



CORRESPONDENCE

**REVISED** **Interpreting cost-effectiveness ratios in a cost-effectiveness analysis of risk-tailored prostate screening: A critique of Callender *et al.* [version 2; peer review: 2 approved, 1 approved with reservations]**

James F. O'Mahony

Centre for Health Policy and Management, University of Dublin, Dublin, D02 PN40, Ireland

**V2** **First published:** 13 May 2020, 3:23  
<https://doi.org/10.12688/hrbopenres.13043.1>  
**Latest published:** 20 Oct 2020, 3:23  
<https://doi.org/10.12688/hrbopenres.13043.2>

**Abstract**

Callender *et al.* recently published a model-based cost-effectiveness analysis of a risk-tailored approach to prostate cancer screening. It considers the costs and effects of prostate cancer screening offered to all men aged 55-69 without any risk selection and, alternatively, over a range of risk-tailored strategies in which screen eligibility is determined by a varying threshold of disease risk. The analysis finds that the strategy of screening men once they reach a 10-year absolute risk of disease of 5% or more is cost-effective in a UK context. I believe there are several problems with the study, mostly stemming from an incorrect interpretation of the cost-effectiveness estimates. I show that one reinterpretation of their results indicates that screening is much less cost-effective than the original analysis suggests, indicating that screening should be restricted to a much smaller group of higher risk men. More broadly, I explain the challenges of attempting to meaningfully reinterpret the originally published results due to the simulation of non-mutually exclusive intervention strategies. Finally, I consider the relevance of considering sufficient alternative screening intensities. This critique highlights the need for appropriate interpretation of cost-effectiveness results for policymakers, especially as risk stratification within screening becomes increasingly feasible.

**Keywords**

Cost-effectiveness analysis, cancer screening, incremental-cost effectiveness ratio, prostate screening, PSA

**Open Peer Review****Reviewer Status** ? ✓ ✓

	Invited Reviewers		
	1	2	3
<b>version 2</b> (revision) 20 Oct 2020		✓ report	
		↑	
<b>version 1</b> 13 May 2020	? report	✓ report	✓ report

- Paul Carty** , Health Information and Quality Authority, Dublin, Ireland
- Conor Teljeur** , Health Information and Quality Authority, Dublin, Ireland
- Eveline Heijnsdijk**, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands
- Robert Boer** , Allergan plc., Irvine, USA

Any reports and responses or comments on the article can be found at the end of the article.

**Corresponding author:** James F. O'Mahony ([jfomahon@tcd.ie](mailto:jfomahon@tcd.ie))

**Author roles: O'Mahony JF:** Conceptualization, Formal Analysis, Investigation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** Health Research Board Ireland [EIA-2017-054].

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2020 O'Mahony JF. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** O'Mahony JF. **Interpreting cost-effectiveness ratios in a cost-effectiveness analysis of risk-tailored prostate screening: A critique of Callender *et al.* [version 2; peer review: 2 approved, 1 approved with reservations]** HRB Open Research 2020, 3:23 <https://doi.org/10.12688/hrbopenres.13043.2>

**First published:** 13 May 2020, 3:23 <https://doi.org/10.12688/hrbopenres.13043.1>

**REVISED Amendments from Version 1**

The original version of the manuscript has been updated in light of the reviewer comments. Principally, the changes are to better acknowledge caveats made by the authors of the original study regarding their results. This relates to the potential relevance of multiple screening intensities, but that the necessary data to support the analysis of such policies is currently lacking. More generally, I have added a discussion that acknowledges the tension between pragmatic modelling of a restricted set of policy choices within the constraints of currently available data as opposed to a theoretical ideal of a complete and comprehensive comparison of strategies for multiple cohorts that is only possible with extensive data. I have also made changes to the manuscript in order to make the constructive intention of the critique clear in order to address comments that the tone of the original submission was unnecessarily negative. I have also given my reply to the reviewer comments in order to further explain the changes made and to acknowledge the helpful direction offered by the reviewers.

**Any further responses from the reviewers can be found at the end of the article**

**Introduction**

Callender *et al.*<sup>1</sup> recently published a cost-effectiveness analysis (CEA) of a risk-tailored approach to prostate cancer screening. I believe the study's results are not interpreted appropriately and cannot be considered a reliable guide to prostate screening policy. This commentary explains the problems with the results, examines if they can be usefully reinterpreted, and more generally, attempts to elucidate the issues regarding risk-group selection and the interpretation of cost-effectiveness estimates. The purpose of this commentary is, through critical examination of the Callender *et al.*, to offer guidance to research groups conducting such modelling on how their analyses can best answer policy questions. Secondly, it aims to help readers of such research interpret published estimates.

Callender *et al.*'s<sup>1</sup> analysis estimates the total net costs and quality-adjusted life-years (QALYs) of alternative screening approaches. They examine prostate-specific antigen (PSA) based testing every four years between ages 55–69. They consider this strategy when applied to all men within that age range (described as age-based screening) and alternatively, the same strategy starting only when men meet a range of alternative prostate cancer risk thresholds (described as precision screening). They consider 17 alternative risk thresholds ranging from 2% to 10% 10-year absolute risk (10y-AR) in 0.5% increments. Men can reach these thresholds at different ages. This means the men with the greatest total lifetime risk reach any given threshold at an earlier age and the proportion of men having reached any given threshold increases with age. Therefore, relaxing the risk eligibility threshold simultaneously expands the pool of screened men and lowers the age of screening initiation in those screened.

The study reports incremental cost-effective ratios (ICERs) calculated by comparing the total costs and health effects of

both age-based screening and the range of precision screening strategies to no screening. The reported ICERs range from £14,862/QALY for the most conservative risk-based strategy that restricts screening to those with a 10y-AR of at least 10% to £34,952/QALY for age-based screening. They report a 10y-AR of 5% yields an ICER of £19,598/QALY and note this would be a cost-effective policy in a UK context in which a cost-effectiveness threshold of £20,000/QALY applies. Their results are presented with caveats regarding the structure of the analytical model used and parameter uncertainty. Callender *et al.*'s<sup>1</sup> analysis is a very welcome attempt at examining how prostate screening could be better targeted towards those at greater risk, thereby avoiding unnecessary harms to men at lower risk and enhancing programme cost-effectiveness.

**Critique**

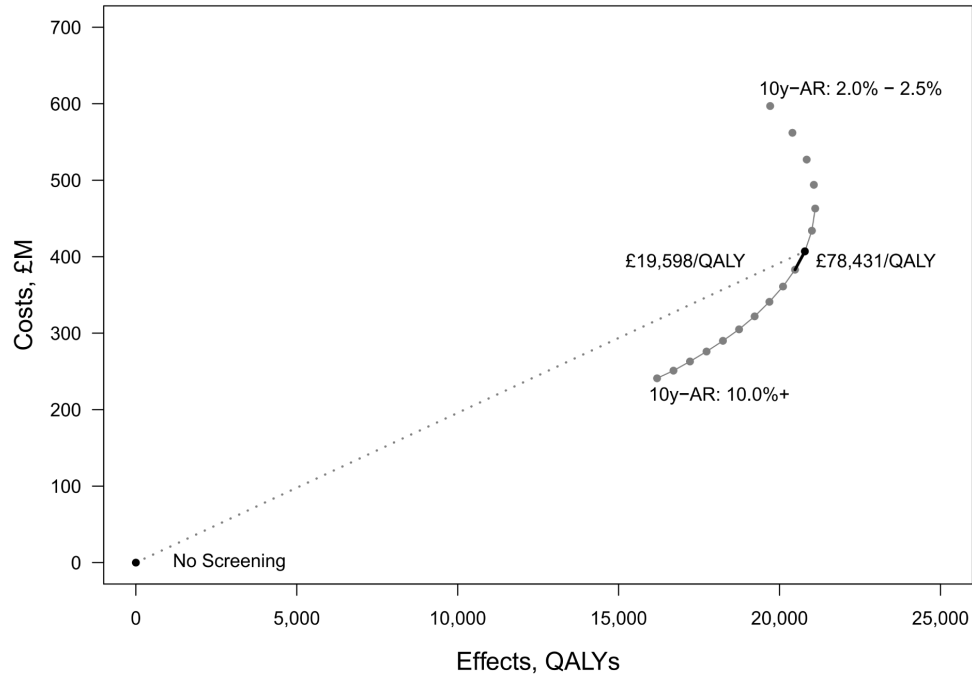
This critique addresses three issues. The first relates to differences between the average and incremental effects of lowering the risk threshold of screening. The second concerns the failure to consider mutually exclusive intervention strategies and the implications this has for finding optimal policies for specific risk-subgroups. The third relates to the relevance of varying the intensity of screening for the estimation of ICERs.

**Incremental analysis of risk threshold adjustment**

Callender *et al.*'s<sup>1</sup> cumulative assessment of the ratio of total costs to total QALYs as the risk threshold is relaxed means the analysis initially includes those men at highest risk who are likely the most cost-effective to screen and then progressively adds those of lower risk that are probably less cost-effective to screen. This cumulative approach to assessing the ratio of total costs to health effects hides the marginal effect of progressively relaxing the risk threshold to include lower risk screenees. The appropriate approach is to examine the incremental change in costs and health effects as the risk threshold is relaxed. Such an incremental analysis identifies what additional health gain is achieved at what additional cost of relaxing the risk threshold relative to the previous, more restrictive threshold.

Reinterpreting the results using an incremental approach indicates that relaxing the risk threshold is less cost-effective than appears under Callender *et al.*'s<sup>1</sup> cumulative analysis. The difference between the cumulative and incremental appraisal is shown in [Figure 1](#). It plots the estimated costs and effects of the 17 precision screening strategies from Callender *et al.*'s<sup>1</sup> analysis. The least costly strategy is for the most conservative risk threshold of 10y-AR of 10%, while the most costly is for the least restrictive 2% 10y-AR threshold. The solid grey line shows the ratio of incremental costs and effects between the risk subgroups. The incremental ratios rise as the risk threshold falls and more men are screened. Beyond a certain threshold total effectiveness falls, implying that some screening becomes harmful to health. That is, it appears that reducing the risk threshold beyond a certain point harms at least some men.

The dotted line in [Figure 1](#) corresponds to the cumulative ratio reported in the by Callender *et al.*<sup>1</sup> as an ICER. In this case, corresponding to screening all men with a 10y-AR of at



**Figure 1.** Cost-effectiveness plane demonstrating the difference between cost-effectiveness ratios calculated on a cumulative and incremental basis.

least 5%, which Callender *et al.*<sup>1</sup> report to be £19,598/QALY. The ratio of the incremental difference of screening men with a 10y-AR of 5.0–5.5% compared to men with a 10y-AR of 5.5% and above is £78,431/QALY. This ratio is shown in Figure 1 as the single thick black segment of the line joining the strategies. Further detail on the estimate is given in Table 1. It includes the ratios reported Callender *et al.*<sup>1</sup> as ICERs (“Reported ICERs”) and an additional cost-effectiveness ratio (CER) calculated as the incremental difference in costs and effects as the risk threshold is incrementally relaxed from 10y-AR of 10% to 2% (“Calculated CERs”). These CERs rise from £14,881/QALY for the highest risk men to £281,553/QALY for men with 10y-AR of 4.0–4.5%. There is no meaningful CER to report below a 10y-AR of 4% once the incremental change in QALY estimates becomes negative.

The policy significance of the difference between the cumulative and incremental analysis can be seen in the context of the UK cost-effectiveness threshold of £20,000/QALY as referenced by Callender *et al.*<sup>1</sup>. Using the originally reported cumulative ratios, the first 11 risk thresholds would be considered cost-effective, as their reported ICERs are within the threshold (shown in italics in Table 1). Using the appropriate incrementally calculated CERs however, only the two most restrictive risk categories fall within the threshold (shown in bold). Accordingly, far fewer men appear cost-effective to screen than originally reported. Furthermore, those screened would start at an older age.

As an aside, it is useful to see how my interpretation of the results presented here is supported by the net monetary benefit (NMB) analysis provided by Callender *et al.*<sup>1</sup> within a

supplementary appendix to their study. The variation of NMB with the risk threshold is presented by them in Figure H (A). It shows that NMB is maximised only when the risk threshold is near its most restrictive around 9.5% to 10% 10y-AR. This contradicts Callender *et al.*'s<sup>1</sup> finding that a 10y-AR of 5% would be cost-effective, as NMB should be maximised at the optimally cost-effective strategy. The observed maximisation of NMB around 9.5% to 10% 10y-AR corresponds with the optimally cost-effective policy (within those simulated) when using the incremental interpretation presented here.

### Non-mutually exclusive strategy comparisons

At this point, I now turn to consider can the reinterpreted CERs reported in Table 1 be used as a reliable guide to screening policy. The above critique draws on the long and widely recognised distinction between the average and incremental cost-effectiveness ratios that is reflected in CEA guidelines<sup>2-6</sup>. Note, however, that I have not described the CERs in Table 1 as ICERs. This is because there are further complications with the incremental reinterpretation of CERs that means they still may not be considered true ICERs and so are not a suitable guide to policy. Moreover, the published results cannot be reinterpreted into ICERs.

The standard interpretation of an ICER is the incremental comparison of costs and effects of mutually exclusive strategies that lie on the efficient frontier of the cost-effectiveness plane<sup>7</sup>. The way the alternative policy choices are specified within Callender *et al.*'s<sup>1</sup> analysis means they fail to constitute mutually exclusive strategies. As mentioned above, the relaxation of the risk threshold simultaneously adds men of lower lifetime risk of disease to the pool of screened men and reduces the age at

**Table 1.** Costs, effects and reported ICERs and reinterpreted cost-effectiveness ratios from Callender *et al.*

Strategy: 10yr-AR risk threshold, %	Effects, QALYs	Costs, £M	Reported ICERs, £/QALY	Calculated CERs, £/QALY
No Screening	0	0	-	-
<b>10.0 +</b>	<b>16,195</b>	<b>241</b>	<b>14,862</b>	<b>14,881</b>
<b>9.5 - 10.0</b>	<b>16,704</b>	<b>251</b>	<b>15,050</b>	<b>19,646</b>
9.0 - 9.5	17,218	263	15,281	23,346
8.5 - 9.0	17,732	276	15,560	25,292
8.0 - 8.5	18,242	290	15,894	27,451
7.5 - 8.0	18,743	305	16,289	29,940
7.0 - 7.5	19,227	322	16,755	35,124
6.5 - 7.0	19,686	341	17,303	41,394
6.0 - 6.5	20,109	361	17,947	47,281
5.5 - 6.0	20,482	383	18,704	58,981
5.0 - 5.5	20,788	407	19,598	78,431
4.5 - 5.0	21,006	434	20,659	123,853
4.0 - 4.5	21,109	463	21,924	281,553
3.5 - 4.0	21,067	494	23,446	SD
3.0 - 3.5	20,844	527	25,290	SD
2.5 - 3.0	20,401	562	27,542	SD
2.0 - 2.5	19,709	597	30,297	SD
Age-based screening	16,416	574	34,952	SD

Source: Callender *et al.* Table 2.

SD: Subject to simple dominance.

first screen for those already within the pool of screened men. Lowering the age of screening initiation in higher risk men is not mutually exclusive of extending screening to lower risk men.

The non-mutually exclusive strategies within Callender *et al.*<sup>1</sup> means the incremental CERs reported in Table 1 correspond to a mixture of different policy choices for men of different lifetime risk. For example, reducing the risk threshold will bring forward the age of first screening for men of high lifetime risk, while it may entail a shift from no screening to one or more screens for a man of lower lifetime risk. While both men may have an equal 10y-AR at the initiation of screening, the differences in both their lifetime risk and the number of lifetime screens they receive mean the policy choices will likely be of different cost-effectiveness. There is no way to disaggregate the published results into a form that would permit policy makers to understand how large any differences in cost-effectiveness may be between men of different lifetime risk or identify what the optimal policies would be.

The implication of non-mutually exclusive strategies for policy is that although most of the incremental CERs in Table 1 are

above the cost effectiveness threshold, it is not necessarily the case that all the corresponding policies are cost-ineffective. For instance, advancing the age at first screening for a man with a high lifetime risk might be less cost-effective than providing one lifetime screen at age 69 to a lower risk man. Accordingly, it would be premature to base policy on the incrementally interpreted CERs I present in Table 1.

To generate mutually exclusive policies within a single analysis, the authors should have held screening intensity constant for all but one sub-group at a time while varying intensity for the remaining sub-group. This policy generation process would have to be repeated for all sub-groups over all alternative strategies considered, resulting in a very large number of mutually-exclusive strategies. A much simpler alternative would be to model the range of screening strategies in separate analyses for each sub-group according to their lifetime risk.

If Callender *et al.*'s<sup>1</sup> analysis were to disaggregate men of different lifetime risks as described, then it would be possible to assess the different simulated screening strategies in each risk subgroup. That would allow the analysis untangle the differences

of intensifying screening in higher risk men from extending screening to lower risk men. Such disaggregation is particularly important when we consider that the reduction in the risk threshold is eventually estimated to harm health. It is important to know which men are harmed by what intensity of screening.

### Screening intensity and ICERs

The third issue of ratio interpretation considered here relates to the range of alternative comparator strategies required to adequately estimate ICERs. ICERs give the ratio of the incremental difference in costs to effects between one strategy relative to the next most effective relevant comparator strategy<sup>7</sup>. For example, the appropriate ICER estimation for a given strategy would typically require comparison to a less intense screening strategy with a lower number of lifetime screens, achieved by comparison to a strategy with either a longer interval or a narrower screening age range<sup>3,7</sup>.

Previous prostate cancer screening CEAs demonstrate the relevance of incremental comparisons between alternative screening frequencies and varied screening age ranges. Heijnsdijk *et al.* show how ICERs rise as the number of lifetime screens increases<sup>8</sup>. While that analysis did not differentiate between risk strata, it does illustrate the relevance of including low intensity strategies as comparators to other strategies with shorter intervals and wider age ranges. They found that in an average risk population in a Dutch context the optimal strategy would be three screens per lifetime at ages 55, 57 and 59. They found no strategy with screening beyond age 59 to be cost-effective, indicating the relevance of considering alternative stopping ages in the case of prostate screening.

This context of previous research and well-established methods guidance tells us that even if Callender *et al.*'s<sup>1</sup> results can be disaggregated into mutually exclusive strategies for separate sub-groups according to lifetime risk, the resulting ratios would still only represent incremental changes to the start age of screening. Ideally, we would like to estimate a range of screening intensities in each sub-group, varying not only the screening start age, but also screening interval and, importantly, the screening stop age. In particular, we would be interested in the potential for low intensity screening to offer at least some prevention to the lowest risk men.

Modelling a wide variety of strategies intervals typically requires simulation of the natural history of disease and the imposition of stage-specific estimates test performance characteristics. Not all models are such "deep" models<sup>9</sup>. Callender *et al.*'s<sup>1</sup> analysis apparently is not one such model and may be restricted to the simulation of quadriennial screening intervals used in the trial that informed the model. While analyses limited to the simulation of one screening interval alone may come with the limitations of not being able to estimate a complete ICER on the basis of a comprehensive set of comparator strategies, these limitations are traded off against the advantage of less reliance on assumptions and reduced parameter uncertainty. There is no clear answer on the optimal balance in this trade-off when informing policy. Nevertheless, results from

analyses without a complete set of comparators can still be useful to policy makers. CERs from analyses with a limited set of strategies that exceed the cost-effectiveness threshold can usefully rule out strategies as cost-ineffective. This is because any strategy with a CER exceeding the threshold within a limited set of comparators can never be cost-effective within a more complete analysis.

It is important to note that Callender *et al.* did clearly acknowledge the relevance of strategies of alternative age ranges and screening frequencies, but explained the data to support a risk stratified analysis was lacking. Accordingly, while the policy choices they simulated may not be those optimal relative to our theoretical understanding of optimally tailored strategies, the latter are, as yet, unsupported by data. Moreover, we should point towards the usefulness of Callender *et al.*'s analysis in informing further research. Their study concludes by stating prospective randomised controlled trials are required to better inform optimal policy. Further analysis of their model could usefully indicate what strategies would be most useful to compare and which parameters estimates are the priority to refine.

### Discussion

The above critique shows that the ratios reported as ICERs by Callender *et al.*<sup>1</sup> should not be used to inform prostate screening policy. Unfortunately there is no way to readily reinterpret the published estimates into policy-relevant guidance. A disaggregation of the model results by the authors could make their analysis more useful to policy makers, but probably only to rule certain strategies out. Further work would be needed to determine if other screening intensities could provide cost-effective screening for lower risk men and what might be the benefit of varying the screening stop age.

Some of the points described in this critique were apparently raised during peer review. The available reviewer comments that accompany the paper show Reviewer 3 noted these issues and explained they could be easily addressed. In reply, the authors give the rationale for their modelling choices, which explain why they did not make the suggested changes. It seems unfortunate that the reviewer's advice was not considered further as it is evident that the authors have done the hard work of constructing, parameterising and implementing their model. It seems a shame that basic changes were not made prior to publication. This highlights the need for both authors and journal editors to ensure that reviewer comments are adequately accounted for. It also serves as a reminder of the fallibility of the peer-review process.

CEA methods for the analysis of screening are well-established and the need for appropriate ICER comparisons between screening strategies of different intensities has been recognised clearly for many years<sup>3,5</sup>. The issues around risk subgroup analysis and how to handle them in screening CEAs have received less attention in the literature<sup>10-12</sup>. Given the increasingly nuanced knowledge of risk subgroups provided by research on genetic and other risk factors, it seems likely that risk-stratified analyses such as Callender *et al.*'s<sup>1</sup> will become more common.



Accordingly, there may be a need for clearer guidelines for analysts.

This commentary is unavoidably critical of the analysis presented by Callender *et al.*<sup>1</sup>. The intention, of course, is not to single out a single study for criticism. Rather, it is to offer constructive guidance to such modellers on how their analyses can be best specified and interrogated. The question of appropriate strategy comparison is not trivial, especially when variation in disease risk is considered, and Callender *et al.* is certainly not the only study to face pitfalls in this respect. Without clear examination of the problems and clear direction of how they should be avoided, subsequent studies will be prone to error. The research question addressed by the authors is important and deserves attention from health economic modellers. It is hoped that the points raised here may inform a revision of the model and the generation of new cost-effectiveness estimates. Such an analysis would usefully inform the tailored provision of prostate cancer prevention according to disease risk, potentially improving the health of men across the UK and beyond.

The criticism made here of Callender *et al.* reflects a broader tension between pragmatic modelling within the constraints of currently available data among a narrow range of policy

alternatives as opposed to the theoretical ideal of modelling many alternative strategies, each considered in disaggregated analyses for separate subgroups. The optimal balance between pragmatism and technical exactitude will always be a matter of debate. We can at least inform this debate by being explicit about the modelling choices made and rationale for them.

## Conclusion

In conclusion, Callender *et al.*'s<sup>1</sup> interpretation of their cost-effectiveness estimates is at odds with accepted CEA practice for several reasons. While a reappraisal of their results suggest that quadrennial screening will likely be cost-ineffective for more men than they suggest, it is not advisable to base any policy recommendations on either the originally published results or the illustrative reinterpretation given here. This example is useful in illustrating some of the methodological considerations surrounding the appropriate handling of risk-subgroup specific cost-effectiveness estimates. Such issues of sub-group specific interpretation of evidence are likely to become more prevalent as increasing knowledge of disease processes permits further disaggregation of screen eligible populations.

## Data availability

All data underlying the results are available as part of the article and no additional source data are required.

## References

- Callender T, Emberton M, Morris S, *et al.*: **Polygenic Risk-Tailored Screening for Prostate Cancer: A Benefit-Harm and Cost-Effectiveness Modelling Study.** *PLoS Med.* 2019; **16**(12): e1002998.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- CADTH: **Guidelines for the Economic Evaluation of Health Technologies: Canada.** 4th Edition: Canadian Agency for Drugs and Technologies in Health; 2017.  
[Reference Source](#)
- Gold MR, Russell LB, Siegel JE, *et al.*: **Cost-effectiveness in health and medicine.** Oxford: Oxford University Press; 1996.  
[Reference Source](#)
- HIQA: **Guidelines for the Economic Evaluation of Health Technologies in Ireland:** Health Information and Quality Authority; 2019.  
[Reference Source](#)
- Neuhauser D, Lewicki AM: **What Do We Gain From the Sixth Stool Guaiac?** *N Eng J Med.* 1975; **293**(5): 226–8.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- NICE: **Guide to the methods of technology appraisal 2013:** National Institute for Health and Care Excellence; 2013.  
[Reference Source](#)
- Drummond MF, Sculpher MJ, Claxton K, *et al.*: **Methods for the economic evaluation of health care programmes.** Oxford: Oxford University Press; 2015.  
[Reference Source](#)
- Heijnsdijk EA, de Carvalho TM, Auvinen A, *et al.*: **Cost-effectiveness of prostate cancer screening: a simulation study based on ERSPC data.** *J Natl Cancer Inst.* 2015; **107**(1): 366.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Knudsen AB, McMahon PM, Gazelle GS: **Use of modeling to evaluate the cost-effectiveness of cancer screening programs.** *J Clin Oncol.* 2007; **25**(2): 203–8.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Dewilde S, Anderson R: **The cost-effectiveness of screening programs using single and multiple birth cohort simulations: a comparison using a model of cervical cancer.** *Med Decis Making.* 2004; **24**(5): 486–92.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hoyle M, Anderson R: **Whose costs and benefits? Why economic evaluations should simulate both prevalent and all future incident patient cohorts.** *Med Decis Making.* 2010; **30**(4): 426–37.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- O'Mahony JF, van Rosmalen J, Zauber AG, *et al.*: **Multicohort models in cost-effectiveness analysis: why aggregating estimates over multiple cohorts can hide useful information.** *Med Decis Making.* 2013; **33**(3): 407–14.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

## Open Peer Review

Current Peer Review Status: ? ✓ ✓

---

### Version 2

Reviewer Report 21 October 2020

<https://doi.org/10.21956/hrbopenres.14261.r28185>

© 2020 Heijnsdijk E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Eveline Heijnsdijk**

Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

I have no further comments.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Evaluation of screening.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

### Version 1

Reviewer Report 23 June 2020

<https://doi.org/10.21956/hrbopenres.14139.r27411>

© 2020 Boer R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Robert Boer** 

Allergan plc., Irvine, CA, USA

The paper by O'Mahony<sup>1</sup> critiques another paper by Callender *et al.*<sup>2</sup> concerning the cost-effectiveness of prostate cancer screening by estimated risk.

I agree with each of the three main three lessons of this critique:



1. Consider incremental cost-effectiveness.
2. Consider mutually exclusive intervention strategies.
3. Include varying intensity of screening.

The author explains the lessons well and was able to show that just following lesson 1) would already make a large difference in the concluded policy recommendation, thus showing that the lessons really matter.

Next to the three lessons that were included in the paper, the author could also have mentioned some other points of critique:

- The Callender *et al.* paper focusses on QALYs as effectiveness measure. Therefore, it seems better to optimize screening based on expected QALYs gained from screening instead of on an estimate of prostate cancer risk. As proxy for QALYs gained, the risk of prostate cancer death seems better than the risk of prostate cancer.
- In order to show the extent of superiority of risk based screening over age based screening, the Callender *et al.* paper should preferably include an efficiency frontier as described by O'Mahony for both age and risk based screening, but at least a scenario that would lead to optimal age based screening at a similar cost as when applying the preferred risk threshold.
- Callendar *et al.*'s estimates for the relationship between risk of cancer, overdiagnosis, and of death, as affected by screening, are quite simplified and should be expected to be considerably less accurate than better models for prostate cancer screening that have been published elsewhere<sup>3</sup>, and therefore less than state of the art. Application of methods that are well below the state of the art, should probably already preclude informing policy, even before considering the presented results.

## References

1. O'Mahony J: Interpreting cost-effectiveness ratios in a cost-effectiveness analysis of risk-tailored prostate screening: A critique of Callender et al. *HRB Open Research*. 2020; **3**. [Publisher Full Text](#)
2. Callender T, Emberton M, Morris S, Eeles R, et al.: Polygenic risk-tailored screening for prostate cancer: A benefit-harm and cost-effectiveness modelling study. *PLoS Med*. **16** (12): e1002998 [PubMed Abstract](#) | [Publisher Full Text](#)
3. National Cancer Institute - Cancer Intervention and Surveillance Modeling Network: Publications by Cancer Site. *Header - Prostate*. [Accessed - 3rd June 2020. [Reference Source](#)

### **Is the rationale for commenting on the previous publication clearly described?**

Yes

### **Are any opinions stated well-argued, clear and cogent?**

Yes

### **Are arguments sufficiently supported by evidence from the published literature or by new data and results?**

Yes

**Is the conclusion balanced and justified on the basis of the presented arguments?**

Yes

**Competing Interests:** Rob Boer is an employee of AbbVie and may hold AbbVie stock.**Reviewer Expertise:** Health economics.**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 09 June 2020

<https://doi.org/10.21956/hrbopenres.14139.r27408>

© 2020 Heijnsdijk E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Eveline Heijnsdijk**

Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

This correspondence criticizes a paper of Callender *et al.* at three points: The first point: using average or incremental cost-effectiveness, is a valid point. The optimal strategy, with the ICER just below the cost-effectiveness threshold, will clearly be different. However, in the original analysis, the decreasing QALYs above the 10-year absolute risk threshold of 4% was also stated. The original paper also clearly describes the harms and benefits, which may be more important than the exact value of cost-effectiveness, because the cost-effectiveness threshold is arbitrarily.

The second and third point (mutually exclusive strategies and range of screening intensities) are also valid points from the cost-effectiveness perspective. This indicates how much additional work is still possible. However, for a first analysis on this subject, the results of the original paper are useful. With all uncertainties in the data, it will be the question if this would be the right moment to perform such an extensive cost-effectiveness analysis.

Concluding, the points of critique in this correspondence are valid. However, the original paper describes a first exploratory analysis on this subject, which can be extended in future.

**Is the rationale for commenting on the previous publication clearly described?**

Yes

**Are any opinions stated well-argued, clear and cogent?**

Yes

**Are arguments sufficiently supported by evidence from the published literature or by new data and results?**

Yes

**Is the conclusion balanced and justified on the basis of the presented arguments?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Evaluation of screening.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 04 June 2020

<https://doi.org/10.21956/hrbopenres.14139.r27409>

© 2020 Teljeur C et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Paul Carty** 

Health Information and Quality Authority, Dublin, Ireland

**Conor Teljeur** 

Health Information and Quality Authority, Dublin, Ireland

The paper represents a critique that addresses three key aspects of an original analysis by Callendar *et al*:

1. the differences between calculating average and incremental cost-effectiveness ratios
2. consideration of policymaking implications when modelled comparators do not represent mutually exclusive strategies in a fully incremental analysis
3. the relevance of varying the intensity of screening for the estimation of ICERs.

We will discuss the merits of these critiques individually before providing an overall assessment of the paper.

- ACERs vs ICERs

The deterministic reinterpretation of the results of the original paper, where the strategies are ranked according to total effects and followed by removal of strategies subject to simple dominance (more costly and less effective than the alternative strategy) are then removed, is straightforward. If conducting a fully incremental analysis then this approach should be adopted when interpreting the CERs. The basis for this approach is well-founded in the economic literature, but is often not followed in published economic evaluations. As asserted in the paper, the NMB presented in the original paper supports this re-calculation of the CERs.

- Mutually exclusive strategies

The second characteristic of a fully incremental analysis is that the interventions under evaluation

are mutually exclusive (i.e. a patient in one patient group can receive only one of the interventions and this intervention is independent of the intervention received in other patient groups). As described in this paper, the original paper did not model the study populations in separate subgroups. As such, the analysis shifts from a comparison of mutually exclusive options to the consideration of independent programmes, each of which incremental costs and effects have already been calculated against the next best alternative.

The reasons for options not being mutually exclusive may relate to data availability, computational burden or simply the decision problem posed by the decision-maker. Therefore, this approach can still provide useful information for the decision-maker.

- Varying the intensity of screening

The critique seems to suggest that to adequately estimate cost-effectiveness, the study should consider all reasonable alternative configurations of the prostate screening programme (meaning different start ages, different stop ages, different frequencies, etc). Indeed, in a complete analysis, all relevant comparators will be assessed relative to the next best alternative. However, as per #2 this may not always be feasible due to limitations in relation to data availability. The question posed in the original paper is straightforward: "to assess the balance of benefit and harms, as well as the cost-effectiveness, of the introduction of a polygenic risk-tailored screening programme for prostate cancer." The question is, if you begin screening when someone reaches a certain minimum level of risk (between ages x and y), what is the cost-effectiveness for a given risk cut-off?

From a policy point of view, the manner in which screening has been assessed in the paper provides useful information: is it prostate cancer screening cost-effective if screening eligibility is decided by risk? It is almost certainly possible to identify variants that are possibly more cost-effective but the complex algorithms associated with them may not be feasible to implement (either because of the difficulties in identifying and calling up individuals, because of the complexity of the public health message, or because of the uncertainty in volumes of people presenting for screening). The paper reduced it down to a simpler question. It is also critical to stress that just because complex pathways for screening can be modelled, it does not mean that the answers are valid or useful. Frequently the parameter data used can only support the analysis conducted on the basis of numerous assumptions with unclear impact on the results. That is, the data for clinical efficacy may be based on a specific frequency or intensity of screening, and when applied to alternative strategies it may have questionable validity. This may be dealt with through uncertainty but that is of limited comfort when the focus is on the "on average" differences.

It's worth bearing in mind the conclusions stated in the original paper are suitably circumspect: "Based on the results of this modelling study, offering screening to men at higher risk could potentially reduce over-diagnosis and improve the benefit-harm trade-off and the cost-effectiveness of a prostate cancer screening program. The optimal threshold will depend on societal judgements of the appropriate balance of benefits-harms and cost-effectiveness." The phrasing is very cautious and it is clear that the authors are aware of the limitations of their approach. It is not presented as an absolute truth, but merely as pointing the way to the potential for a risk-based approach. It would be helpful if the author could acknowledge this, as otherwise it comes across as an unfairly negative commentary.

Overall assessment

Overall, I can't fault the theoretical logic behind the critique which is well-founded in the economic

literature. However, the application in practice is not as straightforward. While the criticisms presented are valid, they are laboured in what would normally warrant only a letter in the journal that published the original article. Decision modellers are faced with the need for pragmatic decision-making according to data availability, technical skillset and decision-maker priorities. When faced with such decisions it is important to consider the relevance of the principle of parsimony (i.e. models are simplifications of reality and should be viewed as such), which is at the core of health economic evaluation. This should be addressed in the paper: modelling does not present the 'truth', it presents a simplification that may or may not approximate the truth. Modelling all possible strategies does not necessarily get you closer to the truth. Indeed, done poorly (even with the best intentions) it may actually get you further from the truth.

The tone of the critique is quite negative considering that the authors of the original paper appear to have appropriately caveated their findings within their limitations and conclusions. The tone could be amended somewhat by providing an examination of the question in the opening paragraph and acknowledging the efforts of the researchers of the original analysis (as per the latter part of third paragraph).

As the paper represents a critique of another paper, I think a letter to the journal would be appropriate than a separate publication. The critique would be a lot stronger if it drew on systematic evidence within the economic literature of these issues rather than focusing solely on one paper. This is even more relevant given that the critiques are intertwined to main overall critique: a fully incremental analysis was not performed.

Finally, the author makes reference to the peer review of the paper, highlighting that one of the issues raised in this paper was also raised in the peer review process of the original paper. It is unfair to draw on specific comments from the peer-review process without a balanced representation of the author's rebuttal to the comments.

**Is the rationale for commenting on the previous publication clearly described?**

Yes

**Are any opinions stated well-argued, clear and cogent?**

Yes

**Are arguments sufficiently supported by evidence from the published literature or by new data and results?**

Yes

**Is the conclusion balanced and justified on the basis of the presented arguments?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Health services research and health economics.

**We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have**

**significant reservations, as outlined above.**

Author Response 04 Sep 2020

**James O'Mahony**, University of Dublin, Dublin, Ireland

### **Preface to Replies**

I thank the reviewers for their thoughtful and thorough comments on the previous version of the manuscript. I welcome the opportunity to revise my manuscript and address their points in response.

### **Re Comments Regarding Mutually Exclusive Strategies**

Regarding the point that simulation of mutually exclusive strategies may be justified by data constraints, computational burdens or the decision-makers decision problem. We acknowledge here that these may all be valid reasons for specifying an analysis in a certain way. I have included a paragraph at the end of the discussion addressing the tensions between pragmatic modelling with constraints as compared to idealised modelling of all strategies for all subgroups. I have kept this discussion brief and relatively general in order to avoid distracting from the principal focus of the critique.

### **Re Comments Regarding Varying the Intensity of Screening**

The reviewers make a series of very fair points here which I agree with. I have revised my manuscript further to reflect the caveats given by the authors regarding the potential relevance of screening strategies of alternative intensity; to emphasise the relevance of data constraints on simulating multiple strategies; and, to further acknowledge the usefulness of results from a model that might simulate less than the theoretical optimal number of strategies.

### **Re "Overall Assessment"**

I fully agree that such a reply would typically be suited to a letter to a journal. Indeed, I wrote to the journal in which the manuscript was initially published (PLOS Medicine) seeking to submit a letter or commentary on the article, but they refused to accept a such submission, instead suggesting that a comment could be added to the article on the journal website. That option seemed inadequate to me as it did not offer the opportunity for a critique to be a fully indexed publication that could be found using research search engines. Furthermore, the proposed option would not offer the opportunity for graphical exposition of the critique. Accordingly, I chose to submit my critique elsewhere.

In part, my response is motivated by my view that a journal publishing work should, as a matter of principle, publish replies to that work. I believe this obligation applies irrespective of whether the journal is an open access journal that charges for publication, as does PLOS



Medicine. I consider that obligation important, especially where published work fails to adequately address points raised during peer review, as I believe is the case here.

I understand why the exposition of my critique appears laboured. This is because the simple question of appropriate incremental comparisons is, unfortunately, intertwined with the knotty issue of mutually exclusive subgroups and multiple screening intensities. Had these issues been separate, it would have been possible to offer an incremental reinterpretation of the published results that resolved the problems of a non-incremental analysis. It seemed appropriate that any critique addressing the issue of incremental comparisons should also address the issue of mutually exclusive strategies within the manuscript. To address this latter problem clearly required a more detailed submission.

I fully agree with the observations made by the reviewers regarding pragmatic use of CEA and the limitations that data present. I have not given these concerns full consideration in my initial manuscript. As mentioned above I have now added a paragraph to the discussion to acknowledge these issues regarding the consideration of multiple screening alternatives.

### **Re Negative Tone**

I understand that the tone of the critique will have read as negative. I have adjusted both the opening of the manuscript and discussion to better relate the constructive intention of the contribution, which was to explain what was wrong with the published results and to explain how future modelling could to address these issues. In accordance with the reviewers' recommendation, to further highlight the useful contribution of the authors I have also amended the discussion to note how Callender et al.'s model could usefully inform trial design in order to better guide future policy.

As a footnote on tone, while my submission may be construed as negative, I think it is important that researchers do not shy away from being clear and unequivocal where possible. A direct articulation of what could be changed appears the most positive and constructive way to contribute to the literature as a whole, even if it requires a somewhat uncomfortable examination of a specific study.

### **Re Placing Critique in Context of Broader Review of Evidence**

I appreciate that the critique would usefully be placed within a broader review. Indeed, I am currently drafting a manuscript assessing the issue of risk group stratification within screening. However, I felt it was necessary to offer a focused critique of this paper for two reasons. The first is that erroneous findings can endure in the literature if not clearly and promptly challenged on publication. This matters, as such results could be used to inform screening policy, leading to suboptimal allocation of resources. The second is that the issues regarding appropriate strategy specification and comparison are not trivial and wanted to present my critique in sufficient detail for readers to reasonably understand the criticisms made. I do not believe I could offer the level of detail required to fairly assess the published study if it was placed within an analysis of many other studies.

### Re Giving a Fair Reflection of the Authors' Reply to Peer Review

This is an important point and I would like to address it clearly. I felt the authors' replies further conveyed a lack of awareness of the relevance of incremental analysis to QALY maximisation and reflected a failure to understand the guidance offered by the reviewer. In reply, they contradict the reviewer's advice that incremental analysis is the standard approach in economic analysis by asserting comparison to no screening is the current paradigm. When the reviewer states that the policies based on their ACER recommendations would not be cost-effective, the authors assert that they are and would maximise QALYs. Both points are clearly erroneous. Given these replies I do not feel a sense of balance would be enhanced by dwelling on them. Accordingly, I have modified my manuscript simply to note that the authors gave their rationale for not amending their analysis. I feel this is suitably neutral and allows other readers to read the reply to reviewers and come to their own judgement.

**Competing Interests:** No competing interests were disclosed.

---

## Comments on this article

### Version 1

Reader Comment 18 Jun 2020

**Tom Callender**, University College London, UK

### Interpreting cost-effectiveness ratios in a cost-effectiveness analysis of risk-tailored prostate screening: A critique of Callender et al. - Reply

#### Authors:

Tom Callender<sup>1</sup>, Stephen Morris<sup>2</sup>, Paul Pharoah<sup>2,3</sup>, Nora Pashayan<sup>1</sup>

#### Affiliations:

<sup>1</sup> Department of Applied Health Research, University College London, London, UK.

<sup>2</sup> Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

<sup>3</sup> Department of Oncology, University of Cambridge, Cambridge, UK.

O'Mahony, in his commentary<sup>1</sup>, makes the criticism that we have incorrectly interpreted the cost-effectiveness estimates in Callender et al.<sup>2</sup> He sets out three issues.

In his interpretation of the cost-effectiveness analysis, O'Mahony reports that our paper finds that screening at a 10-year absolute risk (AR) of developing prostate cancer of 5% or more would be cost-effective, in contrast to his efficiency frontier analysis showing 10-year AR of 9.5%-10% being

*the most* cost-effective. A screening strategy being cost-effective compared to no screening at a willingness-to-pay of £20,000 per quality-adjusted life-year is not the same as being the most cost-effective strategy, and we do not suggest this. In our analysis, to compare all strategies studied against each other, we used net monetary benefit, a widely accepted method in cost-effectiveness analysis of multiple alternatives<sup>3</sup>, and come to the same conclusion as he presents in his reinterpretation, a point which O'Mahony notes.

O'Mahony reports that our analysis is of non-mutually exclusive strategies that “correspond to a mixture of different policy choices for men of different lifetime risk”. This is an incorrect interpretation of our risk-stratified screening strategy. We examined the strategy of eligibility for screening being based on 10-year AR that is dependent on age and polygenic risk and equivalent to the risk threshold for eligibility based on age alone<sup>4,5</sup>. For example, the US Preventive Services Taskforce consider 55 to be the age at which, for some men, there is a net benefit to screening<sup>6</sup>. In England, a 55-year old man has a 10-year AR of being diagnosed with prostate cancer of ~2.5%<sup>2</sup>. Instead of inviting all men from age 55, we invite men to begin screening when their 10-year AR reaches 2.5%, dependent on both their age and polygenic profile. There is a mix-up in O'Mahony's use of the terms higher and lower risk, 10-year absolute risk, and lifetime risk, with no differentiation between lifetime and remaining lifetime risk by age.

There are different approaches for risk-stratified screening<sup>7</sup>. We consider one approach, and O'Mahony appears to propose another, using risk independent of age. That O'Mahony proposes a different risk-stratified screening strategy to the one evaluated in our paper does not negate our analysis.

In modelling different risk-stratified screening intensities, we have acknowledged in our paper the importance of varying the inter-screening interval by risk and discussed in detail why we have not done so.

We show how O'Mahony's critique appears to be based on misunderstanding or misinterpretation of our paper. We think that a more constructive approach would have been for O'Mahony to model the alternative screening strategies he outlines to quantify any difference and provide empirical backing for such screening programmes, furthering the field.

## References

- 1 O'Mahony JF. Interpreting cost-effectiveness ratios in a cost-effectiveness analysis of risk-tailored prostate screening: A critique of Callender et al. *HRB Open Res* 2020; **3**: 23.
- 2 Callender T, Emberton M, Morris S, et al. Polygenic risk-tailored screening for prostate cancer: A benefit-harm and cost-effectiveness modelling study. *PLOS Med* 2019; **16**: e1002998.
- 3 Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes: Fourth Edition*. Oxford, U.K.: Oxford University Press.
- 4 Pharoah PDP, Antoniou AC, Easton DF, Ponder BAJ. Polygenes, Risk Prediction, and Targeted Prevention of Breast Cancer. *N Engl J Med* 2008; **358**: 2796–803.
- 5 Pashayan N, Duffy SW, Chowdhury S, et al. Polygenic susceptibility to prostate and breast cancer: Implications for personalised screening. *Br J Cancer* 2011; **104**: 1656–63.
- 6 Grossman DC, Curry SJ, Owens DK, et al. Screening for Prostate Cancer. *JAMA* 2018; **319**: 1901.

7 Chowdhury S, Dent T, Pashayan N, *et al.* Incorporating genomics into breast and prostate cancer screening: Assessing the implications. *Genet Med* 2013; **15**: 423–32.

**Competing Interests:** None

---