

## Comment

## Open Access

Karin Huwiler\*, Beat Thürlimann, Thomas Cerny and Marcel Zwahlen

## Comment on: ‘Screening’ for Breast Cancer: Misguided Research Misinforming Public Policies, by O. S. Miettinen

DOI 10.1515/em-2015-0022

**Abstract:** Our commentary of the article “‘Screening’ for Breast Cancer: Misguided Research Misinforming Public Policies” has two main parts. First we address some of the methodological points raised by Professor Miettinen. Then we review more specific aspects of the Swiss Medical Board statement on mammography screening for early detection of breast cancer.

**Keywords:** cancer screening, evaluation, methodology, mammography screening, Swiss Medical Board

Professor Miettinen presents his views on how one should assess the benefit of cancer screening and he highlights a few difficulties. He considers, for example, well-known difficulties (Hernan, 2010) of the incidence-density ratio (hazard ratio) of the cancer-related mortality as a useful metric for the benefit of cancer screening. Even in randomized studies of screening versus no screening with perfect compliance, this ratio is not constant over the follow-up time after having started screening. This phenomenon is clearly and easily observable in screening trials in which only one screening is conducted in a lifetime – possibly in colon cancer screening with sigmoidoscopy (Atkin et al., 2010) or colonoscopy. If the screening examination needs to be repeated, more complications arise in using the incidence-density ratio as an efficacy metric.

Professor Miettinen proposes a new metric for determining whether to screen for cancer or not. He states that almost all experts missed addressing the most crucial parameter (of Nature): “the risk/probability of death from the cancer as a function (causal) of its treatment (early vs. late) in a defined domain (preclinical) of applying the compared diagnostic-therapeutic regimens”. So what the patient and the doctor want to know is “the reduction in the cancer’s case-fatality rate resulting from its early treatments following its early – latent-stage, preclinical – diagnoses in lieu of treatments of the cancer when it already is clinically manifest.” He argues that most experts and others were inappropriately interpreting the results of the conducted randomized screening trials. It is not clear whether he questions the conduct of these trials.

In order to simplify the arguments, let us review the situation in which a once only screening is compared to no screening. As clearly highlighted by Professor Miettinen, the isolated screening examination per se has no effect, it is the screening as a starting point of a chain of procedures triggered by a result possibly indicating further diagnostic investigation which might lead to a definite diagnosis followed by treatment (which might subsequently be modified as judged necessary by some clinical experts). Therefore, a randomized trial of screening versus no screening is actually assessing how – over a carefully defined time horizon – results in the population without initial screening examination, differ from results in the population which was screened and underwent this cascade of interventions. Any marginal differences observed in cancer-specific mortality (or other outcomes such as cancer diagnoses and all-cause mortality) after, for example 15 years, are the combined result of starting the cascade versus not starting the cascade. Any attempt to disentangle the relative contribution of the separate steps is probably impossible. If recruitment to the randomized screening trial happens over several years, then possibly some elements of the steps in obtaining a definite diagnosis and what treatments are used might change over time and not

\*Corresponding author: Karin Huwiler, Swiss Cancer League, Bern, Switzerland, E-mail: karin.huwiler@krebsliga.ch

Beat Thürlimann, Breast Center, Cantonal Hospital St. Gallen, St. Gallen, Switzerland

Thomas Cerny, Oncosuisse, Bern, Switzerland

Marcel Zwahlen, Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

 © 2015, Huwiler et al.

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 License.

all persons randomized to screening will be “exposed” to the same elements of the cascade. What will be observed after – say – 15 years is a mixed effect averaged over screening plus the varying cascade elements versus cascade elements in standard care in the no screening group. Even if recruitment time would be short and cascade elements the same for those in the screening arm of the trial, the usually long follow-up will lead to discussions about the final results of the trial. If some elements of the cascade have changed by the time trial results are published (different treatments thought to be better) transferability of results to persons who today undergo screening (compared to no screening) is questionable. This transferability dilemma is inherent in cancer screening due to the rather long time horizon necessary to assess the effects of screening (and the subsequent interventions).

With all these complications in evaluating cancer screenings, it is surprising that Professor Miettinen argues that it should be possible to validly estimate the reduction in the cancer’s case-fatality rate resulting from screening and earlier treatment for cases that would have become clinically manifest. This calls for a very special case of comparing two potential outcomes (Little and Rubin, 2000), namely for a person who would certainly have progressed to clinical disease when undergoing screening (with all the subsequent cascade of interventions) to a situation with the same person not having been screened. The main difficulty is that it seems impossible to determine at the time of screening whether the screened person belongs to those who would have progressed to clinical disease in the absence of screening. Furthermore, it illustrates an important aspect related to overdiagnosis. If the person (to be screened) has a substantial risk of dying from another disease, (s)he will have a low probability to have a clinically diagnosed cancer. This form of overdiagnosis is inevitable (Marcus et al. 2015) and a crucial reason to not recommend screening in older persons or persons with a reduced life expectancy (for example of less than 10 years).

To conclude the methodological arguments: As much as it would be interesting to know whether and by how much cancer screening is reducing “the cancer’s case-fatality rate resulting from its early treatments following its early – latent-stage, preclinical – diagnoses in lieu of treatments of the cancer when it already is clinically manifest” it is an elusive metric which by no study design, even if ideally conducted, can be validly estimated.

As Professor Miettinen illustrates referring to the report on mammography screening prepared by the Swiss Medical Board (SMB), there are other issues to be addressed such as the composition of expert groups. Such groups generally should include multidisciplinary knowledge, especially competent experts with an epidemiological and methodological background as well as competent clinical experts, who have a broad view and the ability to not only defend their own discipline, knowledge of medical but also of public health aspects. Special knowledge and experience in evaluating screening interventions is necessary. The criteria for selection of the evaluation committee are somewhat different to those for the evaluation of relatively straightforward medical procedures because long-term interventions with a public health focus like mammography screening with an effect expected only after years or even decades, require special methodological considerations. Although these thoughts on the composition of such groups are not really groundbreaking, a careful selection has not been the case for that report (and the reports published earlier by the SMB). This resulted – not surprisingly – in methodological flaws of the report (de Koning and Heijnsdijk, 2015) and somewhat incoherent conclusions which might even contradict themselves. By omitting a discussion of pros and cons of opportunistic mammographies outside of organized programs, the SMB indirectly supported the opportunistic approach which has no quality assessment at all. Fortunately, the institution recognized weaknesses, and both the composition of the board and the processes are being adapted. Unfortunately, the report has never been revised.

What was the aftermath of this report in Switzerland? Mostly confusion and irritation of physicians, politicians and especially of women. Some regional mammography screening programs in Switzerland may report lower participation rates after publication of the report, and some cantons (comparable to states in the USA) have postponed the introduction of mammography screening programs whereas others initiated their program despite the SMB report. Whether a decline in the participation rate will persist, remains to be seen. An analysis of other reports published by the SMB showed that such declines – if they occurred – did not last for a long period of time (Eichler et al. 2015).

We do agree with exponents of the SMB that irritation per se might not be detrimental, but in all likelihood a necessary step in the evaluation of interventions. But we believe that in this specific case it was not necessary but rather counterproductive – the SMB report is based on the information, available for years now, from the well-known randomized controlled trials, and on a questionable cost-effectiveness-analysis done by the SMB. Interestingly, the UK panel based its conclusion – that mammography screening programs confer significant benefit and should continue – on the same trials and their own experience with decades of mammography screening (Independent UK Panel, 2012). What the SMB report did *not* lead to is the abandonment of mammography screening programs in Switzerland.

Interpretation of study results and evaluation of interventions is not always straightforward. It starts with the question: What kind of evidence should be considered? Only the evidence from randomized controlled trials (if available), or from observational studies as well? Which models should be used and which assumptions should be made? The approach chosen will undoubtedly impact on the results and conclusions, as will the choice of outcome measure, of statistical analysis etc. But even with the same evidence base, different persons or organizations might come to different conclusions. Ultimately, one has to weigh positive effects against negative ones – and this is up to a certain degree a question of personal preferences and may also be subject to change over time. Coming back to mammography screening and the SMB report, the only positive outcome considered was reduction in breast cancer mortality. As every clinician taking care of women with breast cancer knows, dying from breast cancer is not the only thing women would like to prevent. Avoiding more advanced or even late-stage disease by diagnosing breast cancer early is also a positive outcome for many women. And we should keep in mind that many women are confronted with breast cancer and its sequels as many families know someone affected by breast cancer within the family or among their friends.

As discussed above, an RCT in mammography screening evaluating the case-fatality rate, taking into account overdiagnosis, as proposed by Professor Miettinen, in our opinion would be welcome but is not feasible. RCTs evaluating breast cancer mortality rate would hardly be feasible, in particular (low) accrual and selection bias could be a problem nowadays. Nevertheless, results would be available after years only, and even in properly performed RCTs, generalizability of results will probably be low due to advances in diagnostics and treatment occurring in the meantime. The priority for new studies should be on interventions to reduce harms of mammography screening (including diagnostic workup and treatment), i.e. reducing false positive results and especially reducing overdiagnoses and overtreatment. Quality for breast cancer screening can best – some would say only – be assured if the screening is embedded in organized programs, where quality measurement and measures will help further reduce false positives and negatives. Organized mammography screening programs therefore clearly should be favored over opportunistic mammographies. Furthermore, trials should be conducted which evaluate less invasive treatments for potentially indolent disease, especially small size, low grade DCIS and luminal A invasive cancers. There are indications that e.g. radiation and/or systemic adjuvant therapy can safely be omitted in certain cases (e.g. McCormick et al. 2015; Christiansen et al. 2011). What we also should think about is leaving the one-for-all approach and evaluating risk-stratified screening, meaning e.g. no screening or larger screening intervals for women with low risk for breast cancer and shorter intervals for women with higher risk. Both these approaches could help reducing overdiagnosis and overtreatment. We also expect improvements due to technological advances (e.g. tomosynthesis), but we have to make sure that these will not primarily lead to more diagnoses of even more low-risk (pre-) disease. Risk-adapted mammography screening based on age, breast-density, BMI, family history and molecular markers is investigated in the UK (Warwick et al. 2014).

Finally, we need to better understand how to optimally inform the target population about breast cancer screening and cancer screening in general. Several studies show that there is room for improvement in this area (e.g. Domenighetti et al. 2003; Hersch et al. 2015). We need to know more on how the information should be provided and on how women want to be informed about pros and cons of mammography screening. The aim should be to enable women to make an informed decision on whether or not to participate in mammography screening programs.

Finally, we ask researchers in the field to report study results with caution. Interpretation of results is necessary, but the inherent uncertainty in all results should also be discussed in a balanced way. When publishing and discussing study results, researchers should also think about the potential consequences for the lay public – in the sense of *Quidquid agis, prudenter agas et respice finem!* or in English: Whatever you do, do it wisely and consider the end.

## References

- Atkin, W. S., Edwards, R., Kralj-Hans, I. et al. (2010). Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: A multicentre randomised controlled trial. *Lancet*, 375(9726):1624–1633.
- Christiansen, P., Bjerre, K., Ejlersen, B. et al. (2011). Mortality rates among early-stage hormone receptor-positive breast cancer patients: A population-based cohort study in Denmark. *Journal of the National Cancer Institute*, 103(18):1363–1372.
- de Koning, H. J. and Heijnsdijk, E. A. M. (2015). Swiss medical board mammography screening predictions for Switzerland: Importance of time-periods. *Journal of Medical Screening*. Online first, published on May 27, 2015.
- Domenighetti, G., D'Avanzo, B., Egger, M. et al. (2003). Women's perception of the benefits of mammography screening: Population-based survey in four countries. *International Journal of Epidemiology*, 32:816–821.
- Eichler, K., Hess, S., Riguzzi, M. et al. (2015). [Impact evaluation of Swiss Medical Board reports on routine care in Switzerland: A case study of PSA screening and treatment for rupture of anterior cruciate ligament](#). *Swiss Medical Weekly*, 145:w14140.
- Hernan, M. A. (2010). [The hazards of hazard ratios](#). *Epidemiology*, 21(1):13–15.
- Hersch, J., Barratt, A., Jansen, J. et al. (2015). Use of a decision aid including information on overdetected to support informed choice about breast cancer screening: A randomised controlled trial. *Lancet*, 385:1642–1652.
- Little, R. J. and Rubin, D. B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. *Annu Rev Public Health*, 21:121–145.
- Marcus, P. M., Prorok, P. C., Miller, A. B. et al. (2015). Conceptualizing overdiagnosis in cancer screening. *Journal of the National Cancer Institute*, 107:4.
- McCormick, B., Winter, K., Hudis, C. et al. (2015). [RTOG 9804: A prospective randomized trial for good-risk ductal carcinoma in situ comparing radiotherapy with observation](#). *Journal of Clinical Oncology*, 33:709–715.
- Warwick, J., Birke, H., Stone, J. et al. (2014). Mammographic breast density refines Tyrer-Cuzick estimates of breast cancer risk in high-risk women: Findings from the placebo arm of the International Breast Cancer Intervention Study I. *Breast Cancer Research*, 16:451–455.
- Independent UK Panel on Breast Cancer Screening (2012). The benefits and harms of breast cancer screening: An independent review. *Lancet*, 380(9855):1778–1786.