Medicine®

OPEN

Clarifying the debate on population-based screening for breast cancer with mammography
A systematic review of randomized controlled trials on mammography with Bayesian meta-analysis and causal model

Tony Hsiu-Hsi Chen, PhD[a],[*], Amy Ming-Fang Yen, PhD[b], Jean Ching-Yuan Fann, PhD[c], Paula Gordon, PhD[d], Sam Li-Sheng Chen, PhD[b], Sherry Yueh-Hsia Chiu, PhD[e], Chen-Yang Hsu, PhD[a], King-Jen Chang, PhD[f], Won-Chul Lee, PhD[g], Khay Guan Yeoh, PhD[h], Hiroshi Saito, PhD[i], Supannee Promthet, PhD[j], Chisato Hamashima, PhD[k], Alimin Maidin, PhD[l], Fredie Robinson, PhD[m], Li-Zhong Zhao, MD[n]

Abstract

Background: The recent controversy about using mammography to screen for breast cancer based on randomized controlled trials over 3 decades in Western countries has not only eclipsed the paradigm of evidence-based medicine, but also puts health decision-makers in countries where breast cancer screening is still being considered in a dilemma to adopt or abandon such a well-established screening modality.

Methods: We reanalyzed the empirical data from the Health Insurance Plan trial in 1963 to the UK age trial in 1991 and their follow-up data published until 2015. We first performed Bayesian conjugated meta-analyses on the heterogeneity of attendance rate, sensitivity, and over-detection and their impacts on advanced stage breast cancer and death from breast cancer across trials using Bayesian Poisson fixed- and random-effect regression model. Bayesian meta-analysis of causal model was then developed to assess a cascade of causal relationships regarding the impact of both attendance and sensitivity on 2 main outcomes.

Results: The causes of heterogeneity responsible for the disparities across the trials were clearly manifested in 3 components. The attendance rate ranged from 61.3% to 90.4%. The sensitivity estimates show substantial variation from 57.26% to 87.97% but improved with time from 64% in 1963 to 82% in 1980 when Bayesian conjugated meta-analysis was conducted in chronological order. The percentage of over-detection shows a wide range from 0% to 28%, adjusting for long lead-time. The impacts of the attendance rate and sensitivity on the 2 main outcomes were statistically significant. Causal inference made by linking these causal relationships with emphasis on the heterogeneity of the attendance rate and sensitivity accounted for the variation in the reduction of advanced breast cancer (none-30%) and of mortality (none-31%). We estimated a 33% (95% CI: 24–42%) and 13% (95% CI: 6–20%) breast cancer mortality reduction for the best scenario (90% attendance rate and 95% sensitivity) and the poor scenario (30% attendance rate and 55% sensitivity), respectively.

Conclusion: Elucidating the scenarios from high to low performance and learning from the experiences of these trials helps screening policy-makers contemplate on how to avoid errors made in ineffective studies and emulate the effective studies to save women lives.

[a] Graduate Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, [b] School of Oral Hygiene, College of Oral Medicine, Taipei Medical University, Taipei, [c] Department of Health Industry Management, School of Healthcare Management, Kainan University, Tao-Yuan, Taiwan, [d] BC Women's Hospital, Vancouver, British Columbia, [e] Department of Health Care Management, College of Management, Chang Gung University, Tao-Yuan, [f] Cheng Ching General Hospital, Taichung, Taiwan, [g] Department of Preventive Medicine, College of Medicine, Catholic University of Korea, Seoul, Korea, [h] Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore, [i] Screening Assessment & Management Division, Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo, Japan, [j] Department of Epidemiology, Faculty of Public Health, Khon Kaen University, Khon Kaen, Thailand, [k] Cancer Screening Assessment and Management Division, Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo, Japan, [l] School of Public Health, Makassar University, Makassar, Indonesia, [m] Community Treatment Centre, Universiti Malaysia Sabah, Sabah, Malaysia, [n] Department of Epidemiology, Tianjin Colorectal and Anal Disease Research Institute, Tianjin, China.

[*] Correspondence: Tony Hsiu-Hsi Chen, Graduate Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Room 533, No. 17, Hsuchow Road, Taipei 100, Taiwan (e-mail: chenlin@ntu.edu.tw).

1. Introduction

Evidence in favor of breast cancer screening with mammography has been demonstrated by a series of randomized controlled trials (RCTs) in various countries worldwide. These included the Health Insurance Plan (HIP) of Greater New York in the USA, the 5 Swedish trials, the Canadian Trial, the Edinburgh trial in UK, and the UK age trial. The efficacy of combining the results of various trials has been systematically reviewed by a series of meta-analyses from the 5 Swedish trials in 1993[1] through the Independent UK Panel on Breast Cancer Screening study in 2012.[2] The conclusion drawn from these meta-analyses are by no means consistent. The meta-analyses demonstrating a benefit of breast cancer screening with mammography included the overview of all 5 Swedish trials conducted by Nyström et al in 1993,[1] the 5 Swedish trials combined with HIP, the Canadian trial, and the Edinburgh trial reappraised by Smith et al,[3] and the UK independent review conducted in 2012.[2] The meta-analyses claiming a lack of benefit of breast cancer screening with mammography commenced from a study conducted by Gøtzsche and Olsen in 2000.[4] Since then, the debate over mammographic screening has expanded in the medical literature with evidence from the Swedish Two-County Trial, which showed high benefit,[5] to the Canadian trial, which showed low benefit.[6]

Health policy-makers in other countries are understandably puzzled by such a discrepancy and must consider both ends of these meta-analyses when they are called on to design and plan a population-based breast cancer screening program in response to trends of increasing incidence of breast cancer but low awareness of early detection in Asian countries. The question has been incessantly asked by the health authority of "whether we follow suit to conduct mass screening for breast cancer with mammography, the screening tool developed and strongly recommended for early detection of breast cancer since 1970."

To clarify this issue, we reappraised the content of each trial included in the meta-analyses, without any being excluded at the authors' discretion, by using different statistical criteria. Important information from these meta-analyses is that relevant characteristics across those trials were fraught with heterogeneity in many aspects. These included the characteristics of target population, study design (e.g., interscreening interval), attendance rate, factors related to the quality of screening (e.g., sensitivity, specificity, and over-detection), and treatment and therapeutic components. Although all of the meta-analyses have shown a lack of heterogeneity with statistical criteria, the disparity of different aspects across the trials is difficult to understand merely on the basis of statistical tests for heterogeneity. Putting much emphasis on the aspect of statistical heterogeneity precludes one from understanding the value of each trial's contribution to the elucidation of the benefit and harm of breast cancer screening with mammography.

In order to systematically pinpoint the causes of heterogeneity across trials with emphasis on clinical and public health significance, the objective of this review was to clarify the recent debate on mammography screening when used at the population level by elucidating a cascade of causal relationships between the 2 main parameters of participation rate and sensitivity, and the outcome of advanced breast cancer and breast cancer mortality based on the Bayesian meta-analysis and causal model. It is hoped that this review will provide health policy-makers in countries worldwide a better understanding of why there is disparity across these trials and learn the value of each trial in determining the benefit and harm of the use of mammography to screen for breast cancer.

2. Methods

2.1. Data sources, search strategy, and selection criteria

This review was conducted and reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Statement.[7] Because this study analyzed publicly available tabular data, no protocol review or informed consent was needed.

Data for this review were identified by searching the PubMed/Medline, Embase, Scopus, and Cochrane library databases using the search terms "mammography" AND "breast cancer" AND "screening" AND "randomized controlled trial." Only articles published in English between 1970 and 2015 were included. We also conducted a manual search of references from the retrieved articles and relevant new articles.

The data retrieval was performed by a panel of experts independently. Any discrepancy between reviewers was discussed in a group meeting until a consensus was met. Inclusion criteria applied to this review included the following rules: randomized controlled design on breast cancer screening with mammography; the eligible population is the underlying average-risk women; available tabular data with information on attendance rate and detection mode (e.g., screen-detected breast cancer, interval cancers (ICs), and breast cancer from nonparticipants); reporting the results of advanced breast cancer and breast cancer mortality after the follow-up of trials. Exclusion criteria were mammography screening applied to high-risk women, the articles related to the main study but used for other purposes (e.g., an epidemiological study on the association between risk factors and the risk of breast cancer) unrelated to the aim of this study; those subsidiary studies limited to subgroup analysis (e.g., age-specific results which had not been designed at the beginning of the trial).

2.2. Data collection and quality assessment

Assessment of abstracted data was performed by a panel of experts involved in the evaluation of breast cancer screening over 15 years. Tabular data from the published articles were assessed to regenerate information used for Bayesian meta-analysis, including the invited population and participants, detection modes, number of breast cancer diagnoses (including non-advanced and advanced breast cancer), number of deaths from breast cancer, study duration, and the interscreening interval.

The quality of these population-based studies of screening for breast cancer with mammography was assessed by using the Jadad scoring system that has been widely used for reporting RCTs.[8] According to this guideline, the Jadad scores are almost equal to 3 for all included studies. Two points were derived from the RCT design, and 1 point was derived from an account of all participants for all trials. It is impossible to blind the participant to the invited and uninvited group by mammography. The 2 points attributed to blinding requested by the Jadad scale could not be assigned. However, because the trials included in this review are large population-based mammography screening studies, the inherent property of the radiological reading before the outcomes ascertained by clinical–pathological diagnosis, as well as the independence between readers and equipment operators who capture the mammogram images, is equivalent to the spirit of blinding. It should be also noted that all of these trials were conducted before the era of guidelines used in RCTs, such as the CONSORT (CONsolidated Standards of Reporting Trials) checklist proposed in the mid-1990s.[9] Moreover, they are population-based trials, and the number of participants is often considerably larger than the number of participants involved in drug clinical trials. It is therefore difficult to use scoring to weight the quality of trials included in this review. In addition, it could be argued that publication bias might exist. Unlike drug clinical trials, population-based RCTs are unlikely to be unreported. The included RCTs in this review encompass all existing evidence on mammography screening. Therefore, the publication bias is not problematic in our meta-analysis.

2.3. Framework of causal model

Figure 1 illustrates a framework of causal relationships between 3 key components, including attendance rate, the sensitivity using the indicator (1-incidence rate of interval cancer/expected incidence rate), and the degree of over-detection, as well as the sequelae of their influences. In addition to conducting Bayesian meta-analyses on the heterogeneity of these 3 key components, we related the heterogeneity of attendance rate and sensitivity across trials to the reduction in advanced stage breast cancer and breast cancer mortality by using Bayesian meta-analyses of Poisson fixed- and random-effect regression models. We then developed a Bayesian causal model for assessing a cascade of causal relationships regarding the impact of both attendance and sensitivity on advanced cancer and death from breast cancer.

2.4. Bayesian meta-analysis of attendance rate, sensitivity, and over-detection

2.4.1. Attendance rate. The meta-analysis to determine the pooled estimate of attendance rate based on data from the retrieved trials was performed using Bayesian conjugated beta-binomial distribution from the first HIP trial to the latest UK age trial, assuming these data were exchangeable (exchangeable assumption) and presented in chronological order. The impact of attendance rate on the rate of advanced stage breast cancer and mortality from breast cancer were then modeled by using Bayesian Poisson regression fixed- and random-effect models,[10] with adjustment for the logarithm of counts of IC, the calendar year of conducting the trial, interscreening interval, and age. The random-effect was incorporated to capture the heterogeneity across trials.

2.4.2. Sensitivity as a function of interval cancer and expected incidence rate. To assess the heterogeneity of performance of mammography screening, particularly sensitivity,
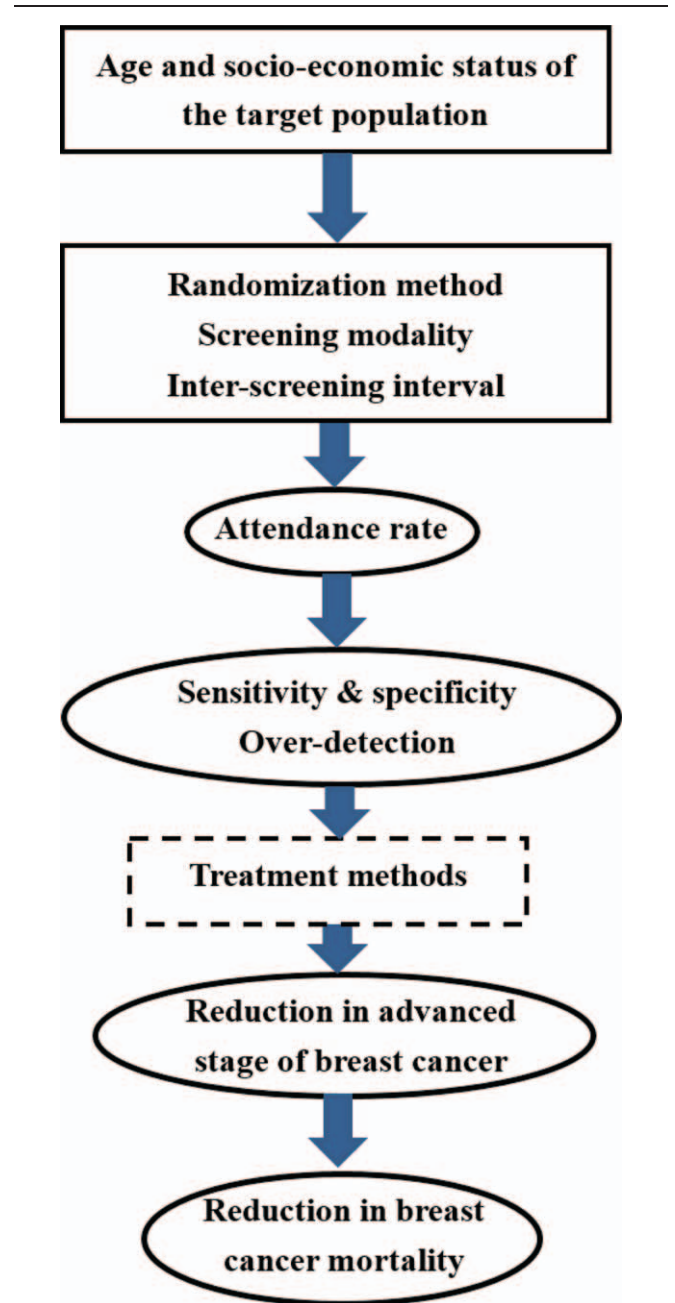
Figure 1. A cascade of components accounting for the heterogeneity across 9 randomized controlled trial.

tabular data on the occurrence of ICs and time since the last negative screen were abstracted from the literature. The expected incidence rate was retrieved from the incidence rate of the control group to calculate the percentage of $[(1 - \text{proportional incidence rate}) \times 100\%]$.

The incidence of IC as a percentage of the expected incidence was computed based on the time since last negative screening (LNS), classified as 0 to 12, 13 to 24, and >24 months. ICs ascertained from the time since last screen within a year represents false-negative cases. It should be noted that IC identified after a year since last negative screen may consist of newly diagnosed incidences of breast cancer rather than only false-negative cases. Therefore, we focused on the comparison of (1-I/E) by 1 year since the last negative screen across trials. A

meta-analysis of sensitivity [1-incidence of interval cancer/ expected incidence (I/E)] was conducted to get a summarized estimate of I/E using a Bayesian Gamma–Poisson conjugated distribution with the presentation in chronological order (see Statistical Methods Section). The meta-analysis of Bayesian Poisson regression fixed-effect model was applied to evaluate the effect of the occurrence of IC on advanced stage breast cancer and breast cancer mortality.

2.4.3. Over-detection.

To obtain the possible range of over-detected cases with follow-up time, we assessed the absolute rate of over-detected breast cancer by subtracting the reduction in the difference of advanced stage breast cancer between the invited and the uninvited group from the difference of overall incidence of nonadvanced breast cancer cases. This is a low estimate that was equivalent to the excess of overall incidence of breast cancer often reported in the literature by comparing the cumulative incidence of breast cancer between the 2 arms. The high estimate was computed by subtracting the reduction in the difference of advanced stage breast cancer from the nonadvanced stage breast cancer in the invited group. The former assumes some non-advanced stage breast cancer cases have not progressed to advanced stage breast cancer, allowing for long lead-time adjustment, until the close of the trial period, whereas the latter assumes all nonadvanced stage breast cancer cases could progress to advanced stage breast cancer in the control group during the trial period. The reason for the proposal of high and low estimates is that the magnitude of over-detection is highly dependent on the follow-up time that determines relative contribution between lead-time related screen-detected breast cancer cases and over-detected cases. The longer the follow-up time after randomization, the less likely it is that the excess of breast cancer can be explained by lead-time and more likely to be influenced by over-detection. Therefore, the average estimates of both scenarios were computed to represent the absolute rate of over-detected breast cancer cases and the percentage of over-detection in the study group in comparison with the control group. The number of screens required for over-detecting (NSO) 1 case of breast cancer was computed by taking the inverse of absolute rate of over-detection. The Bayesian Gamma–Poisson conjugated distribution was applied to the meta-analysis of over-detection from the first HIP trial to the latest UK age trial, with the assumption that these data were exchangeable.

2.5. Indicators for evaluation

1. *Attendance rate*: This was calculated by the number of attendees divided by the number of invited subjects. It is highly dependent on demographic features (e.g., age and gender) and socioeconomic status.
2. *(1-Incidence of interval cancer/Expected Incidence (I/E)) × 100% (sensitivity)*: It has long been proposed to indicate both the test sensitivity to reflect the performance of the screening tool (e.g., mammography) and the adequacy of the inter-screening interval to reflect the progression of breast cancer from the preclinical detectable phase (PCDP) to the clinical phase. This indicates the program sensitivity and may need to be adjusted with the sojourn time distribution using Day method[11] or estimating the sensitivity and the mean sojourn time simultaneously using the Chen method.[12]
3. *Advanced stage breast cancer*: Each trial had reported different tumor attributes, such as tumor size, nodal involvement, and histological differentiation; stage II; or

severe. Here, we used tumor size larger than 2 cm in diameter, node status, or stage II or severe depending on what sort of information was available to represent advanced stage breast cancer. Tabular data on these dichotomous variables (Yes/No) by study arms and detection modes were retrieved from the literature. We used advanced stage breast cancer defined by stage II, severe, node positive, or tumor size larger than 2 cm in diameter to evaluate the effectiveness of using mammography to screen for breast cancer in each trial. This indicator has been also thought of as a surrogate endpoint for breast cancer mortality.

4. *Breast cancer mortality*: This is for evaluation of effectiveness of breast cancer screening with mammography for each trial.
5. *Over-detection*: The over-detection of breast cancer resulting from mammographic screening was also evaluated by comparing the cumulative incidence of breast cancer between the study group and the control group after adjusting for lead-time as indicated above.

3. Statistical methods

To summarize the overall attendance rate based on the trials, we applied a beta-binomial conjugated distribution. A beta-distribution, beta(α_0, β_0), where α_0 and β_0 represent the numbers of participants and nonparticipants, respectively, is often chosen as the prior distribution of a parameter (p), representing the attendance rate. It can be conjugated with the empirical RCT data on the binary outcome to form the likelihood function using the binomial distribution denoted by Bin (n,p), where n is the number of invited women and p is the attendance rate. It can be shown that the posterior distribution of the parameter on the attendance rate forms another similar beta-distribution, beta ($\alpha_0 + y$, $\beta_0 + n - y$), where y is the number of events arising from n invited women, with the exchangeable assumption. The details of this statistical technique are given in the Appendix on statistical methods, Part I, http://links.lww.com/MD/B507.

A Gamma-distribution, Gamma (α_0, β_0), where α_0 and β_0 represent the numbers of ICs and expected incident breast cancers, respectively, is often chosen as the prior distribution of a parameter (γ), representing I/E ratio. It can be conjugated with the empirical data to form the likelihood function using Poisson distribution denoted by Poi(μ), where μ is the mean number of IC, $\mu = rE$. In this study, a noninformative prior, Gamma(1,1), was first chosen. Thus, Bayesian Gamma–Poisson conjugated distribution was used to derive the posterior distribution of approximate sensitivity with (1-I/E) × 100%. It should be noted that the posterior results of the meta-analyses based on the Bayesian conjugated distribution on sensitivity (1-I/E) were presented in chronological order, namely using the resulting posterior as a prior for the model on the next study, assuming these trial data were exchangeable,[13] because the quality of mammographic examination on the enhancement of sensitivity may improve with time from a 1-view to 2-view technique and from a single reading to double reading. The procedure of obtaining the updated posterior distribution in chronological order was repeated to get the final updated posterior distribution of (1-I/E) ratio. The similar Gamma–Poisson conjugated distribution was applied to the Bayesian meta-analysis of over-detection. The details of this statistical technique are given in Appendix on statistical methods, Part II, http://links.lww.com/MD/B507.

The meta-analyses of the impacts of attendance rate and ICs related to sensitivity on advanced stage and death from breast

cancer were modeled by using Bayesian Poisson fixed- and random-effect models under the context of classical Bayesian statistics.[10] The random effect was captured by a normal distribution with zero mean value and sigma of variance, denoted as N (0, σ^2). Bayesian meta-analyses on the relationship between advanced breast cancer and breast cancer mortality based on these 9 trials was also conducted.

We developed a Bayesian causal framework similar to the previous studies used in chronic disease[14,15] by linking together a cascade of causal relationships as a whole from the uptake of screening (attendance rate), the performance of screening (sensitivity) to the yields of reducing advanced breast cancer and breast cancer death. Technically, this Bayesian meta-analysis of causal model was built up by using the Bayesian directed acyclic graphic (DAG) diagram as the backbone. The Markov property was applied to modeling data on repeated rounds of screen. The conditional independence for the parameters and variables implicated in a cascade of causal chains, as often used in the DAG model,[13] was assumed. We used the Bayesian Markov Chain Monte Carlo (MCMC) simulation with Gibbs samples (see Appendix Fig. 1, http://links.lww.com/MD/B507) to estimate the parameters of interest, such as the effectiveness of reducing advanced stage and death from breast cancer attributed to the attendance rate and sensitivity. The details of this statistical technique are given in the Appendix on statistical methods, Part III, and Appendix Table 1, http://links.lww.com/MD/B507. All of the statistical assumptions made for the Bayesian meta-analyses are summarized in the Appendix on statistical methods, Part IV, http://links.lww.com/MD/B507.

4. Results

4.1. Profiles of the included trials

The present study included 9 RCTs, which were conducted between 1963 and 1991.[16–33] Appendix Table 2 http://links.lww.com/MD/B507 shows the basic characteristics of each trial. The first study, the HIP trial, was conducted from 1963 with screening modalities alternating between 2-view mammography and annual physical examinations, with 31,000 women in each arm. Note that breast cancer screening was offered to women with coverage by a health insurance plan, rather than from the community where they resided, as in most other trials. For this reason, selection bias in HIP was different from that in other studies, as the refuser group in the HIP had a lower breast cancer mortality rate than the screened group.[34] The Malmo trial started in 1976 by inviting approximately 40,000 women, aged 45 to 70 years, with equal numbers being randomly assigned to the study control arms. Women assigned to the study group underwent 2-view mammography at intervals of 18 to 24 months. The largest Swedish population- and community-based breast cancer screening originating in 2 counties of Sweden, Kopparberg and Ostergotland (abbreviated as two-county), was launched in 1977 by inviting women, aged 40 to 74 years, to participate 1-view mammography. Subsequently, 3 other Swedish population-based RCTs were conducted in various regions of Sweden between the late 1970s and mid-1980s with similar but slightly different age bands and interscreening intervals. The Edinburg trial invited 45,130 women, aged 45 to 64 years, between 1979 and 1981 in the UK. The unique characteristic of this trial is that the attendance rate varied with age and socioeconomic status. However, such a disparity decreased over subsequent rounds of screens. The Canadian National Breast Screening Study (NBSS) was implemented between 1980 and 1987; it stratified

women by 2 age groups, 40 to 49 years (abbreviated as NBSS1) and 50 to 59 years (NBSS2). It should be noted that recruitment of participants in the NBSS trial included personalized letters of invitation, as well as publicity through advertisements in newspapers, radio, and television. Thus, the conventional definition of attendance rate as number of attendees divided by number of invitees could be different for the NBSS compared to the insurance-based HIP trial or other community-based trials in Europe. This may account for why the attendance rate was high in the NBSS trial. The sensitivity analysis on the influence of the attendance rate was done by including and excluding the NBSS data. The recent RCT in the UK was focused on young women who, at 39 years of age, commenced having annual mammograms (study group) and were compared with women, aged 50 to 69 years, who received triennial mammography. The age to begin mammography was from 40 years onwards for all the studies, with the exception of the Malmo and Edinburgh trials; the age for commencing mammography for those studies was 45. The age at termination of screening varied among the trials: 64 years for HIP, Stockholm, and Edinburgh trial, 59 years for NBSS and Gothenburg, 74 years for the two-county trials and 69 years for Malmo. The interscreening interval ranged from 12 to 33 months.

Appendix Table 2 http://links.lww.com/MD/B507 shows the heterogeneity across the trials with respect to age range, interscreening interval, sample size, screening modalities, and the ages of initiation.

4.2. Bayesian meta-analysis of the impact of attendance rate

It is evident that attendance rate was heterogeneous across trials (see the last column of Appendix Table 2, http://links.lww.com/MD/B507). Generally speaking, the attendance rate in the 5 Swedish trials was higher than that in places outside of Sweden, except the Canadian trial that combined various invitation methods as indicated above. Using Bayesian beta-binomial conjugated analysis, the overall attendance rate during the period from 1965 for the HIP trial to 1991 for the UK age trial was found to be approximately 78.70% (95% CI: 78.60–78.90%).

Table 1 shows the relationship of the attendance rate to breast cancer mortality and advanced stage breast cancer in univariate and multivariate analyses with fixed- and random-effect Bayesian models, respectively. The inverse relationships between the attendance rate and advanced stage and death from breast cancer were noted in both univariate and multivariate analyses. The results of the latter on breast cancer mortality (regression coefficients = −0.0259; relative risk (RR) = 0.97 (95% confidence interval (CI): 0.968–0.981)) and advanced stage of breast cancer (regression coefficients = −0.0321; RR = 0.968 (95% CI: 0.963–0.974)), after adjustment for age, calendar year, interscreening interval and the logarithm of the number of ICs, suggest that an increase in 1% of the attendance rate led to an approximately statistically significant 3% reduction in advanced stage and death from breast cancer.

Regarding the Bayesian random-effect model that allows for the variation of attendance rate across trials, the result was statistically significant for the estimate of heterogeneity (sigma, σ), which indicated the heterogeneity of the relationship between the attendance rate and the rate of advanced stage breast cancer and breast cancer mortality. After adjusting for such a heterogeneity, the effects of the attendance rate on both advanced stage breast cancer (RR = 0.976, 95% CI: 0.94–1.02) and death

Table 1

Effect of attendance rate on the breast cancer death and advanced stage of breast cancer using Bayesian Poisson fixed- and random-effect regression model.

Model	Variables	Estimate	SE	RR	95% CI	
Fixed-effect model						
	Breast cancer death					
	Univariate					
	Attendance rate (%)	−0.0188	0.0026	0.981	0.976	0.986
	Multivariate					
	Attendance rate (%)	−0.0259	0.0034	0.974	0.968	0.981
	Log (IC)	0.4314	0.0615	1.539	1.365	1.736
	Year	−0.0137	0.0049	0.986	0.977	0.996
	Age	0.0384	0.0076	1.039	1.024	1.055
	Interscreening interval	0.3905	0.0690	1.478	1.291	1.695
	Advanced stage of breast cancer					
	Univariate					
	Attendance rate (%)	−0.0142	0.0021	0.986	0.982	0.990
	Multivariate					
	Attendance rate (%)	−0.0321	0.0031	0.968	0.963	0.974
	Log (IC)	0.3637	0.0513	1.439	1.301	1.592
	Year	0.0026	0.0046	1.003	0.994	1.012
	Age	0.0418	0.0066	1.043	1.029	1.056
	Interscreening interval	0.6562	0.0611	1.927	1.712	2.175
Random-effect model						
	Breast cancer death					
	Univariate					
	Attendance rate (%)	−0.0253	0.0182	0.975	0.940	1.010
	Sigma (σ)	0.6189	0.1736	—	0.382	1.050
	Multivariate					
	Attendance rate (%)	−0.0279	0.0199	0.973	0.932	1.012
	Log (IC)	0.3784	0.3763	1.460	0.665	3.022
	Year	−0.0176	0.0365	0.983	0.913	1.057
	Age	0.0251	0.0436	1.025	0.940	1.119
	Interscreening interval	0.3175	0.4520	1.374	0.562	3.490
	Sigma (σ)	0.6385	0.2497	—	0.331	1.284
	Advanced stage of breast cancer					
	Univariate					
	Attendance rate (%)	−0.0243	0.0213	0.976	0.936	1.017
	Sigma (σ)	0.7167	0.2016	—	0.440	1.219
	Multivariate					
	Attendance rate (%)	−0.0337	0.0201	0.967	0.927	1.006
	Log (IC)	0.5002	0.3357	1.649	0.843	3.294
	Year	−0.0057	0.0320	0.994	0.932	1.061
	Age	0.0350	0.0389	1.036	0.958	1.121
	Interscreening interval	0.6696	0.3954	1.953	0.892	4.402
	Sigma (σ)	0.5470	0.2352	—	0.272	1.151

CI = confidence interval, IC = interval cancer, RR = relative risk, SE = standard error.

from breast cancer (RR = 0.975, 95% CI: 0.94–1.01) were not statistically significant, but the effect size in the reduction of advanced stage and death from breast cancer remained 2.5%. The similar findings were noted in multivariable regression analysis.

It is very interesting to note that the results excluding the CNBSS trial provided similar findings on the reduction in breast cancer mortality, but showed a statistically significantly larger benefit of reducing advanced stage breast cancer (11%) (Appendix Table 3, http://links.lww.com/MD/B507).

4.3. Bayesian meta-analysis of sensitivity in chronological order

Table 2 lists the estimated I/E of 1 year since the last negative screen for each trial, yielding the estimated sensitivity, ranging from 57.26% (95% CI: 45.14–66.70%) of the UK age trial (age 40–41) to 87.97% (84.30–90.77%) of the two-county trial (age 40–74), indicating the heterogeneity of sensitivity. A lower sensitivity in the UK age trial was primarily due to the enrollment of the youngest women (age 40–41 years) rather than 10-year age band (aged 40–49 years). In addition to low sensitivity for young women in the UK age trial, the young women aged 50 years or below had poorer sensitivity (63.57% (95% CI: 50.11–73.40%)) compared with the older women aged 50 years or older (74.01% (95% CI: 64.05–82.21%)) in the CNBSS trial. Similar findings were noted for the Gothenburg trial with the same 2 age bands, but the absolute estimates were higher than those found in the CNBSS trial. It is very interesting to note that the young women in the Gothenburg trial had better sensitivity (82.28% (95% CI: 69.85–89.59%)), albeit still lower than the older women (87.68% (95% CI: 77.56–93.24%)), compared with those young women in other trials. It should be noted that because age ranges were different between the trials, the comparison

Table 2

Results of the ratio of interval cancer to the expected incidence cases in 9 randomized controlled trials on mammography screening.

Trial	Year	Age group	Interscreening interval, m	Interval cancers since LNS	Expected incident cases since LNS	Sen (=1-I/E) (95% CI) [Using interval cancer within 1 y since LNS]		Posterior* Sensitivity (=1-I/E)	95% CI
HIP	1963	40–64	12	41	112.85	63.67%	(48.06%, 74.59%)	63.67%	(51.73%, 73.93%)
Malmo	1976	45–69	18–24	34	131.16	74.08%	(62.20%, 82.22%)	69.26%	(61.93%, 75.82%)
Two-county	1977	40–74	24–33	61	506.87	87.97%	(84.30%, 90.77%)	81.89%	(78.72%, 84.80%)
Edinburg	1979	45–64	12	14	89.01	84.27%	(72.37%, 91.05%)	82.14%	(79.17%, 84.88%)
CNBSS-1	1980	40–49	12	53	145.49	63.57%	(50.11%, 73.40%)	79.40%	(76.47%, 82.14%)
CNBSS-2	1980	50–59	12	46	176.97	74.01%	(64.05%, 81.21%)	78.58%	(75.84%, 81.16%)
Stockholm	1981	40–64	24–28	22	105.10	79.07%	(66.86%, 86.78%)	78.62%	(76.00%, 81.09%)
Gothenburg	1982	39–49	18	16	90.32	82.28%	(69.85%, 89.59%)	78.86%	(76.35%, 81.24%)
Gothenburg	1982	50–59	18	12	97.40	87.68%	(77.56%, 93.24%)	79.45%	(77.06%, 81.72%)
UK trial	1991	40–41	12	88	205.90	57.26%	(45.14%, 66.70%)	76.70%	(74.32%, 78.97%)

CI = confidence interval, HIP = Health Insurance Plan, I/E = incidence of interval cancer/expected incidence, LNS = last negative screening.
* The prior distribution for I/E was Gamma(0.001, 0.001), and updated as posterior distribution with observed incidence cases and expected incident cases one by one in chronological order.

should be made with great caution. Nonetheless, it is still obvious that the sensitivity (74.01% (95% CI: 64.05–82.21%)) for the older women in the Canadian trial was lower than the other trials with the same age band. The sensitivity estimates were also poorer for the 2 earlier studies, the Malmo trial (74.08%) and the HIP trial (63.67%), than other trials with the similar age band.

The summarized estimated sensitivity after a meta-analysis using Gamma–Poisson conjugated distribution was 79.45% (95% CI: 77.06–81.72%) and 76.7% (95% CI: 74.32–78.97%) from 1963 to 1991 excluding and including the UK age trial, respectively.

Table 2 also presents the posterior results of sensitivity in chronological order. The sensitivity improved from 63.6% in 1963 during the HIP trial, 69.3% in 1976 during Malmo trial, to 82% around 1980 and became stable between 77% and 79% since then. Note that a lower sensitivity was noted in 1990 because of the UK age trial that was targeted at young women.

4.4. Bayesian meta-analysis of the impact of sensitivity on advanced stage of breast cancer

Appendix Figure 2 http://links.lww.com/MD/B507 shows scatter plots between (1-I/E) and the relative rate of mortality from and advanced stage of breast cancer, both indicating the higher the sensitivity, the lower the relative rates of both mortality and advanced stage breast cancer (i.e., a negative association).

Using Bayesian meta-analysis based on Poisson regression model to model the effect of the logarithm of ICs on the number of breast cancer deaths. Table 3 shows an increase in the logarithm of 1 IC led to a 26% (RR = 1.26, 95% CI: 1.14–1.39) and 55% (RR = 1.55, 95% CI: 1.37–1.75) elevated risk for dying from breast cancer in the univariate analysis and multivariable model with adjustment for calendar year (RR = 0.98, 95% CI: 0.967–0.986). There was a statistically significant decreasing trend, being a higher risk with advancing age (RR = 1.048, 95% CI: 1.034–1.062), and the predetermined interscreening interval (RR = 1.171, 95% CI: 1.049–1.307) being an elevated risk with a longer interscreening interval. Similar findings were noted when the counts of ICs taking the logarithm of 3 years since LNS were modeled.

The corresponding results on the impact of IC on advanced stage breast cancer are also presented in Table 3. Risk was significantly elevated by 18% (RR = 1.18, 95% CI: 1.09–1.28) per 1 IC taking the logarithm function in the univariate analysis

and by 48% (RR = 1.48, 95% CI: 1.34–1.63) in multivariable analysis while calendar year (RR = 0.99, 95% CI: 0.984–1.002), age (RR = 1.06, 95% CI: 1.05–1.07), and interscreening interval (RR = 1.41, 95% CI: 1.29–1.55) were controlled. Age and interscreening interval, but not calendar year, were statistically significantly associated with the risk for advanced stage breast cancer. Similar results were seen while the counts of IC taking the logarithm of 3 years since last negative screen were considered.

4.5. Bayesian meta-analysis of over-detection

Table 4 estimates the absolute rate and the percentage (RR between the 2 groups) of over-detected breast cancers resulting from mammographic screening for each trial. There are low and high estimates, depending on which definition of lead-time in relation to follow-up time is adopted. The low end is obtained if the excess cases due to lead-time have not been washed out by the end of follow-up, and the upper end is obtained if the excess cases due to lead-time have been washed out by the end of follow-up. Taking the average of the 2 estimates, the estimated absolute over-detected breast cancer rate was lowest (22 per 10^5) in the Gothenburg trial for women aged 39 to 49 years, and highest (167 per 10^5) in the Canadian trial. The range of the estimated absolute rate for each trial suggested a lack of significant over-detection in the HIP trial, the two-county trial, the Stockholm trial, and the Gothenburg trial for young women. Similar findings were noted for the percentage of over-detection being the highest in the Canadian trial for women aged 40 to 49 years (61.0%) and the lowest (12.5%) in the Gothenburg trial. The number of screens required for over-detecting (NSO) 1 breast cancer ranged from 597 in the Canadian trial for women aged 50 to 59 years to 4482 in the Gothenburg trial for women aged 39 to 49 years (Table 4). It should be noted that the percentage of low estimates of over-detection reported here are very similar to that of over-detected cases reported with the difference of cumulative incidence between the 2 arms.

4.6. Meta-analysis of advanced breast cancer and death from breast cancer

Table 5 shows the projection of advanced stage breast cancer with the application of attendance rate, proportional incidence rate of IC, and detection mode (prevalent screen, subsequent screen, IC, and refuser), and stage distribution by detection mode.

Table 3

Bayesian meta-analysis using Poisson regression model on the effects of interval cancer, time since last negative screening, conducting year, and age groups on breast cancer death/advanced stage of breast cancer in the population-based screening program of mammography.

Variables	Estimate	SE	RR	95% CI	
Breast cancer death					
(A) IC within 1-y					
Univariate					
Log (IC)	0.2317	0.0503	1.261	1.142	1.392
Multivariate					
Log (IC)	0.4370	0.0614	1.548	1.373	1.746
Year	−0.0237	0.0049	0.977	0.967	0.986
Age	0.0471	0.0069	1.048	1.034	1.062
Interscreening interval	0.1578	0.0560	1.171	1.049	1.307
(B) IC within 3-y					
Univariate					
Log (IC)	0.4587	0.0380	1.582	1.468	1.704
Time since LNS	0.2651	0.0324	1.304	1.224	1.389
Multivariate					
Log (IC)	0.4941	0.0443	1.639	1.503	1.787
Time since LNS	0.0347	0.0360	1.035	0.965	1.110
Year	−0.0262	0.0038	0.974	0.967	0.981
Age	0.0679	0.0053	1.070	1.059	1.082
Interscreening interval	−0.0608	0.0459	0.941	0.860	1.029
Advanced stage of breast cancer					
(A) IC within 1-y					
Univariate					
Log (IC)	0.1624	0.0414	1.176	1.085	1.276
Multivariate					
Log (IC)	0.3887	0.0507	1.475	1.335	1.629
Year	−0.0070	0.0046	0.993	0.984	1.002
Age	0.0550	0.0058	1.057	1.045	1.069
Interscreening interval	0.3455	0.0461	1.413	1.292	1.547
(B) IC within 3-y					
Univariate					
Log (IC)	0.4022	0.0311	1.495	1.407	1.589
Time since LNS	0.3087	0.0266	1.362	1.293	1.434
Multivariate					
Log (IC)	0.3193	0.0361	1.376	1.282	1.477
Time since LNS	0.0197	0.0288	1.020	0.964	1.079
Year	−0.0057	0.0035	0.994	0.987	1.001
Age	0.0744	0.0044	1.077	1.068	1.087
Interscreening interval	0.1343	0.0376	1.144	1.063	1.231

CI = confidence interval, IC = interval cancer, LNS = last negative screening, RR = relative risk, SE = standard error.

Table 4

The estimated results of over-detection and number needed to screen for one over-detected case in the population-based screening for breast cancer with mammography.

Trials	Women-years		Invasive breast cancer cases					Absolute rate of over-detection (per 10^3)			NSO			Percentage of over-detection		
	Study	Control	Study	Control	Study, adv	Control, adv	(Adv breast cancer)	Low	High	Average	Low[*]	High	Average	Low (%)	High (%)	Average (%)
HIP	179,472	180,816	334	352	162	200	Stage 2+	−0.09	0.76	0.34	0	1323	2983	1.4	39.1	17.5
Malmo	185,983	186,674	486	396	190	231	Stage 2+	0.49	1.38	0.93	727	2036	1072	23.5	65.1	44.3
Two-county	652,706	476,864	1303	996	524	555	Stage 2+	−0.09	0.83	0.37	0	1201	2703	0.3	40.0	17.8
Edinburg	157,946	147,854	355	261	228	221	Stage 2+	0.48	0.75	0.62	1331	2079	1623	27.7	43.1	35.4
CNBSS-1	124,621	124,943	286	232	96	63	Nodes +	0.44	1.79	1.11	558	2284	898	24.1	97.0	60.5
CNBSS-2	96,626	97,061	341	274	92	86	Nodes +	0.71	2.64	1.68	379	1415	597	25.5	94.1	59.8
Stockholm	201,590	99,715	385	203	172	97	Stage 2+	−0.13	0.94	0.41	0	1068	2468	1.2	46.5	20.4
Gothenburg, 39–49	81,750	99,335	124	184	39	73	Nodes +	−0.33	0.78	0.22	0	1280	4482	0.2	42.6	12.5
Gothenburg, 40–59	49,564	78,369	147	231	46	71	Nodes +	0.02	2.06	1.04	486	51813	962	4.8	70.3	35.7
Age trial	312,957	622,127	409	755	124	276	Nodes +	0.09	0.86	0.48	1158	10595	2087	8.3	71.3	39.6
Overall								0.19	1.21	0.70				9.9	62.4	

HIP = Health Insurance Plan, NSO = number of screenee required for over-detecting.

[*] The low estimate of NSO is truncated to 0 while the absolute rate is negative.

Table 5

The estimated relative risk of advanced stage of breast cancer for the screen arm compared to the control arm, and the projected relative risk of breast cancer death for the 7 randomized trials on mammography screening and for the sensitivity analysis of various scenarios.

Study	Age groups	RR (adv breast cancer)	Projected RR (breast cancer death)
Empirical trials			
HIP		0.89 (0.68, 1.15)	0.82 (0.67, 0.97)
Malmo		0.81 (0.66, 0.98)	0.76 (0.65, 0.88)
Swedish two-county	40–49	0.92 (0.68, 1.22)	0.83 (0.67, 1.02)
	50–59	0.60 (0.49, 0.73)	0.62 (0.52, 0.73)
	60–69	0.66 (0.55, 0.78)	0.66 (0.56, 0.76)
	70–74	0.94 (0.65, 1.32)	0.84 (0.65, 1.07)
	40–74	0.69 (0.61, 0.78)	0.69 (0.61, 0.77)
Edinburg		0.82 (0.68, 0.98)	0.77 (0.67, 0.87)
NBSS	40–49	1.38 (0.84, 2.19)	1.09 (0.79, 1.53)
	50–69	1.13 (0.68, 1.79)	0.95 (0.67, 1.32)
	40–59	1.22 (0.86, 1.69)	1.01 (0.79, 1.28)
Stockholm		0.77 (0.59, 1.00)	0.74 (0.60, 0.88)
Gothenburg	39–59	0.65 (0.52, 0.81)	0.66 (0.54, 0.78)
Sensitivity analysis of various scenarios [attendance rate, Sen]			
90%, 95%		0.67 (0.58, 0.76)	0.67 (0.58, 0.76)
60%, 95%		0.78 (0.70, 0.86)	0.74 (0.67, 0.81)
30%, 95%		0.89 (0.82, 0.96)	0.81 (0.76, 0.87)
90%, 75%		0.79 (0.66, 0.93)	0.75 (0.65, 0.84)
60%, 75%		0.86 (0.76, 0.96)	0.79 (0.72, 0.87)
30%, 75%		0.93 (0.85, 1.01)	0.84 (0.78, 0.90)
90%, 55%		0.93 (0.76, 1.13)	0.84 (0.72, 0.96)
60%, 55%		0.95 (0.83, 1.10)	0.85 (0.77, 0.95)
30%, 55%		0.98 (0.89, 1.08)	0.87 (0.80, 0.94)

HIP = Health Insurance Plan, NBSS = National Breast Screening Study, RR = relative risk.

The predicted relative rate of advanced stage breast cancer for the invited group versus the uninvited group was the lowest (highest efficacy) in the two-county trial (RR = 0.69; 95% CI: 0.61–0.78) and the highest (lowest efficacy) in the NBSS trial (40–59 years) (RR = 1.22; 95% CI: 0.86–1.69).

Based on the relationship between the logarithm of RR of advanced stage breast cancer (X_{adv} = log(rate of advanced stage breast cancer)) and the logarithm of RR being death from breast cancer (Y_{mort} = log(mortality rate of breast cancer)): Y_{mort} = −0.1261 + 0.6783X_{adv} obtained from a previous study,[35] the projected mortality rates attributed to breast cancer given the rate of advanced stage breast cancer are also listed in Table 5, being the lowest (good efficacy) in the two-county trial (RR = 0.69, 95% CI: 0.61–0.77) and the highest (poor efficacy) in the Canadian trial (NBSS 1 + NBSS 2) (RR = 1.01, 95% CI: 0.79–1.28).

4.7. Bayesian meta-analysis of causal model

The Bayesian causal model can be used to provide an insight into how the 2 key components, attendance rate and sensitivity, affect the rate of advanced stage breast cancer and breast cancer mortality based on the empirical findings using data from the 9 RCTs of Western countries. Table 5 also shows the effectiveness of mammographic screening in 3 scenarios, low, medium, and high attendance and sensitivity based on the average estimate of the data from the 9 trials.

Three scenarios of low, medium, and high sensitivity groups gave the estimated effectiveness on the RR of the rate of advanced stage breast cancer and breast cancer death for 9 scenarios with the attendance rate in combination with sensitivity. The best scenario (90% attendance rate and 95% sensitivity) yielded 0.67 (95% CI: 0.58–0.76) and 0.67 (95% CI: 0.58–0.76) of RR for

being the rate of advanced stage breast cancer and breast cancer mortality, indicating a 33% (95% CI: 20–43%) reduction in both advanced stage breast cancer and breast cancer mortality, whereas the poor scenario (30% attendance rate and 55% sensitivity) gave 0.98 (95% CI: 0.83–1.16) and 0.87 (95% CI: 0.77–0.98), indicating only 2% reduction in advanced stage breast cancer rate and 13% reduction in breast cancer mortality.

5. Discussion

The present study revisited the literature on 9 RCTs and reanalyzed tabular data with systematic approaches using Bayesian meta-analyses and causal models to assess the impact of the attendance rate and quality assurance indicators, proportional incidence rate and over-detection on the outcomes of interest, advanced stage breast cancer and breast cancer mortality. Evaluation of these RCTs in a systematic way is not only very helpful for assessing whether each screening program works, but also provides insight into how and why some screening trials were effective but others were not. The former is often answered by systematic review and meta-analyses of all published trials while making allowances for heterogeneity, but the latter cannot be completely solved by it. Better understanding of the latter aspect is necessary for countries worldwide that are seeking to learn from the experience in Western countries and develop a plan for mass screening in each country. Heterogeneity with respect to age range, interscreening interval, sample size, screening modalities, and the ages of initiation, as shown in Appendix Table 2, http://links.lww.com/MD/B507, prompted us to do a systematic evaluation of how the underlying factors related to heterogeneity contributes to the reduction in advanced stage breast cancer and breast cancer mortality. Therefore, our systematic review is different from previous systematic

meta-analyses. We used a systematic framework analysis to explore how 3 key components (attendance rate, sensitivity, over-detection) account for heterogeneity across trials and how they affect the result of breast cancer stage and mortality in order to clarify the recent debate on the use of mammography screening as the preferred screening tool for population-based breast cancer screening.

5.1. Heterogeneity of breast cancer screening with mammography

5.1.1. Attendance rate. Based on the analyses in this step-by-step systematic evaluation, we have a deeper understanding of the underlying mechanism leading to the primary outcome of breast cancer mortality. We start from the key component of attendance rate. It can be shown that the higher the attendance rate, the lower the mortality, which implies the poor mortality rate is caused by the lower attendance rate. In the RCT design, this is related to self-selection bias, which is often solved by intention to treat analysis. Individually, the attendance rate was lower for the Malmo trial among the 5 Swedish RCTs and the earliest HIP trial had lower attendance.

The HIP had a lower attendance rate but risk in the nonparticipants was lower than in participants.[34] The Malmo trial had the lowest (74%) attendance rate, which may account for the lower mortality rate when compared to other Swedish trials. The Canadian trial had a higher attendance rate but the invitation method was different from that adopted by European trials. Moreover, there was a problem of control contamination[36] that would further complicate the impact of the attendance rate. After excluding this trial, the results on the relationship between the attendance rate and the outcomes of advanced stage breast cancer and breast cancer mortality were similar.

Although the majority of attendance rates were higher than 70%, and the pooled estimate obtained from Bayesian conjugated analysis reached up to 79%, the heterogeneity across trials with respect to attendance rate still exists.

5.1.2. Sensitivity. After studying the attendance rate, we then focused on an indicator (1-I/E) (i.e., proportional incidence of IC) to reflect the quality of the performance of mammography. We performed Bayesian meta-analysis with conjugated Gamma–Poisson distribution in chronological order, which used the resulting posterior distribution based on earlier conducted trials as the prior distribution for the likelihood function based on next trial. Such an approach has been widely used in the synthesis of data on RCTs in a chronological order as proposed by Spiegelhalter et al.[13] The reason for doing so was because the quality assurance of mammography for the next trial would learn from those of the current trial. Therefore, the posterior distribution would be updated with time to reflect the dynamic sensitivity of mammography with time. Compared to conventional Bayesian meta-analysis formed by one prior (noninformative prior in this case before these trials) and one likelihood (including all of the studies). Conducting the meta-analysis with Bayesian conjugated distribution in this way may be better suited for the integration of those data on RCTs conducted between the late 1970s and 1990s, particularly for the parameter of sensitivity. Although the overall effect size would be the same because of the exchangeable assumption that renders all the trial data be conjugated with the same Gamma distribution between prior and posterior distribution. This exchangeable assumption may not be unreasonable, as the screening tool used was based on the same mammography worldwide during the trial period.

Presenting the effect sizes in chronological order is therefore more informative than presenting the effect size of combing data from all studies with one prior (noninformative prior) and one likelihood. Table 2 clearly shows this merit. The sensitivity improved from 63.6% in 1963 to 82% around 1980 and stabilized since then.

We demonstrated that the higher the (1-I/E), the lower the rate of advanced stage breast cancer and the breast cancer mortality rate. It is plausible that a partial reduction in the rate of advanced stage and mortality is highly dependent on the quantitative proportion of this indicator, as supported by the Bayesian meta-analysis of the Poisson regression model that quantifies the effect of this indicator on the rate of advanced stage breast cancer. It is very interesting to note that there is substantial heterogeneity of this indicator across trials. The indicator of (1-I/E) is not only affected by test sensitivity, but also the interscreening interval. To eliminate the possibility of the incidence of newly diagnosed breast cancer after the time since the negative screen, this indicator was limited to 1 year from the negative screen to reflect test sensitivity only. The alternative method was to use the modeling approach to relax this assumption to consider the sensitivity and mean sojourn time (capturing the newly diagnosed breast cancer) as done by Chen et al.[12] The application of the Markov process to the Swedish two-county trial and the Canadian trial have been reported previously.[12,37] The estimated mean sojourn time for women aged 40 to 49 and 50 to 69 were approximately 2.5 and 4 years, respectively. The corresponding figures for women aged 40 to 49 and 50 to 59 in the Canadian trial were 2.5 and 3 years, respectively. The simultaneous estimation for sensitivity gave estimates for young and older women of 85% and 100% in the Swedish two-county trial, and 61% and 75% in the Canadian trial. The disparity in the parameters pertaining to the disease's natural history between the 2 studies was small because the estimates of the mean sojourn time were identical between the 2 studies, but the sensitivity estimates in the Canadian trial were considerably lower than those in the Swedish two-county trial. These confirmed the finding that our (1-I/E) estimate in the Canadian trial was lower than that of the Swedish two-county trial.

5.1.3. Over-detection. We found that the proportion of over-detection is highly dependent on the follow-up time and the lead time gained for early detection. The shorter the follow-up time the more likely to mix up both lead-time related early detected breast cancer and over-detected breast cancer. Our high estimate was based on the premise that all invasive breast cancer would progress from nonadvanced to advanced stage breast cancer during the follow-up time in the absence of screening, whereas the low estimate was based on the premise that not all invasive breast cancer would progress from nonadvanced to advanced stage breast cancer during the follow-up time. It is reasoned that the high estimate can be regarded as the upper limit of over-detected breast cancer but not the base-case estimate of the proportion of over-detection. As mentioned above, the low estimate is actually the often reported over-detected estimate based on the comparison of cumulative incidence of total breast cancer between the 2 arms. Reporting this estimate in this way may not consider the lead-time of nonadvanced breast cancer (long-lead-time) during long-term follow-up after the trial.

5.2. Implications for health policy of breast cancer screening in countries worldwide

There are several main estimates accrued from the meta-analyses of all trials for developing the standard guidelines for Asian

countries. It is recommended that screening be conducted in a manner similar to the era of these RCTs. The attendance rate should not be lower than 70% if the age group is between 40 and 69 years. The 1-proportional IC rate should not be lower than 75% to guarantee a lower IC rate representing good sensitivity and the optimal interscreening interval. The over-detection percentage compared with the incidence rate of the control group taking the average of low and high estimate may not be over 20%.

The parameters estimated from the 9 trials can be used for doing empirical simulation of different scenarios given the combination of the attendance rate with sensitivity. The high attendance rate and the higher the sensitivity, the more likely that the effectiveness of breast cancer screening can be achieved. Simulating the effectiveness of mammographic screening by the combination of the attendance rate and the sensitivity would provide a quantitative assessment of how the attendance rate and sensitivity affect the outcomes of advanced stage breast cancer and death from breast cancer.

The limitations of this review and subsequent analyses are several-fold. There was lack of information on contamination of the control group, which is one of factors accounting for the disparity related to the harm of screening across trials. One typical example is the Canadian trial, which has been reported to have a higher proportion of control contamination, 26% in 40 to 49[38] and 17% in 50 to 59.[36]

The second concern is the follow-up time to estimate the proportion of over-detected breast cancers. It should be noted that the greater the effort made to screen the target population with mammography, the greater the number of small (<1 cm) and node negative breast cancers detected, which requires longer follow-up time to differentiate small, screen-detected breast cancers from the over-detected breast cancers, as these small and node negative breast cancers are supposed to have a longer lead-time to progress to advanced breast cancer in the absence of screening. A short follow-up time is apt to mix screen-detected "lethal" breast cancer with long lead time, with theoretically nonprogressive, over-detected breast cancer. Second, it may be argued that the advent of new diagnostic tools and therapies also make contributions to the outcomes of advanced stage breast cancer and breast cancer mortality. However, these factors were unlikely to affect our evaluation; this was due, in part, to the RCT design and the lack of a new target therapy in the era of RCT. Nonetheless, this also means our results cannot be applied to the era of breast cancer service screening programs, which has a series of emerging adjuvant therapies and new diagnostic techniques, such as alternative imaging technique (magnetic resonance image). Our approach is, therefore, different in that it is entirely based on a microsimulation approach to assess how different screening policies and other factors affect the disease progression and prognosis, such as the Cancer Intervention and Surveillance Modelling Network (CISNET) initiative.

Although the controversy about the justification of population-based breast cancer screening with mammography has continued in academic and nonacademic settings in Western countries, we hope the circulation of such information will not affect health policy-makers who get involved with population-based breast cancer screening because tumor staging in the countries where these trials were located are completely different from some countries that have not been exposed to population-based screening, such as Asian countries. Compared with what has taken place in other countries, breast cancer in these countries has been largely down staged after the introduction of mammogra-

phy. Therefore, evaluation of the effectiveness of population-based mammographic breast cancer service screening in Western countries is complicated. It requires a longer follow-up period to distinguish early-detected cases from over-detected ones to yield an unbiased estimate of effectiveness, and is unlikely to be cost-effective compared with the era before the widespread use of mammography. In contrast, evaluating the scenario in other countries that are still at the beginning of mammography screening would be different. Therefore, to elucidate causes accounting for the heterogeneity across trials would be very helpful for the down staging of breast cancer while population-based breast cancer screening is implemented in other countries. It should also be noted that the trials analyzed in the present study are all based on European-ancestry individuals. The etiology of breast cancer in non-European ancestry individuals is likely to be different. The external application to other ethnic groups, such as Asian people, may be limited but the results of heterogeneity from these trials are likely informative for health policy decision-makers and clinicians.

In conclusion, systematic evaluation of the empirical data from the 9 RCTs on breast cancer screening with mammography revealed that the heterogeneity of the primary outcome in the reduction of the rate of advanced stage breast cancer and breast cancer mortality was well explained by the attendance rate, the performance of mammography screening, and over-detection. Such heterogeneous findings are informative to aid countries in launching high quality and effective population-based screening for breast cancer with mammography in order to downstage breast tumors, which, in turn, saves women lives. We stress here that population-based breast cancer screening with mammography should only be adopted conditionally on the premise that the quality of these indicators can be achieved.

References

[1] Nyström L, Rutqvist LE, Wall S, et al. Breast cancer screening with mammography: overview of Swedish randomised trials. Lancet 1993; 341:973–8.

[2] Independent UK Panel on Breast Cancer ScreeningThe benefits and harms of breast cancer screening: an independent review. Lancet 2012; 380:1778–86.

[3] Smith RA, Duffy SW, Gabe R, et al. The randomized trials of breast cancer screening: what have we learned? Radiol Clin North Am 2004;42:793–806.

[4] Gøtzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? Lancet 2000;355:129–34.

[5] Tabár L, Vitak B, Chen TH, et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology 2011;260:658–63.

[6] Miller AB, Wall C, Baines CJ, et al. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. BMJ 2014;348:g366.

[7] Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6:e1000097.

[8] Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996;17:1–12.

[9] Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340: c332.

[10] Congdon P. Bayesian Statistical Modelling. John Wiley & Sons Ltd, West Sussex:2006.

[11] Day NE. Estimating the sensitivity of a screening test. J Epidemiol Community Health 1985;39:364–6.

[12] Chen HH, Duffy SW, Tabár L. A Markov chain method to estimate the tumour progression rate from preclinical to clinical phase, sensitivity and positive predictive value for mammography in breast cancer screening. Statistician 1996;45:307–17.

[13] Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomised trials. J R Stat Soc A 1994;157:357–416.

[14] Eddy DM, Schlessinger L. Archimedes: a trial-validated model of diabetes. Diabetes Care 2003;26:3093–101.

[15] Ness RB, Koopman JS, Roberts MS. Causal system modeling in chronic disease epidemiology: a proposal. Ann Epidemiol 2007;17:564–8.

[16] Shapiro S, Venet W, Strax P, et al. Ten- to fourteen-year effect of screening on breast cancer mortality. J Natl Cancer Inst 1982;69:349–55.

[17] Shapiro S, Goldberg JD, Hutchison GB. Lead time in breast cancer detection and implications for periodicity of screening. Am J Epidemiol 1974;100:357–66.

[18] Shapiro S. Periodic screening for breast cancer: the HIP randomized controlled trial. Monogr Natl Cancer Inst 1997;22:27–30.

[19] Chu KC, Smart CR, Tarone RE. Analysis of breast cancer mortality and stage distribution by age for the Health Insurance Plan Clinical Trial. J Natl Cancer Inst 1988;80:1125–32.

[20] Andersson I, Aspegren K, Janzon L, et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. BMJ 1988;297:943–8.

[21] Ikeda DM, Andersson I, Wattsgard C, et al. Interval cancer in the Malmo mammographic screening trial: radiographic appearance and prognostic considerations. Am J Roentgenol 1992;159:287–94.

[22] Tabár L, Fagerberg CJ, Gad A, et al. Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. Lancet 1985;1:829–32.

[23] Roberts MM, Alexander FE, Anderson TJ, et al. Edinburgh trial of screening for breast cancer: mortality at seven years. Lancet 1990;335: 241–6.

[24] Alexander FE, Anderson TJ, Brown HK, et al. 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. Lancet 1999;353:1903–8.

[25] Miller AB, Baines CJ, To T, et al. Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. CMAJ 1992;147:1459–76.

[26] Miller AB, To T, Baines CJ, et al. The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up. A randomized screening trial of mammography in women age 40 to 49 years. Ann Intern Med 2002;137(5 Part 1):305–12.

[27] Miller AB, Baines CJ, To T, et al. Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. CMAJ 1992;147:1477–88.

[28] Miller AB, To T, Baines CJ, et al. Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50–59 years. J Natl Cancer Inst 2000;92:1490–9.

[29] Frisell J, Lidbrink E, Hellstrom L, et al. Follow up after 11 years—update of mortality results in the Stockholm mammographic screening trial. Breast Cancer Res Treat 1997;45:263–70.

[30] Frisell J, Hellstrom L, et al. Analysis of interval breast carcinomas in a randomized screening trial in Stockholm. Breast Cancer Res Treat 1987;9:219–25.

[31] Bjurstam N, Björneld L, Warwick J, et al. The Gothenburg Breast Screening Trial. Cancer 2003;97:2387–96.

[32] Bjurstam N, Björneld L, Duffy SW, et al. The Gothenburg Breast Screening Trial: first results on mortality, incidence and mode of detection for women aged 39–49 years at randomization. Cancer 1997;80:2091–9.

[33] Moss SM, Cuckle H, Evans A, et al. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. Lancet 2006;368:2053–60.

[34] Walter SD, Day NE. Estimation of the duration of a pre-clinical disease state using screening data. Am J Epidemiol 1983;118:865–86.

[35] Tabár L, Yen AM, Wu WY, et al. Insights from the breast cancer screening trials: how screening affects the natural history of breast cancer and implications for evaluating service screening programs. Breast J 2015;21:13–20.

[36] Goel V, Cohen MM, Kaufert P, et al. Assessing the extent of contamination in the Canadian National Breast Screening Study. Am J Prev Med 1998;15:206–11.

[37] Taghipour S, Banjevic D, Miller AB, et al. Parameter estimates for invasive breast cancer progression in the Canadian National Breast Screening Study. Br J Cancer 2013;108:542–8.

[38] Baines CJ. The Canadian National Breast Screening Study: a perspective on criticisms. Ann Intern Med 1994;120:326–34.