Cancer Epidemiology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

Comparison of cumulative false-positive risk of screening mammography in the United States and Denmark

Katja Kemp Jacobsen [a,*], Linn Abraham [b], Diana S.M. Buist [b], Rebecca A. Hubbard [c], Ellen S. O'Meara [b], Brian L. Sprague [d], Karla Kerlikowske [e,f], Ilse Vejborg [g], My Von Euler-Chelpin [a], Sisse Helle Njor [a]

[a] Department of Public Health, University of Copenhagen, Copenhagen K, Denmark
[b] Group Health Research Institute, Seattle, WA, United States
[c] Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, United States
[d] Department of Surgery, University of Vermont, Burlington, VT, United States
[e] Department of Epidemiology & Biostatistics, University of California, San Francisco, CA, United States
[f] General Internal Medicine Section, Department of Veterans Affairs, University of California, San Francisco, CA, United States
[g] Center of Diagnostic Imaging, Copenhagen University Hospital, Rigshospitalet, Denmark

ARTICLE INFO

Article history:
Received 12 March 2015
Received in revised form 4 May 2015
Accepted 10 May 2015
Available online xxx

Keywords:
Breast neoplasm
False-positive reactions
Methods
Mass screening
Mammography

ABSTRACT

Introduction: In the United States (US), about one-half of women screened with annual mammography have at least one false-positive test after ten screens. The estimate for European women screened ten times biennially is much lower. We evaluate to what extent screening interval, mammogram type, and statistical methods, can explain the reported differences.

Methods: We included all screens from women first screened at age 50–69 years in the US Breast Cancer Surveillance Consortium (BCSC) (n = 99,455) between 1996–2010, and from two population-based mammography screening programs in Denmark (n = 230,452 and n = 400,204), between 1991–2012 and 1993–2013, respectively. Model-based cumulative false-positive risks were computed for the entire sample, using two statistical methods (Hubbard Njor) previously used to estimate false-positive risks in the US and Europe.

Results: Empirical cumulative risk of at least one false-positive test after eight (annual or biennial) screens was 41.9% in BCSC, 16.1% in Copenhagen, and 7.4% in Funen. Variation in screening interval and mammogram type did not explain the differences by country. Using the Hubbard method, the model-based cumulative risks after eight screens was 45.1% in BCSC, 9.6% in Copenhagen, and 8.8% in Funen. Using the Njor method, these risks were estimated to be 43.6, 10.9 and 8.0%.

Conclusion: Choice of statistical method, screening interval and mammogram type does not explain the substantial differences in cumulative false-positive risk between the US and Europe.

1. Introduction

False-positive tests are an unavoidable consequence of mammography screening. Information on the burden of false-positive tests expected from screening is needed for women in order to make informed decisions about screening participation. From the woman's perspective, it is not only the risk of a false-positive test after attending one screen that is important, but her expected risk of a false-positive test after participating in the multiple rounds of screens called for by a screening program.

Studies from the United States (US), following women with ten years of annual mammography screening, have reported cumulative false-positive risks ranging from 43 to 63% [1–4]. Studies from European mammography screening programs report considerably lower risks, ranging from 8 to 21% after ten biennial screens [5–8]. When comparing estimates of false-positive tests, differences in screening organization and choice of statistical methods should be taken into account since these can affect the estimates. Organization of mammography screening differs considerably between the US and Europe. In the US, there are conflicting guidelines for screening, so that age at first screen, screening interval and number of screens in a woman's lifetime vary

* Corresponding author at: Department of Public Health, University of Copenhagen, Øster Farimagsgade 5, DK 1014 Copenhagen K, Denmark
Tel.: +45 35327735; fax: +45 35327383.
E-mail address: kkja@sund.ku.dk (K.K. Jacobsen).

http://dx.doi.org/10.1016/j.canep.2015.05.004
1877-7821/

significantly [9,10]. European screening programs typically offer biennial screening, but also vary in age range, organization and overall program performance [11].

To our knowledge, this is the first study to compare cumulative false-positive risk of screening mammography between the US and Europe using standardized definitions and statistical methods and long-term follow-up. This study had two objectives: to compare empirical cumulative false-positive risk in different settings and to evaluate whether choice of statistical method results in differences in model-based cumulative false-positive risk. To do this, we applied standard definitions and analysis methods to data from the National Cancer Institute-funded Breast Cancer Surveillance Consortium (BCSC) in the US and from the two long-standing, organized population-based mammography screening programs in Denmark.

Abbreviations: BCSC, Breast Cancer Surveillance Consortium

2. Materials and methods

The National Cancer Institute-funded Breast Cancer Surveillance Consortium (BCSC, http://breastscreening.cancer.gov/) is a collaborative network of seven regional mammography registries, with catchment representative of the US female population of mammography screening age [12]. Within the BCSC, screening is performed in a wide range of delivery systems, including traditional fee-for-service, solo and group radiology practices, managed care organizations, hospital-based radiology practices, free standing mammography centers and mobile van programs.

The BCSC reflects screening practice in the US and contains data from slightly more than 5% of the female population of screening age [13].

The organized, population-based screening programs in Copenhagen and Funen started in 1991 and 1993 respectively, inviting women aged 50–69 years to biennial screening [14]. Screening in Copenhagen and Funen took place at two specialized clinics, supplemented by a mobile van in Funen. Women covered by the two screening programs constituted 20% of Danish women aged 50–69 years. Unlike in the US, all service, including work-up and treatment for Danish women, is free of charge.

In both countries, date of birth, date of screening, type of mammogram, screening history and screening results were collected at the time of screening. In BCSC, breast cancers were obtained by linking mammography data to one or more of three sources: regional surveillance, epidemiology and end results (SEER) registries, state cancer registries, and pathology databases. Completeness of cancer ascertainment is estimated to be >94.3% [15]. In Copenhagen and Funen, breast cancers were obtained by linking mammography data to the Danish cancer registry, the Danish breast cancer cooperative group, and the Danish pathology register. Reporting cancer diagnoses to the Danish cancer registry is mandatory by law in Denmark, and the registry is essentially complete for invasive cancers [16], and supplemented by the other registers above, for ductal carcinoma in situ (DCIS) too..

2.1. Study population

We included women who were first screened at age 50–69 years during 1996–2010 in BCSC, 1991–2012 in Copenhagen and 1993–2013 in Funen. We excluded screens from women with breast implants, a previous mastectomy or diagnosis of invasive breast carcinoma or ductal carcinoma in situ (BCSC: $n = 31{,}111$ women, Copenhagen: $n = 3511$ women, Funen: $n = 3025$). In the Danish data, women with breast implants were only excluded if screening was not technically possible, as data on breast implants were not available. After exclusion, the population covered 1–13 screens/woman in BCSC, 1–8 screens/woman in Copenhagen and 1–10 screens/woman in Funen.

2.2. Definitions

In the BCSC, a mammogram was classified as a screening mammogram based on the indication reported by the radiologist [17]. To avoid misclassifying diagnostic mammograms as screening mammograms, we excluded those that were unilateral or obtained within 270 days after a radiological examination. In Denmark, all program mammograms were classified as screening mammograms. Based on the women's screening history, screens were divided into first screens, including only the first screen for a given woman, and subsequent screens, including all other screens.

The BCSC radiologists used the American College of Radiology's Breast Imaging Reporting and Data System (BI-RADS) [18]. A positive or negative screen referred to the result of the initial assessment which included screening views only. Screens was coded as BI-RADS 0–5, indicating the level of suspicion of malignancy, and were considered positive if the initial BI-RADS assessment was 0 (needs additional imaging evaluation), 4 (suspicious abnormality), 5 (highly suggestive of malignancy), or 3 (probably benign finding) when accompanied by a recommendation for immediate evaluation, and negative if the initial BI-RADS assessment was 1 (negative), 2 (benign finding), or 3 (probably benign finding) without a recommendation for immediate evaluation [18]. Denmark did not use BI-RADS, therefore all screens that led to recall for further work-up were referred to as positive screens without further specification.

A false-positive test was defined as a positive screen where no invasive breast carcinoma or DCIS was diagnosed within one year or prior to the next screen (if this took place before one year).

Subsequent screens were stratified by time since last screen into screening intervals 9–17 months (annual), 18–30 months (biennial) and >30 months (triennial).

2.3. Statistical analysis

We computed empirical false-positive risks as the proportion of false-positive tests for each number of completed screens. Empirical cumulative risk of at least one false-positive were computed as the proportion of at least one false-positive among women who had completed 1–10 screens in BCSC and Funen, and 1–8 screens in Copenhagen, and stratified by screening interval (annual or biennial) and mammogram type (film or digital). Data were censored for four reasons: 1) information about time since last screen differed from self-reported information by ≥ six months (to censor screens in women who were screened outside BCSC); 2) time since last screen was >36 months; 3) BI-RADS assessment or result were missing; 4) when stratifying by screening interval we censored screens where the screening interval differed from previous screening intervals. Similarly, when stratifying data by mammogram type, we censored screens with a different mammogram type compared to previous screens.

Model-based cumulative false-positive risks were estimated using two methods, one developed by Hubbard et al., allowing for variation in false-positive risk among women choosing to attend versus not attend mammography screening, and another method developed by Njor et al. not allowing for this variation [4,6]. In contrast to the Hubbard method, the Njor method assumes independence between screens, meaning a woman with a false-positive test has the same false-positive risk at the next screen as women without a prior false-positive test [4,6]. Nevertheless, this assumption only makes sense if personal characteristics of the screened women such as denser breast do not increase the risk of a subsequent false-positive test. This assumption has previously

Please cite this article in press as: Jacobsen KK, et al. Comparison of cumulative false-positive risk of screening mammography in the United States and Denmark. Cancer Epidemiology (2015), http://dx.doi.org/10.1016/j.canep.2015.05.004