CURRENT ONCOLOGY

To screen or not to screen for breast cancer? How do modelling studies answer the question?

R.G. Koleva-Kolarova MSc,* Z. Zhan MSc,* M.J.W. Greuter PhD,† T.L. Feenstra PhD,*‡ and G.H. De Bock PhD*

ABSTRACT

Breast cancer screening is a topic of hot debate, and currently no general consensus has been reached on starting and ending ages and screening intervals, in part because of a lack of precise estimations of the benefit–harm ratio. Simulation models are often applied to account for the expected benefits and harms of regular screening; however, the degree to which the model outcomes are reliable is not clear. In a recent systematic review, we therefore aimed to assess the quality of published simulation models for breast cancer screening of the general population. The models were scored according to a framework for qualitative assessment. We distinguished seven original models that utilized a common model type, modelling approach, and input parameters. The models predicted the benefit of regular screening in terms of mortality reduction; and overall, their estimates compared well to estimates of mortality reduction from randomized controlled trials. However, the models did not report on the expected harms associated with regular screening. We found that current simulation models for population breast cancer screening are prone to many pitfalls; their outcomes bear a high overall risk of bias, mainly because of a lack of systematic evaluation of evidence to calibrate the input parameters and a lack of external validation. Our recommendations concerning future modelling are therefore to use systematically evaluated data for the calibration of input parameters, to perform external validation of model outcomes, and to account for both the expected benefits and the expected harms so as to provide a clear balance and cost-effectiveness estimation and to adequately inform decision-makers.

Key Words Breast cancer, screening, modelling, mortality reduction

INTRODUCTION

Breast cancer screening is a topic of hot debate, and currently no general consensus has been reached on starting and ending ages and screening intervals. The lack of general consensus is mainly attributable to the varying opinions and estimations of the precise benefits and harms related to regular mammographic screening[1–3]. It should also be pointed out that not only are precise benefit–harm ratio estimates absent, but intense controversy continues to surround estimates of overdiagnosis, lead time and mean sojourn time, and (background) breast cancer incidence. Also, most modelling studies have failed to incorporate proper sensitivity analyses (univariate or probabilistic) for sojourn time[4]. Constant attempts are being made to answer the questions about the balance between what women gain when they participate in regular breast cancer screening programs on the one hand and what harms are associated with regular screening on the other.

Modelling is widely applied to study such issues with respect to breast cancer screening. However, the degree to which simulation models produce reliable answers has not been thoroughly investigated. Therefore, in a recent systematic review, we aimed to assess the quality of published simulation models for breast cancer screening in the general population[5]. Here, we summarize the results of that review.

WHAT WE ASSESSED

Our systematic analysis included models that have been applied more than once to assess the mortality reduction and cost-effectiveness of regular screening. We scored the models according to a self-developed framework for qualitative assessment that included model type; input parameters; modelling approach, transparency of input data sources and assumptions, sensitivity analyses, and risk of bias; validation; and outcomes. We also assessed the predicted benefits, harms, and cost-effectiveness. The model-predicted mortality reductions (MRs) were compared with estimates from randomized controlled trials (RCTs)[3], and the cost-effectiveness estimates, with World

Health Organization criteria based on the per-capita gross domestic product of the country[6]. However, we should note that the Gøtzsche review[3] has not itself been without some considerable controversy (other reviews like it from the Independent U.K. Panel on Breast Cancer Screening[7] diverge significantly from the conclusions of the Gøtzsche review as to both mortality reduction and overdiagnosis), and estimates based on individual patient data are considerably more positive with respect to the benefit–harm ratio of mammographic screening[8].

MAIN FINDINGS AND DISCUSSION

Our search identified 7 original models[9–15]. Most models used in mammography evaluations were developed within the U.S. National Cancer Institute's cisnet (Cancer Intervention and Surveillance Modeling Network) framework, whose 6 models can be categorized into 2 invasive-cancer-only models (no ductal carcinoma *in situ*), represented by the Stanford and Dana–Farber models, and 4 invasive and noninvasive models, represented by Erasmus MC, Georgetown, MD Anderson, and Wisconsin, with some fine subdistinctions within those categories[16].

The mrs predicted by the foregoing models (11%–24%) overestimated the mr attributable to screening as estimated by optimal rcts [10%; 95% confidence interval (ci): –2% to 21%][3], but 5 models were within the 95% ci[9–12,14]. Further, the mr emerging from the original models compared relatively well with the mr emerging from suboptimal trials (25%; 95% ci: 17% to 33%) and from all trials (19%; 95% ci: 13% to 26%)[3].

The original models reported neither on harms associated with regular mammographic screening nor on cost-effectiveness. However, the original models were subsequently used in studies that assessed the potential harms of screening as well as the cost-effectiveness of various screening regimens. According to the simulation studies, most screening scenarios met the World Health Organization's criteria for cost-effectiveness[6].

The analyzed original models had several advantages. They all were classified into the individual-level group[17], which is a modelling type generally assumed to be simple and flexible and productive of reliable outcomes. The only shortcoming of this modelling type is that repeated runs of the model are required to obtain a stable outcome. The original models had common input parameters and used the same epidemiology database to calibrate their values, which facilitated internal validation of the models (that is, comparing the model output to the database used for the input parameters) and comparisons between models (that is, cross-validation). With one exception, the original models applied a tumour growth approach to model disease progression and applied the aggregated population breast cancer incidence rate, which was a reasonable way to quantify mortality reduction on the general population level.

Despite those advantages, and estimations of mr that were in range with estimates from rcts, the analyzed simulation models demonstrated some disadvantages that could compromise the reliability of their outcomes. The biggest shortcoming was a lack of external validation:

that is, comparisons of their outcomes with an independent database different from the one used to populate the input parameters of the model. Another disadvantage was that the models lacked systematic selection and evaluation of sources for calibration of their input data. Sensitivity analyses were, however, performed to account for the uncertainty involved in calibrating the input parameters and to test model performance. Further, basing breast cancer incidence only on the age of the simulated populations (that is, an aggregated approach) failed to encompass the change in breast cancer incidence because of other risk factors such as increased age of first birth, alcohol consumption and smoking, oral contraceptive use, and body mass index[18]. In addition, when modelling disease progression, the original models used only the tumour progression model, which could encompass neither the non-chronologic development of real-life tumours nor the differences in lead time for invasive and noninvasive tumours. Furthermore, some tumours that never surface clinically and cause no complications for the health of a woman—the so-called indolent tumours[19]—were not specifically modelled and accounted for in the original model analyses.

The original models did not report on potential harms. On the one hand, they were developed to study expected mr, and the harms of regular screening might potentially be outside the scope of their investigation. However, not taking into account harms such as overdiagnosis[4] and radiation-induced tumours could result in bias when estimating the expected mr.

CONCLUSIONS

We are aware that our systematic review is limited in scope given that it includes only simulation models developed by one research consortium. However, those models have often been used to inform medical decision-making and should in principle produce reliable outcomes. Current simulation models for breast cancer screening are prone to many pitfalls, and their outcomes carry a high overall risk of bias, mainly because of their lack of systematic evaluation of the evidence to calibrate the input parameters and their lack of external validation[5].

Our recommendation concerning future modelling is therefore to select a modelling type that is flexible and that produces stable outcomes; to use systematically evaluated data for calibration of the input parameters; to apply aggregated incidence together with individual risk factors; to allow for changing lead time depending on the type of tumour (that is, ductal carcinoma *in situ*, invasive, noninvasive, indolent); to perform external validation of model outcomes; and to account for both the expected benefits and expected harms so as to provide a clear balance and cost-effectiveness estimation and to adequately inform decision-makers. In addition, there are important groups (such as women 74 years of age and older) for whom minimal or no rct data are available where modelling can be well deployed[20].

CONFLICT OF INTEREST DISCLOSURES
We have read and understood *Current Oncology*'s policy on disclosing conflicts of interest, and we declare that we have none.

AUTHOR AFFILIATIONS

*Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands; †Department of Radiology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands; ‡rivm, Bilthoven, Netherlands.

REFERENCES

1. Djulbegovic B, Lyman GH. Screening mammography at 40–49 years: regret or no regret? *Lancet* 2006;368:2035–7.
2. Glasziou P, Houssami N. The evidence base for breast cancer screening. *Prev Med* 2011;53:100–2.
3. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2013;6:CD001877.
4. Carter JL, Coletti RJ, Harris RP. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. *BMJ* 2015;350:g7773.
5. Koleva-Kolarova RG, Zhan Z, Greuter MJ, Feenstra TL, De Bock GH. Simulation models in population breast cancer screening: a systematic review. *Breast* 2015;24:354–63.
6. World Health Organization (who). *Macroeconomics and Health: Investing in Health for Economic Development.* Report of the Commission on Macroeconomics and Health. Sachs JD, chair. Geneva, Switzerland: who; 2001.
7. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. *Br J Cancer* 2013;108:2205–40.
8. Kaniklidis C on behalf of the No Surrender Breast Cancer Foundation. Beyond the mammography debate: a moderate perspective. *Curr Oncol* 2015;22:220–9.
9. van der Maas PJ, de Koning HJ, van Ineveld BM, *et al.* The cost-effectiveness of breast cancer screening. *Int J Cancer* 1989;43:1055–60.
10. Tan SY, van Oortmarssen GJ, de Koning HJ, Boer R, Habbema JD. The miscan-Fadia continuous tumor growth model for breast cancer. *J Natl Cancer Inst Monogr* 2006;(36):56–65.
11. Mandelblatt J, Schechter CB, Lawrence W, Yi B, Cullen J. The spectrum population model of the impact of screening and treatment on U.S. breast cancer trends from 1975 to 2000: principles and practice of the model methods. *J Natl Cancer Inst Monogr* 2006;(36):47–55.
12. Fryback DG, Stout NK, Rosenberg MA, Trentham-Dietz A, Kuruchittham V, Remington PL. The Wisconsin Breast Cancer Epidemiology Simulation Model. *J Natl Cancer Inst Monogr* 2006;(36):37–47.
13. Plevritis SK, Sigal BM, Salzman P, Rosenberg J, Glynn P. A stochastic simulation model of U.S. breast cancer mortality trends from 1975 to 2000. *J Natl Cancer Inst Monogr* 2006;(36):86–95.
14. Lee S, Zelen M. A stochastic model for predicting the mortality of breast cancer. *J Natl Cancer Inst Monogr* 2006;(36):79–86.
15. Berry DA, Inoue L, Shen Y, *et al.* Modeling the impact of treatment and screening on U.S. breast cancer mortality: a Bayesian approach. *J Natl Cancer Inst Monogr* 2006;(36):30–6.
16. Mandelblatt JS, Cronin KA, Bailey S, *et al.* on behalf of the Breast Cancer Working Group of the Cancer Intervention and the Surveillance Modeling Network. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Ann Intern Med* 2009;151:738–47.
17. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ* 2006;15:1295–310.
18. Soerjomataram I, Pukkala E, Brenner H, Coebergh JW. On the avoidability of breast cancer in industrialized societies: older mean age at first birth as an indicator of excess breast cancer risk. *Breast Cancer Res Treat* 2008;111:297–302.
19. Feinleib M, Zelen M. Some pitfalls in the evaluation of screening programs. *Arch Environ Health* 1969;19:412–15.
20. Kramer BS, Elmore JG. Projecting the benefits and harms of mammography using statistical models: proof or proofiness? *J Natl Cancer Inst* 2015;107:pii:djv145.