CORRESPONDENCE

Benefit and Risk of Mammography Screening

Considerations from an Epidemiological Viewpoint

by Prof. Dr. rer. nat. Nikolaus Becker, Prof. Dr. med. Hans Junkermann
in volume 08/2008

Inaccurate Estimation of Everyday Risks

Conditional probabilities versus absolute frequencies: what is this really all about?

We know from psychological research that the manner in which facts are presented has a major influence on the thoughts and actions of the persons affected by them. Facts are assessed differently if they are presented as a potential gain or a potential loss. This is the central point: whoever controls the manner in which study data are presented will essentially determine whether a proposed screening program will be accepted.

The authors object to presenting the benefit of screening with absolute frequencies and advocate using conditional probabilities instead. They postulate that the general population is well versed in the use of conditional probabilities because certain risks that are permanently present in everyday life require constant assessment (e.g., building insurance).

I do not share this opinion. Many studies in social psychology have shown that most people generally estimate everyday risks inaccurately. Gigerenzer (1) has shown in several studies that neither laypersons nor experts can draw correct conclusions from conditional probabilities. This being known, the authors' argument that the general acceptance of building insurance reflects a correct understanding of conditional probabilities seems rather weak. It is, furthermore, good scientific practice to express study data in terms of frequencies. No critically thinking physician would assess a drug trial, for example, on the basis of relative risk reduction alone; rather, the actual benefits and potential risks would need to be assessed in the light of the frequency parameters NNT (number needed to treat) and NNH (number needed to harm). Why should this be any less obvious when a screening test comes under scrutiny?

DOI: 10.3238/arztebl.2008.0419a

REFERENCES

1. Gigerenzer G: Das Einmaleins der Skepsis. Über den richtigen Umgang mit Zahlen und Risiken. Bvt Berliner Taschenbuch Verlag 2004.

Dr. med. Reinhard Beise
Brunnenstr. 21, 85570 Ottenhofen, Germany
Beise@mail.biosign.de

Overestimation of Reduction in Mortality

Becker and Junkermann mention a 10% overdiagnosis rate as a possible harm resulting from mammography screening. In stating this, the authors have done a much better job than the current mammography information sheet of the German Joint Federal Committee, which is supposed to educate women about mammography but does not mention this very serious harm that can result from an early detection program.

The authors, however, fail to take overdiagnosis into account when they write that the mortality from breast cancer is 31 of 100 unscreened women compared to 20 of 100 screened women, corresponding to a 35% reduction in mortality.

(1) If the overdiagnosis rate is 10%, then, of 100 tumors detected by mammography, 10 would not have reached medical attention. The 20 deaths among these 100 women would therefore all be in the group of 90 women with potentially fatal tumors, which the authors compare to a group of 100 women whose tumors did reach medical attention. If we renormalize the 20 deaths among 90 women to a group of 100 women, we arrive at a figure of 22 deaths, and therefore to a 29% (not 35%) reduction in mortality.

(2) The authors have taken the 10% overdiagnosis rate from the article by Zackrisson et al. (1). These authors looked at the cumulative incidence 15 years after the end of the Malmö study. Immediately after the end of the study, however, the corresponding figure was 24%. Zahl et al. arrived at an even higher figure of 30% (2). The 10% overdiagnosis figure also seems too low because the overdiagnosis rate of in situ carcinoma alone is 13% (3). If the true overdiagnosis rate is 20%, then there would be 25 deaths resulting from 100 potentially fatal tumors detected by mammography, yielding a 20% reduction in mortality compared to no screening. If the true overdiagnosis rate is 30%, the corresponding numbers are 28 deaths and a 10% reduction in mortality. This uncertainty deserves mention just as much as the phenomenon of overdiagnosis itself.

DOI: 10.3238/arztebl.2008.0419b

REFERENCES

1. Zackrisson S, Andersson I, Janzon L et al.: Rate of over-diagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. BMJ 2006; 332: 689–92.
2. Zahl PH, Strand BH, Mæhlen J: Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. BMJ 2004; 328: 921–4.
3. Koch K, Weymayr Ch: Kritik der Krebsfrüherkennung. Der Onkologe 2008; 14: 181–8.

Dr. rer. nat. Christian Weymayr
Schaeferstr. 22, 44623 Herne, Germany
Christian.Weymayr@web.de

Judgement and Decision Processes

According to the authors, many women may choose to participate because they have followed the actuarial principle of multiplying the low probability of the adverse event in question (death from breast cancer) by the high magnitude, as it were, of the event (death).

It is certainly true that many women would still choose to undergo mammography even if they knew that a healthy woman's chance of avoiding death from cancer with the aid of early detection is low. Yet precisely this fact is withheld

from women in the informational materials that they are generally given in the German-speaking countries. Thus, they are not in a position to make the "actuarial" evaluation described. The likelihood of a catastrophically adverse event, such as overdiagnosis followed by overtreatment, is likewise low but obviously should also be part of the evaluation and decision process.

It has been well documented empirically that decisions depend on the nature and extent of the information provided. Presenting the effect of the intervention only in terms of relative risk creates an unrealistically positive impression (1). Patients consider themselves well-informed if they know absolute numbers and absolute risks (2). Varying presentations of identical facts – "half empty" versus "half full" – can lead patients to make different decisions. I think there can only be one conclusion: the judgement about the probabilities of benefit and harm should be left up to each individual woman on the basis of comprehensive information that does not predetermine her ultimate decision. There is no good reason to withhold understandable information from women, as has been done up to the present. Understandable information that should be communicated includes at least the following: the risk of breast cancer in the individual woman's age group, the likelihood of avoiding death from cancer by early detection, the rates of false positive and false negative findings, and the danger of overdiagnosis leading to overtreatment (3).

DOI: 10.3238/arztebl.2008.0420a

REFERENCES

1. Covey J: A Meta-analysis of the Effects of Presenting Treatment Benefits in Different Formats. Med Decis Making 2007; 27: 638–54.
2. Koch K, Scheibler F: Einstellungen und Informationsstand zur Früherkennung: Informiert und doch getäuscht? In: Böcken J, Braun B, Amhof R: Gesundheitsmonitor 2007. Verlag Bertelsmann Stiftung 2007.
3. Koch K, Weymayr C: Kritik der Krebsfrüherkennung. Der Onkologe 2008; 14: 181–8.

Prof. Dr. med. David Klemperer
Deutsches Netzwerk Evidenzbasierte Medizin e.V.
– Fachbereich Patienteninformation und Patientenbeteiligung –
Hochschule Regensburg
Seybothstr. 2, 93053 Regensburg, Germany
http://www.dnebm-patienteninformation.de
david.klemperer@soz.fh-regensburg.de

Shared Decision-Making

Women would like to know what they have to gain from screening. Around the world, this question is answered with data on the number of deaths prevented per number of women screened. It thus seems strange that the authors would claim that the reference figures presented by Mühlhauser and Höldke are somehow artificial.

Women would certainly also like to know what the chance of successful treatment would be in the event that a tumor is discovered. The authors answer this question as follows: over the course of 10 years, out of 100 screened women found to have breast cancer, 20 will die instead of 31. Thus, there will be 11 fewer deaths out of 100 screened women in 10 years, or, rounding off, 1 woman saved out of 10. Who could fail to be impressed by this?

The authors' response to the question of treatment can be made still more precise once the histological findings are available, because the stage of disease determines whether the chance of survival will be greater or smaller.

The benefit of screening (not of potential later treatment) can be calculated from the article's table 1. Out of 100 000 screened women, the number who go on to die of breast cancer in the next 10 years is reduced from 155 (the figure among non-screened women) to 101, i.e., a reduction by 54/100 000 in 10 years. Rounding off, this corresponds to 1 woman saved out of 2000 in 10 years. Does anyone still want to be screened?

This discrepancy between the internationally standard manner of presentation and that chosen by Becker and Junkermann encapsulates the entire problem of the way we choose to deal with participative decision-making.

If we wish to free the concept of shared decision-making from pure political correctness, then we must tell women what individual benefit they can expect from screening, as well as what harm might come to them from it (something that the article does not completely address) – even if doing so carries with it the danger that fewer women will choose to be screened.

Society's desire for a high rate of screening, so that the maximal public health benefit can be achieved, is another matter. This desire, too, has a rationale behind it. We physicians must choose which side we prefer to stand on.

DOI: 10.3238/arztebl.2008.0420b

Prof. Dr. med. Heinz-Harald Abholz
Leiter der Abteilung für Allgemeinmedizin
Heinrich-Heine-Universität Düsseldorf
Moorenstr. 5, 40225 Düsseldorf, Germany
Abholz@med.uni-duesseldorf.de

Discussing the Benefit-Harm Debate

Becker and Junkermann's review article sheds light on a very important matter of health policy. Yet the problems begin as early as the title, which sets up a contrast between benefit and risk rather than between benefit and harm. The impression is conveyed that the benefit of mammography screening stands against a (merely statistical) risk. This, of course, is not the case; both the "benefit" and the risk are statistical extrapolations of an expected benefit and an expected harm. We consider the authors' assessment of benefit and harm to be imbalanced for a number of other reasons as well:

The authors illustrate the benefit of screening with a fictitious sample of 100 000 women who undergo screening every 2 years a total of 10 times. The harm, however, is illustrated with a fictitious group of only 1 000 women. For a group of 100 000 women in which screening would lead to the prevention of 540 deaths from breast cancer, the following types of harm would be expected to occur:

- 22 300 to 36 300 women would have false positive mammograms,
- 500 women would be overdiagnosed with breast cancer,
- 6300 women would have a false positive indication for a breast biopsy,
- about 500 women would go on to have a breast operation with a benign histological finding,
- and there would be 10 to 240 cases of radiation-induced breast cancer.

Thus, every third woman undergoing screening every 2 years between the ages of 50 and 69 would be subjected to unnecessary worry, many women would undergo invasive diagnostic or therapeutic procedures on the basis of a false positive finding or an overdiagnosis, and a not inconsiderable number of breast cancers would actually be induced. There is no critical discussion of the quality of life of these unnecessarily disturbed women and their families; such a discussion is indispensable if the benefits and harms of screening are to be considered fairly.

Thus, overall, the benefit-harm debate seems to lack balance.

DOI: 10.3238/arztebl.2008.0421a

Prof. Dr. rer. nat. Karl-Heinz Jöckel
Dr. med. Barbara Hoffmann, MPH
Institut für Medizinische Informatik, Biometrie
und Epidemiologie
Medizinische Fakultät
Universität Duisburg-Essen
Hufelandstr. 55, 45147 Essen, Germany
k-h.joeckel@uk-essen.de

Prof. Dr. med. Andreas Stang, MPH
Sektion Klinische Epidemiologie
Institut für Medizinische Epidemiologie, Biometrie und Informatik
Medizinische Fakultät
Martin-Luther-Universität Halle-Wittenberg
Magdeburger Str. 8, 06097 Halle (Saale), Germany

Stage-Dependent Lethality

Becker and Junkermann's article on the benefits and risks of mammography screening is heralded on the title page with the words, "Mammography: Screening Lowers Breast Cancer Mortality by 35%." This statement, which is also one of the authors' three main conclusions, is misleading in two respects:

(1) The authors are speaking, not of mortality in the usual epidemiological sense of the specific probability of death in a defined population group, but rather of lethality, i.e., the conditional probability of death over a certain period of observation after a particular diagnosis has been made and (in the normal situation) followed up with treatment (1).

(2) The authors attempt to circumvent the statistical problem of lead-time bias, i.e., earlier diagnosis in the screened than in the unscreened population, by measuring the effect of screening in terms of mortality/lethality. They write: "Mortality is the only variable quantifiable without bias for studies on the effectiveness of early detection activities." Yet the endpoint that they consider, i.e., lethality after the diagnosis of breast cancer of any type, is variable depending on the stage of the tumor (according to the WHO tumor classification) at the time the diagnosis is made. The distribution of cases in the screened population is selected in favor of earlier and prognostically more favorable stages, and a slower course of disease, in comparison to the unscreened population. Thus, a consideration of mortality/lethality after the disease is diagnosed does not eliminate lead-time bias. To do this, the authors would have had to present and discuss stage-dependent lethality rates over a defined period of time.

DOI: 10.3238/arztebl.2008.0421b

Prof. Dr. med. Friedrich Wilhelm Schwartz
Direktor des Instituts für Epidemiologie,
Sozialmedizin und Gesundheitssystemforschung
MHH
30623 Hannover, Germany
Epidemiologie@mh-hannover.de

Randomized Controlled Trials Are Needed

The authors' attempt to show why mammography screening should be evaluated on the basis of epidemiological studies rather than randomized controlled clinical trials (RCTs). Yet it is precisely because an epidemiological assessment is not possible that RCTs are needed according to internationally accepted standards. Furthermore, RCTs do indeed exist for mammography screening, as does a high-quality Cochrane Review by an independent group of authors (1).

Epidemiological analyses tend to overstate the benefit of screening for the following reasons (2):

(1) Screened women differ from unscreened women. They are a priori healthier and better educated and therefore have a better chance of survival, both in general and for the particular disease for which they are being screened.

(2) Screening tends to detect relatively benign and slowly growing tumors.

(3) Early diagnosis often merely means prolonging the time that the patient lives with breast cancer, rather than improving the prognosis.

(4) Screening detects cases of breast cancer that would never have come to attention without it (overdiagnosis). It is thus inappropriate to compare a group of women whose cancers were diagnosed by screening to a group of unscreened women whose cancers came to medical attention in other ways. It should come as no surprise, therefore, that the Cochrane Review estimates the benefit of mammography screening to be much less, and the resulting harm to be much greater, than stated by Becker and Junkermann (1, 3).

The authors argue that the harm due to breast cancer itself ought to be taken into consideration. When such calculations are made, however, overdiagnoses must be considered also, as well as the overall death rate due to cancer and the overall mortality. These are not improved by screening (1, 3). The low sensitivity of mammography screening in the German pilot projects is not mentioned either (3); this reduces the benefit of screening.

DOI: 10.3238/arztebl.2008.0421c

REFERENCES

1. Gøtzsche P, Nielsen M: Screening for breast cancer with mammography. Cochrane Database of Systematic Reviews 2006, Issue 4.
2. Sackett DL, Haynes RB, Guyett GH, Tugwell P: Clinical epidemiology. A basic science for clinical medicine. 2nd edition. Little, Brown and Company. Boston, Toronto, London 1991.
3. Mühlhauser I: Ist Vorbeugen besser als Heilen? Dtsch Arztebl 2007; 104: A 1805–7.

Prof. Dr. med. Ingrid Mühlhauser
Universität Hamburg
MIN Fakultät, Fachwissenschaft Gesundheit
Marthin-Luther-King Platz 6, 20146 Hamburg, Germany
ingrid_muehlhauser@uni-hamburg.de

In Reply:

Our article had three main points:

(1) The effect of mammography screening in increasing treatment effectiveness and prolonging survival should be presented from the viewpoint of the individual woman invited for screening, while the scientific demonstration of its efficacy must be based on the observation of mortality.

(2) Some authors' attempts to present the effect of screening intelligibly have resulted not just in altered formulations, but in substantially changed and inaccurate information.

(3) The screening debate fails to take account of the different definitions of risk which may play a role in the assessment of benefit and harms, which differ as to whether the severity of damage associated with the undesired event is taken into account (as is the case with insurance) or not (as is the case in epidemiology).

Mühlhauser's letter repeats the familiar argument that the effect of screening can only be demonstrated scientifically by means of randomized epidemiological studies with mortality as endpoint, which we address (1–3).

We agree in general with Beise's comments. We did point out in our article, however, that the effect of screening is, by definition conditional: if an as yet undetected malignancy is present, then it might be advantageous to detect it as early as possible by screening and to treat it (3). Our proposed solution is that the benefit of screening in case such a disease is present should be presented in terms of absolute frequencies, just as Beise suggests. The reference to insurance has nothing to do with conditional probabilities, but rather with the varying definitions of "risk": in epidemiology, risk is the probability of an adverse event, regardless of the magnitude of the associated damage; in insurance, the definition of risk takes this magnitude into account.

Weymayr fails to recognize that our presentation is the computational transformation of the results of the randomized studies with mortality as an outcome into absolute rates of survival. Thus, net effects are presented. Overdiagnosis need not be subtracted, because one does not die from a disease which remains subclinical for the remainder of one's life.

Klemperer's letter from the German Network for Evidence-Based Medicine concurs with our goal of yielding patient information that is as accurate and as comprehensive as possible. We believe it is wrong, however, to take the "magnitude of the damage" in case of a cancer diagnosis limited to a possible lethal outcome, but that the possible implications of treatment and their effects on the patient's quality of life must also be taken into account. The probability of disease should be communicated to the patient as a separate quantity, unaffected by screening.

Abholz extends the discussion of mortality reduction by the early detection of existing disease to individuals who are free of the disease and therefore not at risk of dying from it. This is exactly what we criticized in our article. The probability of getting the disease is relatively low (about 5%); however, given the disease, 31 of 100 will die of it within 10 years without screening, 23 if screening is offered, and 20 if they participated.

Jöckel et al. raise objections to the title of the article and the scaling for the used quantities. The title of the article arose in the course of pre-publication review; originally, we proposed: "Risk communication in mammography screening." Colleagues in epidemiology objected to the term "damage" (or harm, "Schaden") as being not entirely free of emotional connotations; therefore, in the article as printed, we consistently referred to "disadvantageous effects," which we contrasted with benefits. As for the scaling of quantities, due to small numbers we gave incidence figures in the usual epidemiological form, i.e., per 100 000 individuals (5); for more frequent events, we used the scaling usual in that area, i.e., per 1000 individuals (4). The benefit in terms of avoided deaths was transformed into the per 1000 scale in table 1 (bottom) and was contrasted in the text on this scale to the disadvantageous effects "breast cancer surgery with benign findings" and "overdiagnosis." We also pointed out that the most common disadvantageous effect by far is a false positive finding whose frequency was shown in table 2. The authors of the letter present these figures on a larger scale which makes them obviously no more "balanced," but merely larger. As for the effect of false positive findings on the quality of life, we have already dealt with this issue in multiple previous publications (1, 2). The focus in the current article was on the correct interpretation of study results.

Schwartz's first assertion is correct: it was indeed our intention to point out the conditional nature of figures on mortality reduction by screening, and to translate them computationally into lethality figures. Since this was not based on empirical data, lead-time bias, relevant for empirical evaluation, was omitted from consideration (1–3). Selection bias in relation to disease stage is referred to as length bias and does not affect the outcome of offering screening, as opposed to actual participation. DOI: 10.3238/arztebl.2008.0422

REFERENCES

1. Becker N: Screening aus epidemiologischer Sicht. Radiologe 2002; 42: 592–600.
2. Becker N: Die Rolle der epidemiologischen Qualitätsparameter im Mammographie-Screening. Radiologe 2006; 46: 984–992.
3. Morrison AS: Screening in Chronic Disease. Second edition. Monographs in Epidemiology and Biostatistics, Volume 19. Oxford University Press, New York, Oxford 1992.
4. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L, Puthaar E (eds.): European guidelines for quality assurance in breast cancer screening and diagnosis – Fourth Edition. Luxembourg: Office for Official Publications of the European Communities 2006.
5. Robert-Koch-Institut und die Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID) (eds.): Krebs in Deutschland 2003–2004. Häufigkeiten und Trends. Berlin 2008.

Prof. Dr. rer. nat. Nikolaus Becker
Deutsches Krebsforschungszentrum, Abteilung Krebsepidemiologie
Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
n.becker@dkfz.de

Dr. med. Hans Junkermann
Universitätsfrauenklinik, Sektion Senologische Diagnostik
Voßstr. 9, 69115 Heidelberg, Germany

Conflict of interest statement
The authors of all letters and the reply declare that they have no conflict of interest as defined by the International Committee of Medical Journal Editors.