Author’s response to reviews

Title: Development of breast cancer mortality, considering the implementation of mammography screening programs - a comparison of western European countries

Authors:

Yukio Iwamoto (yukio.iwamoto@uni-heidelberg.de)
Simone Kaucher (simone.kaucher@uni-heidelberg.de)
Eva Lorenz (eva.lorenz@bnitm.de)
Till Bärnighausen (till.baernighausen@uni-heidelberg.de)
Volker Winkler (v.winkler@uni-heidelberg.de)

Version: 1 Date: 14 Mar 2019

Author’s response to reviews:

We also uploaded our response as a separate file,

Notes:

- All changes related to the comments were highlighted in yellow in the revised manuscript.
- We also added page and line numbers in the comments to avoid any misunderstanding.
- Additionally, we kept the track-changes for the important parts to be able to reconstruct the changes.
- Minor changes in wording etc. are not track-changed with the purpose to facilitate visualization and readability of the revised manuscript.
- A summary of additional improvements made to the manuscript can be found at the bottom of this letter.

Reviewer 1: Jacques Fracheboud, M.D.

1. Similar analyses have been done before for single countries, among others by Otten et al. for the Netherlands (Int J Cancer 2008; surprisingly not referred to), who came to the same conclusion: a negative (decline) trend change in mortality related to the implementation of the mammography screening programme.
We had cited Sankatsing et al. (Int J Cancer 2017) on p.8 line 42 instead, which is a very similar analysis to Otten et al. (Int J Cancer 2008) but with more recent data and the same conclusion of a declining trend in the Netherlands. However, we have now included the paper from Otten et al. (2008) on p. 8 line 42, as well.

2. Introduction, P3, one but last sentence: "in small clinics" can be questioned; opportunistic screening is also provided by large and even university clinics. But I wonder if the authors do not mean "small (reading) volume", as this is an important difference between opportunistic and organized screening in countries with a strongly centralized screening organisation.

Our intention here was to point out that there is a wide range and discrepancy in quality standards within the clinics conducting opportunistic screening (which include small clinics as well as university clinics). We clarified the sentence on p. 3 in line 29 onwards.

3. Table 1: A 1-year implementation period is not very realistic. Geographical coverage, e.g. everywhere in the country, screening has started, and target population coverage (all targeted women has been invited at least once) might be mixed up. Age ranges changed after (Netherlands from 50-69 to 50-74; UK, the latter starting since several years before the age of 50) or during the implementation (France started also with 50-69).

Thank you for pointing that out, the term “implementation period” was not defined specifically, and we added a note in the legend of table 1 for this purpose. For large countries, a one-year implementation period is indeed not realistic. In table 1, we have three countries for which we have found a one-year period of implementation in the literature and all three are relatively small (population wise). We added the information regarding age range changes for the Netherlands, the UK and France in the legend.

4. Methods, P4/5: The first analysis is done country-specific, whereas the second and third take all countries into account. This should somewhere be mentioned. I am not sure that I fully understand “interrupted time series”, but this is possibly due to my limited knowledge of the Joinpoint application and my statistical skills.
This is an important point which was not clear enough. We added a phrase on p.5 line 18 onwards to clarify that the interrupted time series (ITS) and consequent multivariable Poisson regression were conducted on the entire dataset of all 14 countries together. We also included this information in the respective paragraphs of the results-section. Additionally, we’ve moved the Reference 23 (which explains ITS) to the beginning of the paragraph and added a short sentence on its purpose. In general, the ITS analysis is increasingly being used to evaluate the effect of population-based interventions retrospectively. The ITS is appropriate when the intervention (such as in our case the policy to offer an MSP) was implemented at a specific point in time on a population-level. Therefore, the investigated time series is “interrupted” by the implementation of the intervention. A more detailed explanation of the ITS can be found in Bernal, Cummins & Gasparrini (2017).


5. Results, P5/6: the description of Figure 1 looks a bit random. Countries as Italy, Norway and Sweden were not mentioned at all. From all three countries observational studies were available that found a positive impact on breast cancer mortality. In particular Sweden, the pioneer in breast cancer screening that started before opportunistic screening became widespread is interesting as it looks like having no effect. As Sweden has the longest period of "years with MSP" and the impact of this variable seems to be not so small, this should be discussed in the Discussion section.

This is true, we hadn’t mentioned all countries from figure 1 in the text. We included some sentences regarding the mortality trends from the countries which were missing so far on p. 6 line 34 onwards.

Since we were not able to use mortality data on a regional level, we could not define “year of MSP implementation” as accurate as others who have used regional-level data. This may have blurred the effect of MSP implementation in our results, especially for countries with very long implementation periods and regional heterogeneities (like in Sweden, Norway and Italy). Thank you very much for pointing that out. We have now included a new paragraph to the discussion section, where we discussed our results for Sweden, Norway and Italy from this point of view and compared the results with previous findings (see p.8 line 27 onwards).

After updating the analyses with the most recent WHO mortality data from December 2018, we actually see an accelerated decreasing effect among 50 to 59-year old women in Norway, four years after implementation of the MSP. We also added this information to the results section.
6. Discussion, P8, last para: what do you mean with "relatively small dataset"?

Even though we are using datasets from 14 western European countries over the time of 38 years, we are still relying on the country-level aggregated data in our ecologic study design. For clarification, we have explained it more thoroughly in the discussion (p.9 line 39 onwards) and it should be clear now.

7. The authors conclude that "the implementation of MSPs has contributed to the reduction of breast cancer mortality". This is, however, not for all countries in Figure 1 the case; some of them with important, large programs of good quality do hardly show any trend change (Italy) in breast cancer mortality after the implementation of screening or not for both targeted age groups (Sweden). This should be discussed in order to justify the general conclusion.

It is true that we did not observe reducing breast cancer mortality rates associated with MSP implementation in all countries from figure 1. However, as we already mentioned above (see response no. 5), results from figure 1 for countries like Sweden and Italy with particularly long implementation periods and regional heterogeneities, need to be treated with caution since we were not working with regional-level data. Therefore, the “year of MSP implementation” does not capture these developments accurately enough. We have added this point to the discussion section (p.8 line 27 onwards).

For the regression analyses we combined data from all 14 countries to be able to draw general conclusions about the MSP implementation and mortality reduction in Europe. The results of these regression analyses bring us to the conclusion of stating that MSP implementations in Europe have contributed to the reduction of breast cancer mortality in Europe. We re-phrased our conclusion on p.10 li.26 onwards to clarify this.

Reviewer 2: Elsebeth Lynge

1. This is a valid attempt to 0 however, very crude, and the results are presented in a way where it is difficult to compare with the outcome of other studies. The key point is that screening, if it works, affects birth cohorts offered screening and not age-groups. I
wonder why the authors have focused on age-groups. Proxy cohort data could be form from the WHO data tabulated by 5-years age groups combined with 5-year calendar periods, starting with year of screening implementation. I acknowledge that the authors' Poisson regression includes a variable "years with MSP", but in the age-group 50-59 this makes limited sense, as the program constantly recruits new 50-years old women.

→ Thank you for your valued opinion. We agree, that an analysis of birth cohorts would be another very interesting approach and possibly more valid to evaluate the MSP itself. However, our analysis focuses on the population-wide effects due to the decision to introduce this health policy. Through additional attention in media, the public discussion of such an extensive health policy and consequences taken by individuals through the awareness, MSPs also affect non-screened age groups. We tried to clarify this aspect by making various changes to the wording (see p.4 line 30 onwards).

→ Furthermore, we want to highlight that an analysis with respect to yearly birth cohorts is not possible with the WHO-dataset since year of birth is not given and age is tabulated in 5-year age groups. The suggested method by the reviewer would lead to a likewise crude analysis of estimated 5-birth year periods, which contain misclassified women at the borders to the previous and the next 5-year period.

→ The variable “years with MSP” tries to explain a possible change in the effectiveness of an MSP since it has been established. E.g. a country, which only has had one year with MSP will benefit in a different way from the program compared to a country where it has been established for ten years. This can be due to improvements in process quality, rise of invitation and participation rates and other factors.

2. The outcome measure as "estimated yearly RR" is also difficult to compare with eg results of the RCTs. Furthermore, if we expect an effect of screening, it is probably not in the form of a slope, where breast cancer mortality would eventually disappear. It is more in terms of an effect of a certain size. Results should be presented eg as estimated effect of 10 years of screening.

→ Thank you for your comment. Our aim was not to directly compare the observed changes in mortality to previously published results from RCTs. This is impossible due to the nature of our ecological data and study design. Our analysis does not even allow a direct evaluation of the MSP as we tried to highlight in our manuscript. The main goal is to investigate if observed trends in breast cancer mortality change with MSP implementation which would hint to positive population-wide effects of MSP. The analysis of time trends is classically
done with the more descriptive Joinpoint analysis as well as the interrupted time series. Even though we know the associated limitations, we think this analysis is appropriate and we tried to highlight its limitations in detail.

3. Given that Autier et al, 2011 presented a somewhat similar analysis coming to the opposite conclusion, the authors should address much more in detail why these studies show different results. Length of follow-up may play a role.

→ The article “Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: trend analysis of WHO mortality database” from 2011 by Autier et al. does indeed use the same WHO database and does also look at Joinpoint trends. However, there are several important differences to our study that make a direct comparison inadequate:

- Autier et al. have much shorter follow-up periods. Especially for some countries (e.g. Belgium) where MSPs were introduced in 2001, they look at the effects only until 2004. In our analyses, we’ve investigated a longer follow-up period of at least 10 years for most of the countries (excluding Austria, which started in 2014 with its nationwide screening) until 2017 and therefore, allowed for a longer time to see possible APCs after MSP implementation.

- We’ve restricted our analyses to the 14 western European countries with more than 4.5 million inhabitants each. Autier et al. looked at a small number of very heterogeneous countries and also included very small countries / regions such as Northern Ireland, Republic of Ireland or Flanders.

- For countries like Sweden or Italy, we also didn’t find any significant APCs after the MSP implementations in the countries. However, we think that these results should be interpreted with caution since the implementation of the MSP took a long time in these countries or since the implementation differed regionally within the countries, it’s actually necessary to use regional-level data to be able to draw conclusions for these countries. We’ve therefore now added a new paragraph to the discussion section (p.8 line 27 onwards), where we’ve discussed this issue in more detail. We think the conclusions regarding MSP efficacy drawn by Autier et al. based on population-level data in Sweden and Norway are doubtful and should be treated with caution.

- Most important difference: Our aim was to investigate the effect of the implementation of an MSP in a country on the breast cancer mortality, the aim of Autier et al. was to assess the effect of an MSP on breast cancer mortality (which is questionable whether this is possible based on population-level data in an ecologic study design).
As mentioned by reviewer 3, this article also tries to conclude the general effectiveness of nationwide MSPs from a very high-level population data basis, ignoring results of RCTs. We therefore chose to exclude this reference, as we want to avoid a direct comparison between this article and our study.

4. The use of crude - and not incidence-based - mortality data should be discussed.

→ Thank you for pointing that out. Since we are using data from the WHO mortality database, we were not able to analyze incidence-based mortality rates. However, our study focuses on effects through the implementation of the MSP policy, rather than the isolated effect of the measure itself. This means that attention or popularity in media, the public discussion of the policy and consequences taken by individuals through the awareness were analyzed as an aggregated effect of MSP implementation. Therefore, we think it is appropriate that we have not used incidence-based mortality rates. We have added a sentence to the discussion section on p.10 line 3 onwards.

5. The introduction should be shortened considerably.

→ Thank you, we have shortened the introduction.

Reviewer 3: Robert Smith

1. Introduction: The authors approach the issue of evaluating mammography screening programs as if it has received little attention, which is odd given the number of manuscripts cited from the Journal of Medical Screening 2012 supplement.

→ We do not think that the evaluation of MSPs has received little attention so far, but we do think that there is an ongoing debate about the effectiveness of MSPs, which involves supporters and critics of MSPs. We agree that there were several studies performed (including RCTs), which evaluated the effectiveness of MSPs, but there are also many manuscripts published, which are questioning the effectiveness of MSPs. Additionally,
results from meta-analyses on different RCTs do not show consistent results. Furthermore, there are results from Swedish RCTs, showing a high effect of MSP (Swedish two-county trial) and results from the Canadian RCT, showing a weak association between MSP and breast cancer mortality.

→ Of course, these different results also depend on important heterogeneities regarding study designs and the conducted MSPs in the different countries, but these different results lead to an ongoing debate about the effectiveness of the MSP.

However, thank you for pointing it out that this was not clear. We have therefore changed the wording in the introduction on page 4 (penultimate paragraph).

2. Only one randomized controlled trial (RCT) is cited (the Swedish Two County Trial), whereas there were 9 RCTs conducted, and many more meta-analyses of the trials.

→ We have included the references from the European RCTs (p.3 line 8-9) in the manuscript, thank you.

3. The authors are correct that it is complex, but they do not describe the complexities in enough detail. To mention a few, there is the 1) time of introduction, rate of population inclusion, and rate of uptake; 2) the age group invited and the attendance rate of the mature program; 3) quality of screening and the screening interval; 4) isolating screened and unscreened cohorts; censoring deaths from women who die in the evaluation period, but were diagnosed prior to the introduction of screening, and prior to their invitation to screening; 5) duration of follow-up; 6) trends in incidence, including birth cohort and period effects, 7) etc, etc, etc. Having said that, many countries have made the effort to evaluation their programs, and the introduction, thin on the RCTs, is further vague on the studies that have taken place in Sweden, Norway, the UK, Netherlands, Denmark, Italy, etc. There are numerous citations, but they are covered in the new IARC Handbook (15).

→ Of course, we agree and are aware that many countries have made the effort to evaluate the effect of their MSPs and considered the factors you have mentioned. However, we did not aim to evaluate the effectiveness of the MSPs of the 14 European countries, but we rather focused on the effect through the implementation of the MSP policy. Different countries implemented policies in different ways (e.g. define different age ranges eligible for screening, start the implementation at a different point in time, etc.) and it is not possible to look at these differences when using data on a population-level in form of an
ecologic study design like we did. We have included some references from European RCTs in the introduction section (p.3 line 8-9), but since reviewer 2 asked us to shorten the introduction we have chosen not to go into more detail in the introduction.

Additionally, we have added a paragraph to the discussion section, where we also cited studies from Italy, Norway and Sweden and their positive effect of MSPs on breast cancer mortality and discussed why we might not have found an effect although they found it (see also response to question 5 & 7 from reviewer 1).

4. It is odd that the 2002 IARC handbook is cited, and the NEJM summary (reference 32) is cited, but not the new handbook. Evaluation methods, specifically incidence-based mortality analysis,2, 3 and the incidence rate of fatal breast cancers,4 address these challenges, and specifically, have been conducted on a population basis with attention to both invitations to screening and exposure to screening in the target group; the difficulty achieving this attention to detail complicates trend analyses and in general can lead to an underestimate of the benefit.

→ Thank you for pointing that out, we added the IARC handbook from 2016 as a reference to the manuscript (Ref. 43). Also, regarding underestimation with our ecologic study design, we have adjusted the manuscript (answer to comment 5 by reviewer 1). We have also clarified the limitations of our study design in more detail (answer to comment 4 by reviewer 2).

5. Further, disentangling wild type screening and the influence of therapy are the only challenges described in the use of trend data….there are many more issues to address, and failure to address these issues leads to stunningly flawed articles with incorrect conclusions, such as reference 33.

→ Thank you for the note. We tried to mention these difficulties in the paragraph starting on p.9 line 49, as our ecologic study design does not allow us to account for all these confounding variables.

6. The further irony here is that reference 33 is exactly the sort of article that, had it even its methodology had been competent, and it had concluded that mammography was beneficial, the methodology would have been judged to be insufficient to adopt mammography screening on a population basis. Only favorable results from an RCT would be acceptable. Well, we have those.
And yet, Autier, et al. challenges the results of the RCTs based on a poorly designed ecological study.

→ For the reasons mentioned above in comment 3 of reviewer 2, we want to avoid a comparison of that study with our study, since the two, which might look similar on the surface, take a completely different approach. We therefore chose to remove this reference.

7. Methodology: There have been good trend studies, and the methodological issues are described by Moss, et al. in reference 21. In this analysis, there is no adjustment for deaths that occur in the evaluation period but were diagnosed before screening was initiated, deaths among women in their 50s due to a diagnosis in the 40s (in most countries women would not have been invited) and there is no adjustment for trends in incidence, nor are these factors mentioned as limitations.

→ Thank you for pointing this out. For the changing incidence trends, we cited Ref 35 and 36 on p.9 line 29 and we added a sentence on p.9 line 19 onwards emphasizing the additional effectiveness for the age group screened. However, we want to point out that in the WHO dataset it is not possible to differentiate which or how many deaths have been diagnosed and when. But our analysis aims to look at the effect due to implementing MSPs rather than the effectiveness of the MSP itself. The implementation will also affect women in their 40s e.g. through raising the awareness. Please, see also comment 4 to Reviewer 2.

8. Discussion: The opening sentence suggests that this analysis is answering a vexing question, although tentatively ("..the implementation of MSPs has contributed to the reduction of breast cancer mortality."). It would be more appropriate to concluded that the analysis of mortality trends further supports the extensive literature on the effectiveness of mammography screening programs, and is in contrast with trend analyses that question the effectiveness of MSPs.

→ We adjusted the concluding sentence (p.10 line 26 onwards) and added references as well.
9. With respect to the influence of therapy, there are studies that have compared women exposed to screening and women not exposed to screening and have shown both the positive effects of new therapies over time in reducing deaths in women who did not attend screening, but more to the point, the morality reductions are considerably greater in the women who attended screening. This observation was recently demonstrated more conclusively in reference 4 below.

→ Thank you for mentioning this very helpful recent reference. We included it in our manuscript on p.9 line 19 onwards as well as in the conclusion.

10. Regarding overdiagnosis, it is not relevant in this instance because the study is focused on mortality. Women who are overdiagnosed do not die from breast cancer.

→ By the enumeration of controversial aspects in the discussion part, we tried to not only cover those aspects that would have been important to consider for our specific analysis but rather also mention those, which are the source of public debate for or against MSPs. For effectiveness, defined as mortality reduction as in our case, overdiagnosis does not matter, but yet this issue can cause psychic and emotional harm for the patients and is therefore important and should be mentioned. We added a phrase on p.9 line 49 onwards that these controversial points are not limited to those directly affecting our analysis.

11. Page 9, line 18. The working group for the IARC Handbook 15 was a different group than the one that worked on Handbook 7.

→ Thank you for the note. By ‘same working group’, we did not mean the exact members belonging to the group of specialists but rather the IARC as the initiator of this review. We changed it accordingly in the manuscript (p. 10 line 19 onwards).

12. I would recommend against citing reference 33. It is deeply flawed and does not credibly challenge the evidence for the effectiveness of mammography.
13. In general, this manuscript should not treat the efficacy of mammography screening as a still open question; rather, it should focus on the ability of population registry data to be used to evaluate MSPs that commonly target less than half of the female population and have variable attendance rates.

As for comment 8, we adjusted the concluding sentence (p.10 line 26 onwards) and added references there.

14. References


Thank you for the suggestion of additional references. We added the IARC 2016 handbook as a reference. The 2nd and 4th suggestion we also added in the discussion part and conclusion to further strengthen the message to support our results. We had already included the 3rd reference in the discussion, but also added it to the conclusion.

Additional changes made:
- Re-ran complete analysis based on the most recent WHO dataset from December 2018. For all countries, more recent data was available, ranging from availability between up to 2015 to 2017. We also adjusted the figures and values in tables and text accordingly
- Added E-mail addresses of all authors to title page
- Changed titles to “Background”, “Methods”
- Added list of abbreviations
- Added and adjusted figure titles and legends in the text as well as at the end of the manuscript
- Adjusted access dates for websites in references