OXFORD

JNCI J Natl Cancer Inst (2020) 112(10): djaa025

doi: 10.1093/jnci/djaa025
First published online February 14, 2020
Commentary

The Value of All-Cause Mortality as a Metric for Assessing Breast Cancer Screening

Martin J. Yaffe, PhD,[1,2,*] James G. Mainprize[1]

[1]Physical Sciences Platform, Sunnybrook Research Institute, Toronto, Canada; and [2]Department of Medical Biophysics, University of Toronto, Toronto, Canada

*Correspondence to: Martin J. Yaffe, PhD, Sunnybrook Health Sciences Centre, Room S657, 2075 Bayview Avenue, Toronto, ON M4N3M5, Canada (e-mail: martin.yaffe@sri.utoronto.ca).

COMMENTARY

Abstract

Although screening mammography has been demonstrated to contribute to reducing mortality due to breast cancer, some have suggested that reduced all-cause mortality should constitute the burden of proof for effectiveness. Using a microsimulation model of the development, detection, and treatment of breast cancer, it is straightforward to demonstrate that this is an unrealistic expectation for trials of practical size and period of observation, even where the reduction of breast cancer mortality is substantial. Estimates of all-cause mortality will depend not only on the efficacy of the screening intervention but also on the alignment between the age distribution of the effect of screening on reduction of deaths and that of the other major causes of death. The size of a randomized trial required to demonstrate a reduction in all-cause mortality will, therefore, depend on the length and timing of the observation period and will typically be at least 10 times larger than the size of a trial powered to test for a reduction in deaths due to breast cancer. For breast cancer, which represents a small fraction of overall deaths, all-cause mortality is neither a practical nor informative metric for assessing the effectiveness of screening.

In assessing the effectiveness of breast cancer screening, some have looked for an effect on all-cause mortality (ACM), that is, death from any cause over a specific time period (1–6). The rationale for this expectation is that if the screening intervention is responsible for a reduction in breast cancer mortality, overall mortality should also be seen to drop, and unless this occurs it implies that the intervention results in an increase in deaths from other causes, for example, cardiotoxicity due to chemotherapy or radiation-induced cancer.

Overall reduction in ACM has not been observed in the randomized trials studying the efficacy of breast cancer screening, and some have interpreted this as an indicator that breast cancer screening is not effective (6–8). We contend that unless a trial is specifically powered to test changes in ACM, it is quite probable that a reduction in ACM will not be observed even though a substantial reduction in deaths due to breast cancer is measured. Furthermore, this is a result of statistical uncertainty or timing of the observation period rather than an increase in other-cause mortality (OCM) caused by the screening intervention. Here, we demonstrate this using estimates from a microsimulation model of breast cancer growth, detection, and treatment.

For this purpose, we employed the OncoSim-Breast model recently developed by Statistics Canada (9). OncoSim borrows concepts from the Wisconsin CISNET Breast Cancer Model (10,11) and considers the effects of different cancer subtypes, and population immigration and emigration. OncoSim "follows" each woman in a large cohort through life, and each year calculates—by "pulling" a random number for each process—the probability of development of breast cancer, detection, and survival after treatment, updating the tally of women who are alive or die from breast cancer or other causes. Competing risks of death are built into these probability calculations.

The estimated reduction of both breast cancer mortality and ACM (the effect sizes) in a randomized controlled trial (RCT) will depend on the intrinsic efficacy of the intervention and compliance with the trial protocol and the randomization assignment. Importantly, it will also be influenced by the period of observation and how it aligns with the period of the screening intervention.

To illustrate the importance of trial size and choice of observation period, let us consider four scenarios applied to the following cohorts of women: those who received no breast cancer screening (the reference cohort), an idealized extreme where breast cancer deaths were completely eliminated, women who were screened annually between ages 40 and 74 years, and women who received annual screening but only between ages 40 and 49 years. For the screened scenarios, let us assume that

Received: September 26, 2019; **Revised:** January 9, 2020; **Accepted:** February 5, 2020

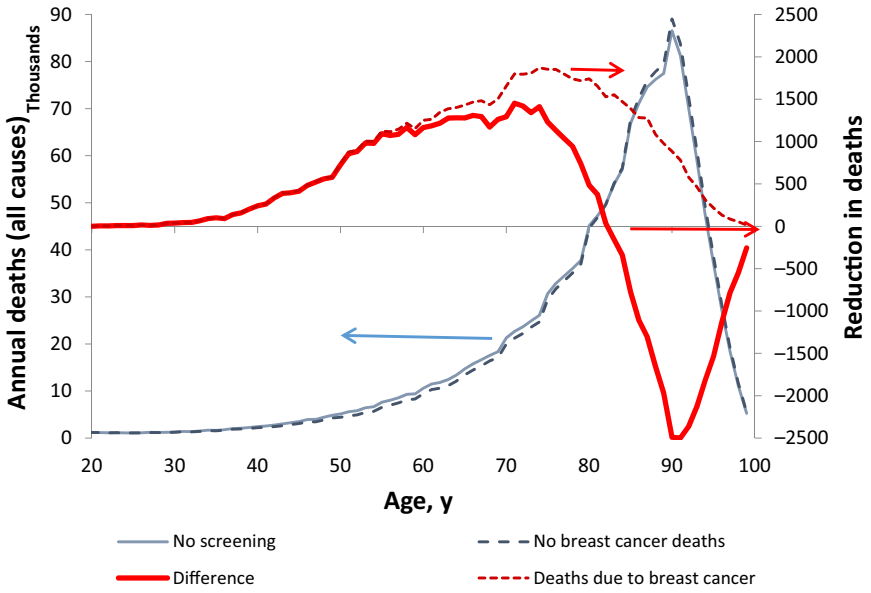

Figure 1. All-cause deaths and breast cancer deaths vs age. **Left axis**: Annual deaths at each age from all causes for a cohort of 2.1 million women who (**solid gray line**) did not receive breast cancer screening (reference cohort) and (**dashed gray line**) one for which there are no deaths due to breast cancer. The curves appear to be nearly superimposed. **Right axis, solid black line**: The difference between the numbers of deaths in these cohorts, expanded for visibility. Positive values indicate fewer all-cause deaths in the cohort without breast cancer deaths. Also plotted (**dashed black line**) are the annual number of deaths from breast cancer in the reference cohort.

modern mammography was used and that women complied perfectly with the annual screening interval.

For each woman in a cohort, the model predicts whether she developed breast cancer at some point or not, the age at death, and whether death was attributable to breast cancer or some other cause. Results from OncoSim-Breast model runs are provided in Supplementary Table 1 (available online).

For the extreme example of comparing an unscreened population with one (also unscreened), in which all breast cancer deaths had hypothetically been eliminated (Cohort 2 vs Cohort 1), the model estimates of deaths at each age from all causes for these scenarios are shown in Figure 1 for 2.1 million women in each cohort.

Although breast cancer is the second highest cause of cancer death in North American women, over a lifetime, it is responsible for only about 2.6% of total deaths (12). For this reason, the curves for the two scenarios are almost indistinguishable, but a close look at results from these very large cohorts reveals that ACM is slightly lower in the cohort without breast cancer deaths for younger women, with the situation reversing after age 82 years. For better visibility, the difference is shown against the right-hand axis. This curve is positive, indicating a reduction in ACM, from around age 30 years when breast cancer deaths begin to occur. The deaths due to breast cancer in the reference cohort are also plotted, and it is seen that up to about age 60 years the reduction of ACM is well explained by the absence of breast cancer deaths in Cohort 2. But although the number of breast cancer deaths continues to increase until about age 75 years, the reduction in ACM, although still positive, begins to decline until age 82 years, after which ACM begins to increase. The increase continues beyond age 99 years, where the model calculation ends. It reflects an increase in OCM that gradually offsets and then overwhelms the effect of the absence of breast cancer deaths. This is an inevitable result of the fact that, in the absence of breast cancer deaths, there will be more women alive to succumb to other causes of death. We all die of something, usually before age 100 years. The application in the model of standard life tables (from which breast cancer deaths have in this case been removed), not the sequalae of treatment for breast cancer, drives this increase in OCM.

This example illustrates that even with the complete elimination of breast cancer deaths, because these mortality curves cross, the ability to observe a reduction in ACM will depend on the timing of the observation period over which deaths are counted. The effect size depends on the relationship between the age-dependent probabilities of the aversion of breast cancer deaths and the occurrence of the other major causes of death. Figure 1 clearly illustrates the large difference in the shape of these distributions, the peak for ACM occurring 15 years after that for deaths due to breast cancer.

The model can be used to emulate an RCT comparing cumulative ACM and breast cancer mortality between two cohorts of equal size for a specified observation period. The effect size for both breast cancer and all-cause deaths can be quantified as the differences between each scenario and the reference in the proportion of deaths that occurred during the observation period. The statistical significance of the effect is characterized by the *P* value for a difference of proportions test (13), with cohort size adjusted until *P* becomes just less than .05. In Table 1, for the comparison of the scenarios, the required sizes (total of both arms) of randomized trials to compare the effects of the three interventions against the unscreened reference group are shown for various observation periods, which here were chosen to begin at age 40 or 50 years with durations between 10 and 60 years.

Table 1 shows that the effect size for ACM depends heavily on duration and situation of the observation period. For an observation period of 20 years (ages 40–59 years), the predicted (relative) reduction in ACM in the extreme example of eliminating all breast cancer deaths would be 13.2%, and this would require an RCT with 10 400 participants (5200 women in each

Table 1. Estimates from OncoSim-Breast model of the relative reductions in deaths due to breast cancer and from all causes and the total number of participants required in a randomized trial to observe these effects. Reference cohort consists of women who receive no screening intervention

Scenario	Duration of observation period (beginning age 40 y), y*	Predicted reduction in breast cancer deaths, %	Required breast cancer death trial size (both arms)	Predicted reduction in ACM, %	Required ACM trial size (both arms)
No breast cancer deaths	10	100	2580	12.1	39 800
	15	100	1250	13.4	17 050
	20	100	750	13.2	10 400
	30	100	380	10.7	6450
	40	100	234	7.2	5600
	50	100	178	2.8	10 050
	60	100	168	1.1	23 800
	10*	100	1050	13.9	13 750
Annual screening 40–74 y	10	18.2	141 700	2.2	1 254 000
	15	27.8	27 560	3.7	230 000
	20	33.7	10 950	4.4	95 050
	30	42.7	3270	4.6	35 600
	40	48.0	1530	3.5	24 000
	50	46.5	1260	1.3	47 200
	60	45.5	1250	0.5	122 100
	10*	40.1	10 340	5.6	89 450
Annual screening 40–49 y	10	18.2	141 700	2.2	1 254 000
	15	25.9	32 300	3.5	268 000
	20	25.5	20 100	3.3	170 000
	30	21.8	14 200	2.3	140 800
	40	17.9	13 050	1.3	186 000
	50	15.9	12 750	0.4	450 000
	60	15.6	12 600	0.2	1 180 000
	10*	28.5	21 950	3.9	180 250

*For the final entry in each section, observation begins at age 50. ACM = all-cause mortality.

arm). Note that effect on ACM in this example is a reduction from 5.3% to 4.6%, in absolute terms a reduction of 0.7%.

The effect of observation times is interesting in that it involves a tradeoff between the difference in the proportion of deaths and the actual number of deaths in the scenarios being compared. With increased observation time, there will be more deaths, reducing the statistical uncertainty (lower *P* value) in comparing the proportion of deaths. However, the increasing OCM with age reduces the relative difference in proportion of deaths between the two cohorts. As shown in the upper right of the table, trials with less than 20 years of observation would have to be larger to achieve statistical significance, as would a trial with 60 years of observation; in the former case because there would be an insufficient number of deaths counted and in the latter because increasing OCM would reduce the effect size. In this example, the "sweet spot" in the tradeoff occurs at a lengthy 40 years of observation (ages 40–79 years), where a trial with a total of only 5600 patients would be powered to detect the 7.2% decrease in ACM.

Most reports of RCTs are limited in observation time, at least at their first publication. To achieve statistical significance for a shorter observation period of 10 years, a trial would require many more patients, almost 40 000, to observe the 12% reduction in ACM if it began at age 40 years. But if observation began at age 50 years, a total of 13 750 patients would be adequate to detect the only slightly larger reduction of 14%. Recall that this is for the complete elimination of breast cancer deaths.

Now a more realistic example—the comparison of the unscreened cohort with one that receives annual mammography between ages 40 and 74 years. Results from OncoSim-Breast are shown in Figure 2, again for cohorts of 2.1 million women. Here, the annual difference in ACM, that is, lives saved, as well as the annual breast cancer deaths averted as a result of screening are plotted on the left-hand axis and the cumulative reductions in ACM and breast cancer deaths are shown at the right. Again, the reduction in ACM in midlife is almost completely explained by the reduction in breast cancer deaths. The predicted reduction in breast cancer mortality has a broad age distribution, beginning shortly after age 40 years when screening begins, centered at age 73 years just before screening ends, and continuing beyond age 90 years. On the other hand, the increase in OCM with age gradually diminishes the reduction in ACM. Although this results in an increase in annual ACM after age 82 years, the cumulative reduction persists past age 99 years.

Figure 2 illustrates again the importance of the observation period on the assessment of effects on ACM. Both the time at which observation starts and the duration of the observation period will strongly influence the estimate of the effect on ACM, but here the timing of the screening intervention is also a major factor. The observed effect will be greatest when the window of observation is aligned with the period during which the effect on reduction of breast cancer deaths is greatest, but also before mortality unrelated to breast cancer begins to rise sharply and become dominant.

From the right-hand columns of Table 1, it is seen that the minimum required trial size of 24 000 occurs at 40 years of observation (ages 40–79 years), where a 3.5% drop in ACM is predicted. Here, the value of ACM falls from 26.9% to 25.9%, only 1% in absolute terms. This is logical because it allows much of the

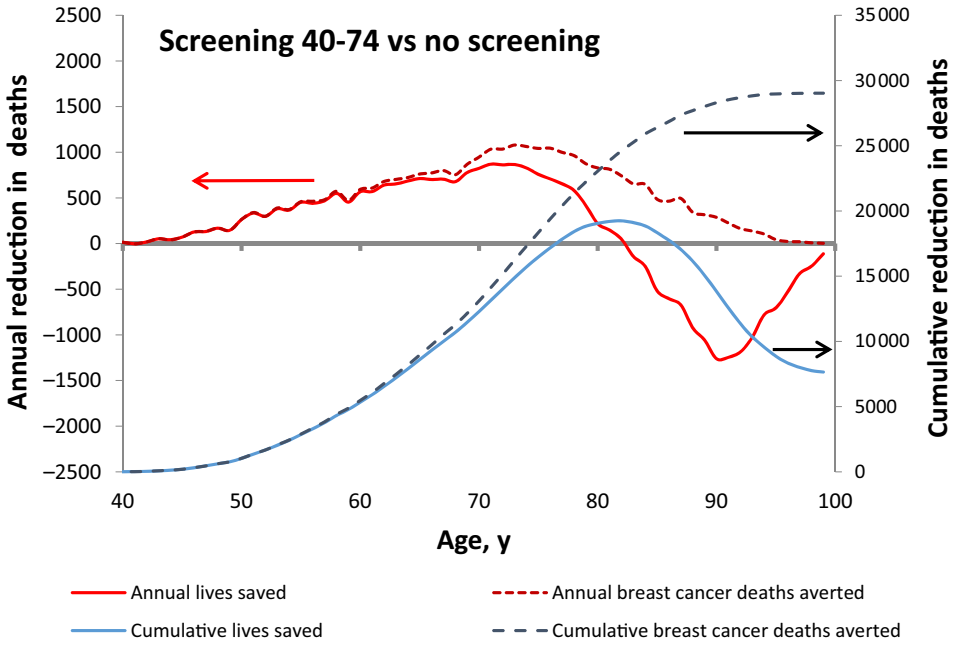

Figure 2. Modeled annual and cumulative effects on breast cancer deaths and all-cause mortality (ACM) because of mammography screening. **Left axis**: Estimated reduction in breast cancer deaths as a result of annual mammography screening between ages 40 and 74 years and annual reduction ACM. **Right axis**: Estimated cumulative effect of this intervention on reduction of breast cancer deaths and deaths from all causes.

benefit from all 35 annual screens to be captured while not overly emphasizing the OCM, which peaks around age 90 years. Although an observation period of 15 years provides a similar relative reduction in ACM, to accrue sufficient deaths for statistical significance an RCT that is almost 10 times larger would be required, in part because the benefit of only 15 of the 35 screens would be measured.

For the 10-year screening period of 40–49 years, much larger trials are required because the probabilities of death due to breast cancer or OCM are low. However, breast cancer is the dominant cause of death in this age group. Because the time period between diagnosis and death can be several years, it is actually considerably more efficient to follow the women screened in their 40s from age 50 to 59 years to observe the 3.9% reduction in ACM with an RCT of 180 000 women.

How does the size of RCTs required to detect the reduction in breast cancer deaths compare with the size needed when reduction of ACM is the endpoint? These data are given in the left-hand columns of Table 1. It is interesting that even for the complete elimination of breast cancer deaths, with 20 years of observation a trial of 750 women would be needed to detect that 100% reduction (Table 1), because fewer than three breast cancer deaths would have occurred in the reference arm. To detect reduction of ACM in the three comparisons illustrated, Table 1 shows that trials must be larger by factors of between 8 and 142 than the size required to observe reduction in breast cancer-specific mortality. Therefore, if a trial is not initially powered to detect a change in ACM, it is highly unlikely that such a change will be seen.

The observation period that allows the smallest trial to detect a change in ACM as the outcome is often different from that for breast cancer-specific mortality. This occurs because the optimal observation period for the former type of trial depends on the alignment of the temporal distributions of both the aversion of breast cancer deaths and that of other-cause deaths, whereas that for the latter depends only on the distribution for aversion of breast cancer deaths. Too short an observation period may miss part of the effect of the intervention, whereas an excessively long period may dilute its effect due to the influence of OCM.

Although breast cancer is the second largest cancer killer of women, over 97% of women die from other causes. Several authors responding to Black et al. (1) have suggested that this limits the value of ACM as a metric for assessing the effect of breast cancer screening by necessitating RCTs that would be impractically large to conduct (14–21). Even with the extreme situation of comparing an unscreened cohort of women with one in which there has been a complete eradication of deaths due to breast cancer and the idealized elimination of fluctuations due to population migration and other factors, the effect on ACM is difficult to observe. Given the high cost of RCTs and the desire to obtain results within reasonably short observation times, it is unlikely that a trial on breast cancer screening will ever be powered to detect reduction of ACM.

Conclusion

ACM is neither a sensitive nor a particularly informative measure for assessing the effect of an intervention for a disease that is responsible for a small proportion of total deaths. Much larger trials are required to test for reductions in ACM than for reductions in breast cancer deaths, and the optimal durations and starting ages for observation may differ. The required trial size for ACM becomes even larger if potentially confounding factors such as crossover, variability in socioeconomic or lifestyle factors, or the type, quality, accessibility, or compliance with the intervention are in place.

When the cohort of interest consists only of those who have been diagnosed with breast cancer, 20–30% of whom eventually die of their disease, ACM becomes a more logical metric as much smaller cohorts are required. In a Swedish study with

120 000 women randomly assigned and an average of 11.6 years of observation, Tabar et al. observed a 13% reduction in ACM in women who participated in mammography screening and were diagnosed with breast cancer, for whom breast cancer mortality was reduced by 31% (20).

Most interventions are designed to reduce mortality or morbidity from a specific cause. Demonstration of reduction of ACM is not a reasonable expectation. If an intervention is successful in reducing cause-specific mortality, the only additional criterion to apply is that it should not be responsible for appreciably increasing morbidity or mortality from another cause.

Notes

OncoSim is led and supported by the Canadian Partnership Against Cancer, with model development by Statistics Canada, and is made possible through funding by Health Canada. We are grateful to Dr Claude Nadeau at Statistics Canada, who provided valuable assistance explaining the inner workings of OncoSim-Breast and helping us to configure the model runs. We also thank the OncoSim Breast Technical Working Group, whose members are: Dr Anthony B. Miller (chair), Dr Andrew Coldman, Dr Michael Wolfson (University of Ottawa), Dr Nicole Mittmann (Canadian Agency for Drugs and Technologies in Health CADTH), Dr Tallal Younis (Medical Oncology, Dalhousie University), Dr John Bartlett (OICR), Saima Memon, Nicolas Iragorri, Drs Natalie Fitzgerald and Jean Hai Ein Yong (Canadian Partnership Against Cancer), and Drs Keiko Asakawa and Rochelle Garner (Statistics Canada). They provided useful advice and access to the model for this work. Dr Kathleen Pritchard reviewed the manuscript and provided valuable clinical insight.

The authors have no conflicts of interest directly related to this work. Martin Yaffe has a research collaboration on other topics in breast cancer imaging with GE Healthcare through Sunnybrook Research Institute and with The Canadian Partnership Against Cancer and is also a shareholder of Volpara Health Technologies.

The authors would be pleased to entertain reasonable requests for access to the data used to support this Commentary for academic research purposes.

References

1. Black WC, Haggstrom DA, Welch HG. All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst*. 2002;94(3):167–173.
2. Prasad V, Lenzer J, Newman DH. Why cancer screening has never been shown to "save lives"—and what we can do about it. *BMJ*. 2016;352:h6080.
3. Baum M. Harms from breast cancer screening outweigh benefits. *BMJ*. 2013; 346:f385.
4. Saquib N, Saquib J, Ioannidis JP. Does screening for disease save lives in asymptomatic adults? Systematic review of meta-analyses and randomized trials. *Int J Epidemiol*. 2015;44(1):264–277.
5. Nelson HD, Tyne K, Naik A, et al. *Screening for Breast Cancer: Systematic Evidence Review Update for the US Preventive Services Task Force*. Rockville, MD: Agency for Healthcare Research and Quality (US); 2009. https://www.ncbi.nlm.nih.gov/books/NBK36392/. Accessed September 17, 2019.
6. Moher D, Little J, Barbeau P, et al. Breast Cancer Screening: Part A. An Evidence Report to Inform an Update of the Canadian Task Force on Preventive Health Care. 2011. https://canadiantaskforce.ca/wp-content/uploads/2019/02/Systematic-Review-Evidence-Report_v2_FINAL.pdf. Accessed August 13, 2019.
7. Olsen O, Gotzsche PC. Cochrane review on screening for breast cancer with mammography. *Lancet*. 2001;358(9290):1340–1342.
8. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography (review). *Cochrane Libr* 2013;6:CD001877.
9. Gauvreau CL, Fitzgerald NR, Memon S, et al. The OncoSim model: development and use for better decision-making in Canadian cancer control. *Curr Oncol*. 2017;24(6):401–406.
10. Fryback DG, Stout NK, Rosenberg MA, et al. The Wisconsin Breast Cancer Epidemiology Simulation Model. *J Natl Cancer Inst Monogr*. 2006;36:37–47.
11. Alagoz O, Ergun MA, Cevik M, et al. The University of Wisconsin Breast Cancer Epidemiology Simulation Model: an update. *Med Decis Mak*. 2018; 38(1_suppl):99S–111S.
12. SEER. Lifetime risk (Percent) of dying from cancer by site and race/ethnicity: females, total US, 2012–2014 (Table 1.20). https://seer.cancer.gov/csr/1975_2014/results_merged/topic_lifetime_risk.pdf. Accessed November 13, 2019.
13. StatTrek.com. https://stattrek.com/hypothesis-test/difference-in-proportions.aspx. Accessed September 17, 2019.
14. Church TR, Ederer F, Mandel JS. Re: All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst*. 2002;94(11):861–860.
15. Gail MH, Katki HA. Re: All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst*. 2002;94(11):862.
16. Kopans DB, Halpern E. Re: All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst*. 2002;94(11):863.
17. Begg CB, Bach PB. Re: All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst*. 2002;94(11):864.
18. Weiss NS, Koepsell TD. Re: All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst*. 2002;94(11):865.
19. Freedman DA, Petitti DB, Robins JM. On the efficacy of screening for breast cancer. *Int J Epidemiol*. 2004;33(1):43–55.
20. Tabar L, Duffy SW, Yen MF, et al. All-cause mortality among breast cancer patients in a screening trial: support for breast cancer mortality as an end point. *J Med Screen*. 2002;9(4):159–162.
21. Stang A, Jöckel K. The impact of cancer screening on all-cause mortality. What is the best we can expect? *Dtsch Arztebl Int*. 2018;115(29-30):481–486.