly maternal health benefits to identifying and treating women with syphilis, so while this may not be the intended effect of the programme, it is in line with best practice to include all relevant outcomes. Maternal QALYs and relevant costs should therefore be modelled as a sensitivity analysis. 3. Perinatal mortality is assumed to be unaffected by this intervention, and it is unclear whether this assumption is explored in the sensitivity analysis. Other studies have reported a relationship between preterm birth and perinatal mortality (Callaghan et al, “The Contribution of Preterm Birth to Infant Mortality Rates in the United States”, 2006), and CS and perinatal mortality (Hersh 2018). The potential contribution of perinatal mortality to overall cost-effectiveness should be explored in the SA. Minor comments: Table 1. The sources of evidence for the diagnostic accuracy are inadequate. Reference 22 is an analysis of newborn rather than prenatal testing, and makes no reference to either EIA or congenital syphilis. Reference 23 is a case series. Neither paper reports diagnostic accuracy values for syphilis in pregnant patients. The diagnostic accuracy reported is exceptionally high with very narrow credible intervals. A strong justification is required, or alternatively a wider credible interval should be used for both the one way sensitivity analysis and the PSA. The DA reported here is substantially higher than in both the Hersh and Albright studies. The tests may be different, but this emphasises the lack of evidence supporting such high reported DA. Table 1. There appear to be multiple issues with the references used in table 1. Reference 21 is used to support the probability of developing syphilis between the 1st and 3rd trimesters, but in fact refers to a costing study of lifetime cerebral palsy costs. Likewise, the parameters that refer to a cerebral palsy costing study instead point to reference 27, which is an incidence study. Finally the Hersh study is cited twice, as a full paper (ref 16) and a conference

	abstract (ref 26). Page 12. A utility value of 1 is used for non-CS infants. This is unrealistic, particularly as children become adults, and then elderly adults. UK population norms (eg, Kind, 1999) should be used to obtain age-adjusted utilities which can then be applied to calculate expected QALYs. Discussion. The authors reference two previous studies, reporting that Hersh shows the intervention to be cost-saving, while Albright reports it to not be cost-effective. However Albright actually defines cost-effectiveness in terms of whether the intervention is cost-saving. What these two studies actually say is that the intervention may or may not be cost-saving. The authors should rephrase this sentence to reflect this.

VERSION 1 – AUTHOR RESPONSE

#	**Reviewer comment (grey) and our response/action**
1.1	While utility discounting of 3.5% is standard in the UK, I have found it very informative to vary the discounting from 0 to 5% in a sensitivity analysis. Utility discounting has never made as much sense to me as cost discounting. **Might you consider doing and reporting this sensitivity analysis**?
	Response + Action: Thank you for your suggestion. We have included an additional DSA where utility discounting is varied from 0-6% (since 6% is typically used by NICE for DSA.) The Methods section (p11, paragraph 3) now states: “In DSA, discounting of utilities was varied from 0 to 6%.” The results of this DSA are included in the Supplementary Material (Table S15), and in the main results (page 15, paragraph 3) we have now stated: “In DSA, the ICER was £32,716 and £205,600, when discounting of utilities was 0% and 6% respectively (see **Table S15** in online supplement)”. This is also mentioned as follows in the discussion (top of page 17): “For this analysis, costs and utilities were discounted by 3.5%, in line with NICE guidelines for England [32]. When no discounting of utilities was assessed in DSA, the cost per QALY gained was £32,716, just above the £30k threshold. Even if a lower discounting rate were considered by NICE in the future, it is unlikely that 0% discounting would be used.”
1.2	Please state the standard NICE willingness to pay threshold in your methods.
	Response + Action: We have now stated the standard NICE willingness to pay threshold in the methods section when describing Scenario 7 (page 10): “Examining the per screen cost required to meet the standard National Institute for Health and

	Care Excellence (NICE) cost-effectiveness threshold of £20,000-£30,000 [28]".
2.1	The authors perform a cost-utility analysis using QALYs but relegate this to a sensitivity analysis and instead make their primary outcome cost-per-CS case avoided. However, there is no standard cost-effectiveness threshold for cost-per CS case avoided, so all of their cost-effectiveness judgements are made on the basis of the cost-utility estimate. This is also how the majority of readers will approach the paper. As a result, the cost-per-QALY estimate should be presented as a co-primary outcome and the PSA should be repeated using QALYs, and a CE plane should be presented. The authors say in the discussion that the uncertainty surrounding long-term costs and QALYs is too great to present them as anything other than sensitivity analysis, but that is simply an argument for conducting proper PSA with sensible credible intervals that fully capture that uncertainty. The paper would be greatly improved by conducting this PSA and reporting this as a cost-effectiveness acceptability curve across a range of thresholds. This would also address concerns – briefly raised in the paper – surrounding how to value health gains and costs that occur far in the future.
	Response: We agree with the reviewer that a lack of a threshold for cost per CS avoided presents some difficulty for interpretation. We have included the following additional text in the limitations section to clarify this point (Strengths and weaknesses: end Page 17). Nevertheless, we have made the cost-effectiveness judgements around both the cost per CS avoided and the cost utility estimate. "There were sparse UK data available on pregnancy outcomes in women treated for syphilis or in infants born with CS. <u>There are no published EQ-5D scores for CS and a lack of evidence on changes to utility and health and social care costs over time for infants born with CS</u>. We therefore used the additional lifetime cost of CP, estimated in a study from Denmark [27], as a proxy for the lifetime cost of CS. As such, the primary focus of the analysis was the short-term costs and CS cases avoided since it was difficult to have confidence in the estimate used for lifetime CS cost <u>or utility</u>." Our preference would be not to promote the cost per QALY to the primary outcome as we do not have confidence in the data on long term costs and utilities. Unfortunately, there are no published EQ-5D scores for CS and there is a lack of evidence on changes to utility and health and social care costs over time for infants born with CS. A PSA shows parameter uncertainty but would not capture the structural uncertainty of the long-term CE analysis. Inclusion of a PSA for cost-per-QALY would not provide any additional information that would alter the decision of whether to recommend universal repeat screening since the ICER is so far above the threshold for cost-effectiveness. We would be concerned that promoting the cost per QALY would shift the reader's focus to the cost utility analysis only, regardless of the many caveats we include about the inadequacy of the evidence for long-term costs and utilities.
2.2	Maternal outcomes are not included in the analysis. This exclusion is mentioned but not justified. There are clearly maternal health benefits to identifying and treating women with syphilis, so while this may not be the intended effect of the programme, it is in line with best practice to

	include all relevant outcomes. Maternal QALYs and relevant costs should therefore be modelled as a sensitivity analysis.
	Response: Thank you for your suggestion, we have added an additional sentence in the Methods (page 11, paragraph 2) to justify this. "Parental HRQoL were not considered as this would add complexity to the model. There are no published data on many of the utility scores for each of the pregnancy outcomes, for maternal syphilis diagnosis or receiving a false negative result. These would need to be based on uncertain estimates or from expert opinion due to limited evidence on HRQoL." We agree that there may be maternal health benefits from repeating screening. However, the primary purpose of a repeat screening programme for syphilis would be to prevent adverse pregnancy and new-born outcomes, therefore, we focussed our model on what is most clinically relevant. We believe that assessing maternal QALYs is outside the scope of the model. Taking into account the maternal utilities for each of the different pregnancy outcomes would also add huge complexity and further uncertainty. There is insufficient published evidence around utility scores for each of the pregnancy outcomes and for false positive results etc. Therefore, many of these utilities would contain huge uncertainty as they would be based on uncertain estimates or expert opinion. Furthermore, in the short-term, there would be no difference in utility for women with asymptomatic infection and estimating how many of these women would go on to develop sequalae and the associated utility would be equally complex and uncertain as there are no data.
2.3	Perinatal mortality is assumed to be unaffected by this intervention, and it is unclear whether this assumption is explored in the sensitivity analysis. Other studies have reported a relationship between preterm birth and perinatal mortality (Callaghan et al, "The Contribution of Preterm Birth to Infant Mortality Rates in the United States", 2006), and CS and perinatal mortality (Hersh, 2018). The potential contribution of perinatal mortality to overall cost-effectiveness should be explored in the sensitivity analysis.
	Response: Thank you for this comment – it has highlighted the fact that IUFD and neonatal death are not clearly defined in the paper. Perinatal mortality is typically defined as 'stillbirths plus early neonatal deaths (under 7 days)' Patient.info. The model takes into account perinatal mortality since it takes into account both intrauterine fetal demise (IUFD) i.e. stillbirths and neonatal deaths. Action: Neonatal death and IUFD are now defined in the footnotes of Table 1 (page 9). "IUFD refers to the death of a baby in the uterus at ≥20 weeks gestation i.e. stillbirth. Neonatal death refers to the death of a baby within the first 28 days after birth."

2.4	Table 1. The sources of evidence for the diagnostic accuracy (DA) are inadequate. Reference 22 is an analysis of new-born rather than prenatal testing and makes no reference to either EIA or congenital syphilis. Reference 23 is a case series. Neither paper reports DA values for syphilis in pregnant patients. The diagnostic accuracy reported is exceptionally high with very narrow credible intervals. A strong justification is required, or alternatively a wider credible interval should be used for both the one-way sensitivity analysis and the PSA. The DA reported here is substantially higher than in both the Hersh and Albright studies. The tests may be different, but this emphasises the lack of evidence supporting such high reported DA.
	Response: Thank you for flagging this. There were some issues introduced with the referencing for this table when it was turned into a pdf, which we have now fixed. Table 1 references two papers related to DA. Reference 22 is to Morshed (2015). This paper is a mini review of recent trends in the serologic diagnosis of syphilis. It outlines in Figure 1, three testing algorithms for syphilis. The paper states that the first algorithm (typically used in the US) has a sensitivity of 75.8% (p.140) whereas the other two algorithms (akin to the algorithm used in the UK) have sensitivity 99.38% - 99.85% and specificity 99.98% - 100%. This goes some way to explain why the DA used in the US papers are much lower. Reference 23 was supposed to be Binniker (2011), a paper which compares seven different treponemal assays. However, we have now replaced this with reference to data provided by the lead microbiologist at the National Reference Lab who listed the 5 assays currently used by laboratories in the UK. The specificity and sensitivity of these assays was used to inform the model inputs and the low and high values used in the sensitivity analysis. There is no published evidence of the DA of testing for syphilis taking into account the laboratory tests plus the decision making by the clinician as to whether women are categorised as having active syphilis that requires treatment or not. We favoured using high DA based on the performance of the assays as this would bias the outcome to being more cost effective. Action: Table 1. Reference 22 has been corrected and reference 23 changed. Table 1 now states that these DA inputs are ‘Based on the average test sensitivity of five EIA assays used in the UK.’ We have included additional text in the limitations section (Page 19, paragraph 3): “Since there are no published estimates of the diagnostic accuracy (DA) of the syphilis screening process, accounting for the diagnostic accuracy of laboratory assays and the diagnosis decision making by clinicians, average sensitivity and specificity of EIA assays used in UK laboratories were used. The DA values used here are considerably higher than those used in the US models, where a different testing algorithm is used and would, if anything, bias repeat screening results towards being more cost-effective.”
2.5	Table 1. There appear to be multiple issues with the references used in Table 1. Reference 21 is used to support the probability of developing syphilis between the 1st and 3rd trimesters, but in

	fact refers to a costing study of lifetime cerebral palsy costs. Likewise, the parameters that refer to a cerebral palsy costing study instead point to Reference 27, which is an incidence study. Finally, the Hersh study is cited twice, as a full paper (Ref 16) and a conference abstract (Ref 26).
	Action: This error in Table 1 was introduced when the document was turned into a pdf. All the references for Table 1 have been checked and are now cited correctly. The probability of developing syphilis: Townsend (2017); Shiber (2013). Cerebral palsy costs: Kruse (2009). The Hersh paper is now correctly cited throughout (i.e. only the paper is cited, not the conference abstract).
2.6	A utility value of 1 is used for non-CS infants. This is unrealistic, particularly as children become adults, and then elderly adults. UK population norms (eg, Kind, 1999) should be used to obtain age-adjusted utilities which can then be applied to calculate expected QALYs.
	Response + Action: We understand the concern, however, EQ-5D value for CS was not available and there are no data on how utility changes over time for infants born with CS, therefore, we assume that the difference in utility between these two groups remains the same throughout childhood and adulthood (i.e. an adult who had congenital syphilis at birth compared to an adult who did not). We have added an explanation and have clarified this in the methods section: The methods section (top page 11) now states that this utility value is a relative value: infants born with CS (0.74) and infants with no CS (1.00). We have added an additional sentence in the methods section (page 11, paragraph 1): “In the absence of data on changes to utility for infants born with CS, it was assumed that the difference in utility between infants born with CS and infants with no CS remained constant through childhood and adulthood.”
2.7	Discussion. The authors reference two previous studies, reporting that Hersh shows the intervention to be cost-saving, while Albright reports it to not be cost-effective. However, Albright actually defines cost-effectiveness in terms of whether the intervention is cost-saving. What these two studies actually say is that the intervention may or may not be cost-saving. The authors should rephrase this sentence to reflect this.
	Action: Thank you for flagging this. The wording in the discussion (page 17, paragraph 2) has been changed to clarify the main findings from these two US health economic evaluations:

	"Only two previous economic evaluations have assessed universal repeat syphilis screening in pregnancy compared to single screening in early pregnancy, both in the United States (US). Albright *et al.* reported that repeat third-trimester screening would prevent 60 CS cases per 4 million women costing $419,842 per case avoided, concluding that repeat screening was not cost-effective [17]. Hersh *et al.* found that repeat screening would prevent 41 CS cases per 3.9 million women and result in total cost savings of $52 million [16]. Neither study accounted for late presentation to antenatal care - syphilis prevalence and incidence were considerably higher than in the UK as were healthcare costs."

VERSION 2 – REVIEW

REVIEWER	Catherine Albright University of Washington
REVIEW RETURNED	11-Sep-2020

GENERAL COMMENTS	Overall this is a well-done cost analysis. They have responded appropriately to the revision requests.

REVIEWER	Lucy Abel University of Oxford, UK
REVIEW RETURNED	26-Aug-2020

GENERAL COMMENTS	I thank the authors for taking the time to address my comments, which they have responded to fully. I believe the paper is in good shape for publication.