PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

Title (Provisional)

Health economics evaluation of diagnostic strategies for gastroesophageal reflux disease with reflux symptoms in China: a modelling study

Authors

Yan, Xiaxiao; Li, Xiaoqing; Chen, Yang; Ouzhu, Meiduo; Guo, Ziqi; Lyu, Chengzhen; Yang, Daiyu; Chen, Hongda; Xie, Feng; Wu, Dong

VERSION 1 - REVIEW

Reviewer	**1**
Name	**Ugliono, Elettra**
Affiliation	**University of Turin, Department of Surgical Sciences**
Date	**01-Oct-2024**
COI	**None**

I would like to thank for the possibility to review this interesting article entitled “Health Economics Evaluation of Diagnostic Strategies for Gastroesophageal Reflux Disease with Reflux Symptoms in China”.

The article is well designed and well written and focuses on an interesting topic.

The economic analysis is well conducted, however I feel like an effort towards a more to improve the understanding of the topic should be done.

I have some concerns since the population in study is about patients with typical GERD symptoms; as a wide body of literature show, patients with typical symptoms are generally associated with acid reflux at pH-impedance esophageal monitoring and, therefore, respond well to PPI treatment.

If I don’t go wrong, however, in the decision tree analysis only patients with alarm symptoms and patients without alarm symptoms but ineffective PPI response are considered to undergo endoscopy, while patients without alarm symptoms, no CA and good response to PPI treatment (which should be the majority of GERD patients with typical symptoms) are not supposed to undergo endoscopy. Therefore I’m not sure on how the effectiveness of the

endoscopic approaches is evaluated if the majority of the patient population considered don't undergo endoscopy, it could be underestimated.

The economic results should be better explained to enhance generalizability.

The outcome of the study is the rate of GERD diagnosis. The authors state in the results section that (page 5, lines 38-42): "The rate of correct diagnosis of GERD was 0.45 and 0.52 for the stratified strategy and the endoscopy-first strategy, respectively"

The point is: is this difference statistically significant? If there is no statistically significant difference, then a cost-minimization approach should be used (therefore favoring the stratified endoscopy strategy). If the results are statistically significant, then it should be better clarified which are the benefits of an endoscopy-first strategy: avoidance of further investigations? Avoidance of drug prescription? Early diagnosis?

The authors state that " The ICER comparing the endoscopy-first strategy with the stratified endoscopy strategy was $440.39 per additional correct case of GERD". There is no accepted willingness to pay threshold for ICER, but why should we choose to pay more in endoscopy-first strategy?

This is even more true for cancer diagnosis, the authors state that: "The rates of detecting upper gastrointestinal CA of the two strategies were 0.0088 and 0.0120. The ICER was $8561.34". The difference in cancer detection in statistically significant? How do you justify the higher costs of endoscopy-first strategy? Early diagnosis? Increased overall survival? Since the big question of diagnostic procedure is the possibility to achieve an modification in the clinical history of the pathology that is clinically relevant.

Furthermore, do the difference in incidence for instance Eastern/Western Countries in the cancer detection do make a difference? Please explain in the discussion

Minor revisions: improve the legends of the figures

Reviewer **2**

Name **Tang, Yurong**

Affiliation **The First Affiliated Hospital with Nanjing Medical University, Gastroenterology**

Date **13-Nov-2024**

COI **None**

This study compared the costs of two recommended diagnostic processes (American College of Gastroenterology clinical guidelines and Chinese expert consensus) for GERD, and complemented the gap in health economics evidence for the expert consensus of GERD diagnosis in China. But there are still some issues that need to be explained by the author:

1. In the abstract, the conclusion of this study “the use of endoscopy for all patients with reflux symptoms was more effective but with an increased cost compared with the strategy recommended in international guidelines” does not seem to surprise us and is quite understandable. Authors should not only focus on the cost but also the cost-effective result.

2. Based on the results of this study, a more cost-effective strategy should be recommended, or modifications to the details of the diagnostic process be suggested, such as which process is more suitable for different age groups.

3. Introduction, line 17-18 “The gold standard and algorithmic approach for diagnosis are not yet defined”. There are standard criteria for diagnosing GERD, but the recommendations for the diagnostic process are inconsistent with domestic and foreign guidelines.

4. In “Figure 1. Decision tree model”, the “without alarm symptom” group is divided into “non-CA” and “CA” groups, but the author didn’t mention how they divided these patients without endoscopy.

Reviewer	**3**
Name	**Kelson, Zoe**
Affiliation	**University of Exeter, Mathematics**
Date	**23-Jan-2025**
COI	**None**

This study aims to compare two recommended diagnostic processes using a Chinese population-based health economic analysis.

Reviewer comments:

"considered a hypothetical cohort of patients with typical reflux symptoms" [Abstract]

and

"Our analysis considered a hypothetical cohort of patients with typical reflux symptoms (heartburn and regurgitation) in China"

Can the authors please clarify whether they consider the assumptions and bases used to create this hypothetical cohort to be representative?

"a decision tree model to compare the two recommended diagnostic processes "

Appropriate modelling methods have been applied.

Can the model formulation please be detailed in the supplementary material?

"Preference was given to the most recent studies based on the Chinese population. When more than one value of the same parameters was reported in multiple studies, the maximum and minimum values, or baseline±20% if insufficient parameters, were included as

the value range. For unavailable parameters, data were obtained through expert consultation or referred to relevant studies from other countries. All input parameters are listed in Table 1"

Can the authors please clarify if all inputs are sampled from a range (i.e. not just the ones with different values from multiple sources)?

Can sensitivity analyses please be conducted that apply the alternative model parameter value attained from different source(s)?

"The base-case analysis estimated the incremental cost-effectiveness ratio (ICER) between the stratified endoscopy strategy and the endoscopy-first strategy. We used the incremental cost per additional correct diagnosis of GERD."

and

"We also evaluated the incremental cost per additional detection of upper gastrointestinal CA (biopsy-confirmed CA was assigned a value of 1, while other results were 0). "

A suitable model assessment strategy has been performed.

"Sensitivity analysis was performed to evaluate a range of cost and probability estimates on costs and health outcomes over a one-year time horizon from the health care system perspective" [Abstract]

and

"To evaluate the robustness of the results of the decision tree analyses, we explored broad distributions around uncertain parameters using one-way sensitivity analysis. Each parameter varied within the value range to explore the potential factors affecting the optimal strategy, and the results were shown in the tornado diagrams"

Can a probabilistic sensitivity analysis please also be undertaken?

Can what-if scenario analyses please be explored?

"The total expected costs were $122.51 for the stratified endoscopy strategy and $150.12 for the endoscopy-first strategy. The ICER comparing the endoscopy-first strategy with the stratified endoscopy strategy was $440.39 per additional correct case of GERD. The rates of detecting upper gastrointestinal CA of the two strategies were 0.0088 and 0.0120, and the ICER was $8561.34." [Abstract]

and

"Table 2. The total expected costs were $122.51 for the stratified endoscopy strategy and $150.12 for the endoscopy-first strategy. The rate of correct diagnosis of GERD was 0.45 and 0.52 for the stratified strategy and the endoscopy-first strategy, respectively. The ICER comparing the endoscopy-first strategy with the stratified endoscopy strategy was $440.39 per additional correct case of GERD. The rates of detecting upper gastrointestinal CA of the two strategies were 0.0088 and 0.0120. The ICER was $8561.34. A total of 47.4% of patients

underwent endoscopy, and 25.8% finished reflux monitoring in the stratified endoscopy strategy. In the other strategy, where all patients underwent endoscopy, 25.7% needed reflux monitoring."

Can uncertainty intervals please be reported?

"The one-way sensitivity analysis is shown in Figure 2. The most sensitive parameters were the probability of RE in patients without alarm symptoms, the probability of true positives in the PPI test, and the cost of endoscopy. "

Can the interpretation of Figure 2 please be expanded on in the Results text?

"This study had some limitations. One of the significant limitations is the one-year time horizon. The study did not measure the costs and outcomes related to treatment, survival, and disability. Costeffectiveness was not measured in terms of cost per disability-adjusted life year averted, which is a more robust measure of cost-effectiveness. Moreover, our model is structured based on several assumptions and parameter estimates. Parameter estimates were extracted from multiple sources with different evidence quality. Considering that the prevalence also varies considerably in various regions of China, these results are bound to change with changes in prevalence rates from other populations."

The authors have provided a discussion on the study limitations.

Thanks for providing a copy of the CHEERS checklist.

VERSION 1 - AUTHOR RESPONSE

Reviewer Reports:

Reviewer: 1

Dr. Elettra Ugliono, University of Turin, Politecnico di Torino

Comments to the Author:

I would like to thank for the possibility to review this interesting article entitled “Health Economics Evaluation of Diagnostic Strategies for Gastroesophageal Reflux Disease with Reflux Symptoms in China”.

The article is well designed and well written and focuses on an interesting topic.

The economic analysis is well conducted, however I feel like an effort towards a more to improve the understanding of the topic should be done.

I have some concerns since the population in study is about patients with typical GERD symptoms; as a wide body of literature show, patients with typical symptoms are generally associated with acid reflux at pH-impedance esophageal monitoring and, therefore, respond well to PPI treatment.

If I don’t go wrong, however, in the decision tree analysis only patients with alarm symptoms and patients without alarm symptoms but ineffective PPI response are considered to undergo endoscopy, while patients without alarm symptoms, no CA and good response to PPI treatment (which should be

the majority of GERD patients with typical symptoms) are not supposed to undergo endoscopy. Therefore I'm not sure on how the effectiveness of the endoscopic approaches is evaluated if the majority of the patient population considered don't undergo endoscopy, it could be underestimated.

Reply: Thank you for your question. In fact, our decision tree model includes two "branches" and compares these two strategies. In addition to the strategy you mentioned, which is to perform endoscopy only on patients with alarm symptoms, the other strategy, currently recommended by the expert consensus in China, is to perform endoscopy on all patients with reflux symptoms. Our decision tree compares these two strategies. When determining the parameters required for the model, we distinguished between patients with alarm symptoms, those without alarm symptoms, and all patients with reflux symptoms (as shown in Table 1), and investigated the detection rates of various lesions by endoscopy under different circumstances. Of course, there is currently insufficient research data based on the Chinese population for these parameters, so we can only make rough estimates and calculations. We hope that future Chinese researchers will conduct larger-scale epidemiological and endoscopic detection surveys. I hope my response has adequately addressed your question.

The economic results should be better explained to enhance generalizability.

The outcome of the study is the rate of GERD diagnosis. The authors state in the results section that (page 5, lines 38-42): "The rate of correct diagnosis of GERD was 0.45 and 0.52 for the stratified strategy and the endoscopy-first strategy, respectively"

The point is: is this difference statistically significant? If there is no statistically significant difference, then a cost-minimization approach should be used (therefore favoring the stratified endoscopy strategy). If the results are statistically significant, then it should be better clarified which are the benefits of an endoscopy-first strategy: avoidance of further investigations? Avoidance of drug prescription? Early diagnosis?

The authors state that " The ICER comparing the endoscopy-first strategy with the stratified endoscopy strategy was $440.39 per additional correct case of GERD". There is no accepted willingness to pay threshold for ICER, but why should we choose to pay more in endoscopy-first strategy?

This is even more true for cancer diagnosis, the authors state that: "The rates of detecting upper gastrointestinal CA of the two strategies were 0.0088 and 0.0120. The ICER was $8561.34". The difference in cancer detection in statistically significant? How do you justify the higher costs of endoscopy-first strategy? Early diagnosis? Increased overall survival? Since the big question of diagnostic procedure is the possibility to achieve an modification in the clinical history of the pathology that is clinically relevant.

Reply: As a health economics study, we have chosen to compare the ICER between the two strategies. When the effectiveness in the ICER is measured through QALYs, we can compare the ICER with the WTP. Unfortunately, for effectiveness measures other than QALYs, there are no corresponding WTP thresholds available. As you pointed out, these also deeply troubled us during the research process, but due to the imperfection of the methodology, we finally decided to only describe the results without giving a conclusion (as mentioned in the "base-case analysis" section of the methodology). Many researchers have proposed the need for diversified or tailored cost-effectiveness thresholds based on study designs, but there is currently no suitable methodological support for this. We hope that in the future, a method for determining WTP will be established, enabling us to further

compare the ICER results from this study with WTP.

One of the major limitations of this study is that it only analyzes diagnostic relief. Clearly, discussing the outcomes and prognosis (QALYs) of cancer patients is essential. However, there is currently insufficient epidemiological data based on the Chinese population to support further research, which is an area we aim to explore and advance in the future.

Furthermore, do the difference in incidence for instance Eastern/Western Countries in the cancer detection do make a difference? Please explain in the discussion

Reply: Thank you for your suggestion. The endoscopy-first strategy is recommended only in China, due to the lower cost of endoscopic examinations and the higher incidence of upper gastrointestinal tumors. Therefore, our study was designed from the outset to be limited to the Chinese population. Our results cannot support the discussion of the same two strategies in the context of Western countries. However, in different countries, varying endoscopic findings would certainly influence the outcomes of health economic evaluations. We look forward to similar studies and analyses being conducted in Western countries.

Minor revisions: improve the legends of the figures

Reply: Thank you for your suggestion. We have improved the legends of the figures. If there are any other detailed suggestions for the figures, please let us know.

Reviewer: 2

Dr. Yurong Tang, The First Affiliated Hospital with Nanjing Medical University

Comments to the Author:

This study compared the costs of two recommended diagnostic processes (American College of Gastroenterology clinical guidelines and Chinese expert consensus) for GERD, and complemented the gap in health economics evidence for the expert consensus of GERD diagnosis in China. But there are still some issues that need to be explained by the author:

1. In the abstract, the conclusion of this study “the use of endoscopy for all patients with reflux symptoms was more effective but with an increased cost compared with the strategy recommended in international guidelines” does not seem to surprise us and is quite understandable. Authors should not only focus on the cost but also the cost-effective result.

Reply: We sincerely appreciate the reviewer's valuable comment. We agree that the conclusion in the abstract, as pointed out by the reviewer, may appear unsurprising and overly simplistic. As a health economics study, we have chosen to compare the ICER between the two strategies. When the effectiveness in the ICER is measured through QALYs, we can compare the ICER with the WTP. Unfortunately, for effectiveness measures other than QALYs, there are no corresponding WTP

thresholds available, which limits us to describing the results without drawing definitive conclusions (as mentioned in the "base-case analysis" section of the methodology). As for the conclusion, it is not surprise and in line with our expectations. But we think it is necessary to use decision tree model analysis to verify this conjecture.

2. Based on the results of this study, a more cost-effective strategy should be recommended, or modifications to the details of the diagnostic process be suggested, such as which process is more suitable for different age groups.

Reply: We sincerely thank the reviewer for this constructive suggestion. We agree that recommending a more cost-effective strategy or proposing modifications to the diagnostic process, such as tailoring strategies for different age groups, would be valuable. However, due to the lack of a recognized willingness-to-pay (WTP) threshold in China, our results cannot draw definitive conclusions or provide specific recommendations. Instead, our findings can serve as a reference for future expert consensus or guideline development. Additionally, the currently available epidemiological data based on the Chinese population are insufficient to support detailed analyses for different age groups. We acknowledge this limitation and hope to conduct further in-depth investigations and studies across different age groups and regions in the future. Thank you for your insightful comment, which highlights an important direction for our subsequent research.

3. Introduction, line 17-18 “The gold standard and algorithmic approach for diagnosis are not yet defined”. There are standard criteria for diagnosing GERD, but the recommendations for the diagnostic process are inconsistent with domestic and foreign guidelines.

Reply: We sincerely thank the reviewer for pointing out this issue. We agree that the statement “The gold standard and algorithmic approach for diagnosis are not yet defined” could be misleading, as there are indeed standard criteria for diagnosing GERD, even though recommendations for the diagnostic process may vary between domestic and international guidelines. To address this, we have removed the aforementioned statement to avoid any confusion. We appreciate the reviewer's careful attention to detail, which has helped improve the clarity and accuracy of our manuscript. Thank you for this valuable feedback.

4. In “Figure 1. Decision tree model”, the “without alarm symptom” group is divided into “non-CA” and “CA” groups, but the author didn’t mention how they divided these patients without endoscopy.

Reply: We sincerely thank the reviewer for raising this important question. The division of the “without alarm symptom” group into “non-CA” and “CA” groups in Figure 1 does not imply that these patients were categorized using endoscopy. Instead, these groups represent two distinct diagnostic outcomes (terminal nodes) within the model, corresponding to patients without cancer (non-CA) and those with cancer (CA). This division was made to facilitate the input of parameters and the computational process of the model. Actually, both branches subsequently proceed to the PPI test diagnostic pathway. We acknowledge that this aspect of the model may not be immediately clear and appreciate the reviewer's attention to this detail. If the reviewer has any suggestions for improving the clarity of this representation, we would be very grateful to receive them. Thank you for your valuable feedback.

Reviewer: 3

Prof. Zoe Kelson, University of Exeter

Comments to the Author:

This study aims to compare two recommended diagnostic processes using a Chinese population-based health economic analysis.

Reviewer comments:

"considered a hypothetical cohort of patients with typical reflux symptoms" [Abstract]

And "Our analysis considered a hypothetical cohort of patients with typical reflux symptoms (heartburn and regurgitation) in China"

Can the authors please clarify whether they consider the assumptions and bases used to create this hypothetical cohort to be representative?

Reply: We sincerely thank the reviewer for this important question. The hypothetical cohort of patients with typical reflux symptoms (heartburn and regurgitation) in our decision tree model was constructed based on the target populations specified in both international and domestic guidelines or expert consensus. The two diagnostic strategies compared in our model are derived from these recommendations, and the applicable population (or clinical scenario) for these strategies is explicitly defined as patients presenting with reflux symptoms. Therefore, our hypothetical cohort aligns with the populations described in these guidelines and is representative of the clinical context in which these strategies are intended to be applied. We appreciate the reviewer's attention to this detail and hope this clarification addresses their concern. Thank you for your valuable feedback.

"a decision tree model to compare the two recommended diagnostic processes "

Appropriate modelling methods have been applied.

Can the model formulation please be detailed in the supplementary material?

Reply: We sincerely thank the reviewer for their comment. In the Methods section, we have provided a detailed explanation of the model construction, including the rationale behind the model, the selection of parameters for model computation, and the assignment of diagnostic outcomes. Additionally, Figure 1 illustrates the complete decision tree model. At this stage, we do not believe there is additional information that needs to be included in the supplementary material. However, if the reviewer identifies any specific details or aspects that require further clarification, we would be more than happy to address them. Thank you for your valuable feedback.

"Preference was given to the most recent studies based on the Chinese population. When more than one value of the same parameters was reported in multiple studies, the maximum and minimum values, or baseline±20% if insufficient parameters, were included as the value range. For unavailable parameters, data were obtained through expert consultation or referred to relevant studies from other countries. All input parameters are listed in Table 1"

Can the authors please clarify if all inputs are sampled from a range (i.e. not just the ones with different values from multiple sources)?

Reply: We sincerely thank the reviewer for this question. As shown in Table 1, all parameters and ranges in the "Clinical probability" section were derived from existing studies (with corresponding references listed in Table 1). For the "Cost" section, the parameters were based on the pricing of medications and medical services in Beijing (as mentioned in the Methods section), and the ranges were calculated using a ±20% variation. We hope this clarification addresses the reviewer's concern. Thank you for your valuable feedback.

Can sensitivity analyses please be conducted that apply the alternative model parameter value attained from different source(s)?

Reply: We sincerely thank the reviewer for this valuable suggestion. To determine the ranges for sensitivity analysis, we employed several commonly used methods, including: 1) reviewing relevant literature, 2) consulting expert opinions, 3) using the 95% confidence intervals of parameters, 4) considering clinically meaningful ranges (e.g., variables reflecting clinical conditions), 5) applying a fixed percentage variation (e.g., ±20%) around the baseline values, and 6) defining probability distributions within the parameter ranges. The choice of method depended on the nature of the parameter and the availability of data from the literature. By using these approaches, we aimed to explore a wide range of parameter values to assess the robustness of our results and the reliability of our conclusions. We believe this comprehensive approach to sensitivity analysis addresses the reviewer's concern. Thank you for your insightful comment.

"The base-case analysis estimated the incremental cost-effectiveness ratio (ICER) between the stratified endoscopy strategy and the endoscopy-first strategy. We used the incremental cost per additional correct diagnosis of GERD."

and

"We also evaluated the incremental cost per additional detection of upper gastrointestinal CA (biopsy-confirmed CA was assigned a value of 1, while other results were 0). "

A suitable model assessment strategy has been performed.

"Sensitivity analysis was performed to evaluate a range of cost and probability estimates on costs and health outcomes over a one-year time horizon from the health care system perspective" [Abstract]

and

"To evaluate the robustness of the results of the decision tree analyses, we explored broad distributions around uncertain parameters using one-way sensitivity analysis. Each parameter varied within the value range to explore the potential factors affecting the optimal strategy, and the results were shown in the tornado diagrams"

Can a probabilistic sensitivity analysis please also be undertaken?

Reply: We sincerely thank the reviewer for this valuable suggestion. We agree that probabilistic sensitivity analysis would provide a more comprehensive understanding of the uncertainty in our model. However, due to the lack of a recognized WTP threshold, we are unable to conduct a

probabilistic sensitivity analysis at this stage. Instead, we have performed one-way sensitivity analyses to explore the impact of individual parameter variations on the results. While this approach may not capture the full range of uncertainty, we believe it offers preliminary insights and highlights potential directions for future research. We hope that as more data and methodological support become available, probabilistic sensitivity analysis can be incorporated in subsequent studies. Thank you for your constructive feedback.

Can what-if scenario analyses please be explored?

Reply: We sincerely thank the reviewer for this insightful suggestion. We agree that exploring scenario analyses, particularly in the context of healthcare insurance payment, would be highly meaningful. However, due to the vast population size in China and the diversity and complexity of healthcare insurance payment policies, we are currently unable to conduct such an in-depth analysis. Nevertheless, we will continue to monitor this area closely and engage in discussions with methodological experts and policymakers to explore the feasibility of such analyses in future research. We appreciate the reviewer's valuable input and hope to address this aspect in subsequent studies as more data and methodological support become available. Thank you for your constructive feedback.

"The total expected costs were $122.51 for the stratified endoscopy strategy and $150.12 for the endoscopy-first strategy. The ICER comparing the endoscopy-first strategy with the stratified endoscopy strategy was $440.39 per additional correct case of GERD. The rates of detecting upper gastrointestinal CA of the two strategies were 0.0088 and 0.0120, and the ICER was $8561.34." [Abstract]

and

"Table 2. The total expected costs were $122.51 for the stratified endoscopy strategy and $150.12 for the endoscopy-first strategy. The rate of correct diagnosis of GERD was 0.45 and 0.52 for the stratified strategy and the endoscopy-first strategy, respectively. The ICER comparing the endoscopy-first strategy with the stratified endoscopy strategy was $440.39 per additional correct case of GERD. The rates of detecting upper gastrointestinal CA of the two strategies were 0.0088 and 0.0120. The ICER was $8561.34. A total of 47.4% of patients underwent endoscopy, and 25.8% finished reflux monitoring in the stratified endoscopy strategy. In the other strategy, where all patients underwent endoscopy, 25.7% needed reflux monitoring."

Can uncertainty intervals please be reported?

Reply: We sincerely thank the reviewer for this valuable suggestion. In response to your comment, we have added a supplementary file that includes Table S1 (Incremental Cost-Effectiveness Ratios (ICERs) Under Variations in Parameter Values). This table provides a detailed presentation of the ICERs under different parameter variations, which helps to illustrate the uncertainty in our model results. We hope this additional information addresses your concern and enhances the transparency and robustness of our findings. Thank you for your constructive feedback.

"The one-way sensitivity analysis is shown in Figure 2. The most sensitive parameters were the probability of RE in patients without alarm symptoms, the probability of true positives in the PPI test,

and the cost of endoscopy. "

Can the interpretation of Figure 2 please be expanded on in the Results text?

Reply: We sincerely thank the reviewer for this valuable suggestion. In response to your comment, we have expanded the interpretation of Figure 2 in both the Results and Discussion sections. We hope these additions enhance the clarity and depth of our analysis. Thank you for your constructive feedback.

"This study had some limitations. One of the significant limitations is the one-year time horizon. The study did not measure the costs and outcomes related to treatment, survival, and disability. Costeffectiveness was not measured in terms of cost per disability-adjusted life year averted, which is a more robust measure of cost-effectiveness. Moreover, our model is structured based on several assumptions and parameter estimates. Parameter estimates were extracted from multiple sources with different evidence quality. Considering that the prevalence also varies considerably in various regions of China, these results are bound to change with changes in prevalence rates from other populations."

The authors have provided a discussion on the study limitations.

Thanks for providing a copy of the CHEERS checklist.

VERSION 2 - REVIEW

Reviewer	**3**
Name	**Kelson, Zoe**
Affiliation	**University of Exeter, Mathematics**
Date	**06-Mar-2025**
COI	

Thanks to the authors for responding to each comment in turn, providing clarification and their rationale, and revising the manuscript where required.

Author response:

"We agree that probabilistic sensitivity analysis would provide a more comprehensive understanding of the uncertainty in our model. However, due to the lack of a recognized WTP threshold, we are unable to conduct a probabilistic sensitivity analysis at this stage. Instead, we have performed one-way sensitivity analyses to explore the impact of individual parameter variations on the results. While this approach may not capture the full range of uncertainty, we believe it offers preliminary insights and highlights potential directions for future research. We hope that as more data and methodological support become available, probabilistic sensitivity analysis can be incorporated in subsequent studies."

Reviewer requests:

The lack of a recognised WTP threshold preventing a PSA is appreciated.

Can the authors please incorporate this response to reviewer within the main article?

Can the authors please perform bootstrapping to produce confidence intervals for the model outputs?

VERSION 2 - AUTHOR RESPONSE

Reviewer Report:

Reviewer: 3

Prof. Zoe Kelson, University of Exeter

Comments to the Author:

Thanks to the authors for responding to each comment in turn, providing clarification and their rationale, and revising the manuscript where required.

Author response:

"We agree that probabilistic sensitivity analysis would provide a more comprehensive understanding of the uncertainty in our model. However, due to the lack of a recognized WTP threshold, we are unable to conduct a probabilistic sensitivity analysis at this stage. Instead, we have performed one-way sensitivity analyses to explore the impact of individual parameter variations on the results. While this approach may not capture the full range of uncertainty, we believe it offers preliminary insights and highlights potential directions for future research. We hope that as more data and methodological support become available, probabilistic sensitivity analysis can be incorporated in subsequent studies."

Reviewer requests:

The lack of a recognised WTP threshold preventing a PSA is appreciated.

Can the authors please incorporate this response to reviewer within the main article?

Reply: We sincerely appreciate the reviewer's valuable suggestion. As requested, we have now incorporated the additional content (our previous response) into the main article to improve clarity and completeness. Thank you for your thoughtful feedback, which has strengthened our paper.

Can the authors please perform bootstrapping to produce confidence intervals for the model outputs?

Reply: We sincerely appreciate the reviewer's insightful suggestion regarding bootstrapping to estimate confidence intervals for the model outputs. However, as our study is currently based on a theoretical framework without patient-level real-world data, performing a conventional bootstrapping analysis (which requires resampling from empirical datasets) is not feasible at this stage. To ensure transparency, we have explicitly acknowledged this limitation in the revised manuscript (Limitations section, last point) and highlighted the need for future validation using real-world data, which would enable robust uncertainty quantification (e.g., via bootstrapping or Bayesian methods). We agree that such analyses would strengthen the model's credibility and are committed to addressing this in

subsequent work. Thank you for this constructive feedback, which underscores an important direction for improving our model's applicability.
