BMC Cancer

The impact of mammography screening programmes on incidence of advanced breast cancer in Europe: a literature review

--Manuscript Draft--

Manuscript Number:	BCAN-D-17-01814R2
Full Title:	The impact of mammography screening programmes on incidence of advanced breast cancer in Europe: a literature review
Article Type:	Research article
Section/Category:	Epidemiology, prevention and public health
Funding Information:	
Abstract:	Background: Observational studies have reported conflicting results on the impact of mammography service screening programmes on advanced breast cancer rates (ABCRs), a correlation that was firmly established in randomized controlled trials. We reviewed and summarized studies of the effect of service screening programmes in the European Union on ABCR and discussed their limitations. Methods: The PubMed database was searched for English language studies published between 01-01-2000 and 01-06-2018. After inspection of titles and abstracts, 220 of the 8644 potentially eligible papers were considered relevant. Their abstracts were reviewed by groups of two authors using predefined criteria. Fifty studies were selected for full paper review, and 22 of these were eligible. A theoretical framework for their review was developed. Review was performed using a ten-point checklist of the methodological caveats in the analysis of studies of ABCR and a standardised assessment form designed to extract quantitative and qualitative information. Results: Most of the evaluable studies support a reduction in ABCR following the introduction of screening. However, all studies were challenged by issues of design and analysis which could at least potentially cause bias, and showed considerable variation in the estimated effect. Problems were observed in duration of follow-up time, availability of reliable reference ABCRs, definition of advanced stage, temporal variation in the proportion of unknown-stage cancers, and statistical approach. Conclusions: We conclude that much of the current controversy on the impact of service screening programmes on ABCRs is due to observational data that were gathered and/or analysed with methodological approaches which could not capture stage effects in full. Future research on this important early indicator of screening effectiveness should focus on establishing consensus in the correct methodology.
Corresponding Author:	Mireille Broeders NETHERLANDS
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	
Corresponding Author's Secondary Institution:	
First Author:	Mireille Broeders
First Author Secondary Information:	
Order of Authors:	Mireille Broeders
	Prue Allgood
	Stephen Duffy
	Solveig Hofvind
	Iris Nagtegaal
	Eugenio Paci

	Sue Moss
	Lauro Bucchi
Order of Authors Secondary Information:	
Response to Reviewers:	The authors' response letter has been included as a supplementary file.

Title:

The impact of mammography screening programmes on incidence of advanced breast cancer in Europe: a literature review

Authors:

M.J.M. Broeders[1,2], P. Allgood[3], S.W. Duffy SW[3], S. Hofvind[4], I.D. Nagtegaal[5], E. Paci[6], S.M. Moss[3], L. Bucchi[7]

Affiliations:

1 Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, The Netherlands, mireille.broeders@radboudumc.nl

2 Dutch Expert Centre for Screening, Nijmegen, The Netherlands

3 Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, United Kingdom, s.w.duffy@qmul.ac.uk, s.m.moss@qmul.ac.uk, p.allgood@qmul.ac.uk

4 Cancer Registry of Norway, Oslo, Norway, solveig.hofvind@kreftregisteret.no

5 Department of Pathology, Radboud University Medical Center, Nijmegen, The Netherlands, iris.nagtegaal@radboudumc.nl

6 Florence, Italy ---- formerly: Clinical and Descriptive Epidemiology Unit, Cancer Research and Prevention Institute (ISPO), Florence, Italy, paci.eugenio@gmail.com

7 Romagna Cancer Registry, Romagna Cancer Institute (Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori, IRST, IRCCS), Meldola, Forli, Italy, lauro.bucchi@irst.emr.it

Corresponding author:

M.J.M. Broeders, Radboud Institute for Health Sciences, Radboud university medical center, PO Box 9101, 6500 HB Nijmegen, The Netherlands, fax: 00-31-24-3613505, email: Mireille.Broeders@radboudumc.nl

Article category:

Review

ABSTRACT

Background: Observational studies have reported conflicting results on the impact of mammography service screening programmes on advanced breast cancer rate (ABCR), a correlation that was firmly established in randomized controlled trials. We reviewed and summarized studies of the effect of service screening programmes in the European Union on ABCR and discussed their limitations.

Methods: The PubMed database was searched for English language studies published between 01-01-2000 and 01-06-2018. After inspection of titles and abstracts, 220 of the 8644 potentially eligible papers were considered relevant. Their abstracts were reviewed by groups of two authors using predefined criteria. Fifty studies were selected for full paper review, and 22 of these were eligible. A theoretical framework for their review was developed. Review was performed using a ten-point checklist of the methodological caveats in the analysis of studies of ABCR and a standardised assessment form designed to extract quantitative and qualitative information.

Results: Most of the evaluable studies support a reduction in ABCR following the introduction of screening. However, all studies were challenged by issues of design and analysis which could at least potentially cause bias, and showed considerable variation in the estimated effect. Problems were observed in duration of follow-up time, availability of reliable reference ABCR, definition of advanced stage, temporal variation in the proportion of unknown-stage cancers, and statistical approach.

Conclusions: We conclude that much of the current controversy on the impact of service screening programmes on ABCR is due to observational data that were gathered and/or analysed with methodological approaches which could not capture stage effects in full. Future research on this important early indicator of screening effectiveness should focus on establishing consensus in the correct methodology.

KEY WORDS

breast cancer, mammography, screening, advanced stage, review

BACKGROUND

A long follow-up is required to assess the impact of mammography screening programmes on breast cancer mortality. The advanced breast cancer incidence rate (hereafter briefly referred to as advanced breast cancer rate, ABCR) can potentially be used as an earlier indicator of the effectiveness of a screening programme. Moreover, since tumour stage at diagnosis is independent of treatment, except for neoadjuvant therapy, analysis of trends in ABCR allows the effects of early detection to be disentangled from those of improvements in treatment [1]. The correlation between reductions in breast cancer mortality and ABCR has been firmly established on the basis of screening trials [2]. In a pooled analysis of data from eight trials, the decrease in the risk of advanced breast cancer and the decrease in the risk of dying from the disease were approximately proportional [1, 3]. It is clear that screening is associated with a reduction in the proportion of advanced stage cancers [4]. However, observational studies published over the last 15-20 years have yielded conflicting results on the association between the introduction of population-based service screening programmes and changes in ABCR, i.e. the absolute incidence of advanced stage disease [3, 5]. Nevertheless, the evaluation of the change in the incidence of advanced breast cancer cases is relevant in service screening outcome research. An apparent lack of this change has been considered by some as evidence of the lack of mammography screening programmes' effectiveness [5-8].

The objectives of the current study were (a) to review studies of the effect of mammography screening programmes in Europe on ABCR, and (b) to summarize their limitations and the extent to which they contribute to the evidence on screening effectiveness.

METHODS

Search strategy and selection criteria

A systematic search of PubMed with the search terms 'cancer stage', 'screening', 'breast cancer', 'incidence', and 'mammography' was performed to identify papers published from January 2000 until May 2013 (details in Appendix) and later updated to June 2018. Only papers in English evaluating European programmes were reviewed. The search strategy was built using 7 key papers [9-15].

Abstracts from the papers identified were reviewed by two from a group of four reviewers (MB, PA, SM, LB) and papers for full review were selected using the following general criteria: (a) the study represented original data and estimated the impact of a current regional or national population-based screening programme in Europe; (b) definition of advanced disease was based on breast cancer size, nodal status and/or stage at diagnosis of breast cancer; (c) the analysis included at least some of the age groups between 50 and 69; (d) the study used an observational research design comparing rates or proportions of advanced stage cancers; and (e) an uninvited and/or unscreened control population was available. This included the pre-screening years for the population targeted for screening in the study. Comparisons only of attenders vs non-attenders were not included. We focused the review on European programmes to add evidence on advanced breast cancer to the European balance sheet of benefits and harms as an outcome to the work of the Euroscreen reviews of observational mortality studies [16].

Definition of advanced breast cancer

Tumour staging criteria vary across studies and even studies using the UICC TNM classification [17] show little agreement in their definition of advanced breast cancer. Theoretically, the benefit of screening is limited to screen-detected cases, either earlier within the same stage or at an earlier stage. However, using stage in itself has a disadvantage due to the stage migration bias caused by the introduction of sentinel lymph node dissection [18] and by changes in coding and classification practices [19]. In this respect, using only the pT information as a proxy for the diameter of the lesion

is the most direct link to radiological detection and less influenced by trends in missing data and changes in coding and classification practices, even though it cannot show within-stage shifts in diameter. It is therefore the least biased option to define advanced breast cancer detection. Tumour size (measured in mm), even though put forward by some authors as an indicator of diagnostic anticipation [20], has never been confirmed as such and is often inaccurate since pathologists tend to round to the nearest multiple of five (terminal digit preference bias) [21].

Theoretical framework and checklist

We designed an assessment form to extract detailed quantitative and qualitative information, the study design, completeness of information and results from the selected papers in a standardized fashion.

The expected effect of mammography service screening programmes on ABCR is best understood looking at the randomized controlled trials (RCTs) as a reference, as previously described [1-3]. Based on the RCTs, the ABCR in the population invited to screening, usually from age 50, is expected to remain stable or slightly increase when the programme starts. The increasing incidence, in comparison with the prescreening incidence rate, is due to the intra-stage shift. This means that screening will detect advanced cancer cases earlier, but still within the same stage as in the absence of screening. After the prevalence screening, assuming a 100% sensitivity, the advanced cancer cases will be diagnosed as interval cancers, if fast growing, or are expected to be detected earlier at subsequent rounds. For this reason, the expectation is a reduction of the ABCR 2-3 years after the start (Figure 1). The advanced cancer cases that are detected earlier through screening than they would have been in the situation without screening are the ones which should benefit. The ABCR should thus decrease from the time of prevalent screening (time 0) to a lower level than the expected, reaching a plateau after a few years, because screening will move diagnoses of breast cancer cases forward in time as long as the programme continues. If screening stops, e.g. at 65 or 69

years in most European screening programmes, the ABCR is expected to increase again, rising after some years to the prescreening level (age-specific) .

– Figure 1 Insert here –

In order to discern this pattern of occurrence, the ABCR with or without screening will be best observed in a study where individual women are followed over time, and an unconfounded comparison of screening with non-screening incidence is available. In order to assess the extent to which studies achieve or approximate this ideal situation, we developed a ten-point checklist of the main methodological issues with which such studies of ABCR have to contend, logically derived from the above described theoretical framework (Table 2). The checklist is based on epidemiological principles of observational studies as applied to screening [22] and previous research experience, including knowledge of the relevant literature from outside of Europe [6, 7, 23-26] and findings of the Euroscreen reviews of observational mortality studies (trend studies, incidence-based mortality studies, and case-control studies) [27, 28]. The methodological issues identified using the ten-point checklist, their definitions, and their consequences on design, likely accuracy, and results of studies are presented in Table 2. This in turn highlights the main potential departures of studies from the ideal design of a study of the temporal association between mammography screening programmes and incidence of advanced stage breast cancer, and indicates the major issues of interpretation of the results.

The checklist items included: 4 complications related to the timescale of screening introduction, periods of exposure and observation, and transient prevalence screen effects; 3 to endpoint definition, stage migration and completeness of stage data; 1 to difficulties of formal inference; and 2 to the inevitable problem of incomplete information on what the incidence of breast

cancer overall and of advanced disease would have been in the absence of the screening programme.

Presentation of results

Due to the heterogeneity in methodology and endpoints used in the studies, no attempt was made to produce a pooled estimate of the effect of screening on ABCR. Instead, we reported details of methods and results of each study individually in Table 1. We looked for data on screening coverage and attendance rates from other sources as well, if the selected study did not provide that information.

RESULTS

Selection of studies

The search strategy identified 8644 English-language papers of which 220 were considered relevant based on title and abstract (Figure 2), including both studies of incidence rates and those of proportions of advanced cancers.

– Insert Figure 2 here –

Based on the selection criteria, 38 studies were included, and a further 24 were identified as possible inclusions. For the latter group, full papers were assessed by two different reviewers, with arbitration by a third (SD) where necessary, which resulted in the inclusion of 4 studies. In addition, the abstract of one paper suggested by a co-author was assessed and included for review. In total, after adding the 7 key papers, 50 studies were included for full paper review by the two reviewers who had not assessed the abstract. We also manually searched the reference lists of these papers

and identified 10 references that fulfilled the inclusion criteria but had not been identified by the search strategy. Review of the full papers for these references resulted in the inclusion of an additional 5 studies. Differences between reviews were resolved through consensus by all four reviewers. Of the 60 full paper reviews in total, 22 studies were found eligible for inclusion in a comparison of incidence rates as the outcome measure [8, 12-15, 19, 29-44]. A further 9 studies were comparisons of proportions of advanced cancers and not included in the current review. Of the 29 papers excluded, 21 lacked a suitable control group, 3 were not related to population-based screening and 5 were excluded for other reasons (no data for 50-69 (n=2), no tumour stage data (n=1), not European Union (n=1) no original data (n=1)).

Study generalities

These are shown in Table 1. The 22 eligible studies were from Norway (n=5), Italy (n=5), the Netherlands (n=4), Denmark (n=2), Sweden, Finland, Germany, United Kingdom (UK), Ireland, and France. There were 9 nation-wide studies, four from Norway [19, 36, 38, 39], two from the Netherlands [14, 41], two from Denmark [8, 37], and one from Finland [34].

– Insert Table 1 here –

Programme characteristics

In most studies, the target age range was 50-69 years [8, 14, 15, 19, 29, 30, 32, 35-41, 44] or wider [12, 31, 43]. The papers from Finland, the West Midlands region of the UK, and Ireland reported programmes aimed at women aged 50-59 years [34] and 50-64 years [13, 42]. The target age of the Swedish programme varied locally between 40 and 74 years [33]. The size of the target population, often not reported, was between 500,000 and 1,000,000 in the national Dutch study [14], in the Danish studies [8, 37] and in one Italian study [15], and exceeded 1,000,000 in the study from

Sweden [33] and in a second study from Italy [32]. The screening interval was 24 months except in the West Midlands (36 months) [13]. The start of screening programmes ranged from the early/mid 1970s in Florence, Utrecht, and Nijmegen [14, 29] to 2005 in the Münster district (Germany) [40]. The time period of observation of breast cancer incidence was between the second half of 1980s and the first half of the current decade in most studies.

Study design

The methods of analysis varied from the provision of purely descriptive information to the evaluation of the magnitude and statistical significance of observed changes in ABCR. We assigned the design of the studies that evaluated the magnitude of effect to four broad categories:

(1) comparison of ABCR before and after the introduction of screening using different endpoints, i.e., annual percent change (APC), percent reduction in ABCR, absolute reduction in ABCR, incidence rate ratio (IRR), relative risk (RR), excess RR, slope value calculated from a log-linear Poisson regression model, and observed:expected ratio, or simply by juxtaposition of rates [8, 12, 15, 19, 29, 30, 32-40, 43, 44];

(2) comparison of ABCR between each year after the introduction of screening and the prescreening years using the estimated annual percent change (EAPC) [14, 31];

(3) calculation of the EAPC after the introduction of screening without information on prescreening years [13, 41]; and

(4) comparison of ABCR in an invited population vs. a neighbouring uninvited one using the percent reduction in ABCR. This is the case for a single study [42], although the inclusion of neighbouring nonscreening areas is a secondary part of the design of other investigations [8, 36].

The statistical significance of observed changes, if any, was assessed in 17 studies [8, 13-15, 30-34, 36-41, 43, 44].

Some information on the trend (before and after the introduction of screening) for the frequency of unknown-stage cancer was provided by 11 studies [8, 12, 15, 19, 29, 30, 32, 33, 35, 38, 39]. The tumour staging criteria varied. Although 20 studies used the UICC TNM classification there was little agreement in the definition of advanced breast cancer. In one study, incidence was presented for multiple stage categories but the advanced category (or categories) was not explicitly identified [29].

Study results

A significantly favourable impact on ABCR was reported by nine studies. In the national Dutch study, ABCR [T2+ with lymph node (N+) and/or distant metastases (M1)] decreased by 12% [14]. In one regional Dutch study, the annual IRR varied between 0.86-0.82 (T2+ cancer) and 0.83-0.72 (N+ cancer) [31]. In the study from Sweden, RRs were 0.74 (tumour size >2 cm), 0.89 (N+ cancer), and 0.84 (Stage II+ cancer) [33]. In the national Finnish study, the ABCR (non-localised cancer) decreased by 9% [34]. A significant impact on ABCR was observed in three studies from Italy. Paci et al. found a RR (Stage II+) of 0.72 [30]. The figure reported by Foca et al. for T2+ cancer was between 0.81-0.71 [15]. A secondary observation from a more recent Italian cohort study comparing attenders and non-attenders was a significant ratio of 0.83 between the observed number of T2+ cancers in a whole invited cohort and the expected number based on pre-screening rates [44]. In a large French study, the decrease was significant both for T2+ cancer and Stage II+ cancer [43]. In a local study from Germany, Simbrich et al. demonstrated significant decreases of varying magnitude in annual ABCR among women aged 50-69 years [40].

Two studies provided unclear results. A Danish study described a transient increase in incidence of cancers >20 mm in size in early screening regions followed by a decline of N+ cancers in late screening regions [37]. The Italian study of Buiatti et al. was limited to ≤3 screening years for most of the participating subareas. After early significant increases in T2+ cancer rates in two of

them, a moderate reduction was observed 4-6 years after the start of the programme in the area with longer follow-up [32].

Four nationwide Norwegian studies reported contradictory findings. Kalager et al. observed a significant IRR (Stage III+ cancer) of 0.76, but the same figure was found in the not-yet invited population before screening [36]. Also, the reduction was confirmed by a second study but in association with an increase for Stage II cancer [39]. Others reported the opposite, that is, a decrease for Stage II cancer and an increase for Stage III cancer [19]. Another study found significant increases both for Stage II and Stage III cancers and a decrease for Stage IV cancer alone [38]. None of these studies used individual data indicating whether women were diagnosed before or after they were invited to participate.

In addition to the abovementioned studies from France [43] and Germany [40], three investigations used the joinpoint analysis or the Poisson regression analysis. In the West Midlands (UK), the incidence of N+ cancer increased in the first years of screening and then returned to the baseline level but with a significant positive APC of 1.1 [13]. In Denmark, the negative APC in incidence of T2+ cancer was significant but the ratio between post-screening and pre-screening rate was not significantly different from the unity [8]. In another study from the Netherlands, a non-significant negative APC in Stage 2+ cancer rate was observed but the estimate included the whole of women aged 50 or older [41].

Four studies, in addition to one of the abovementioned Norwegian studies [19], presented no assessment of significance of observed changes in ABCR (if any). One Italian study reported a 8.7% decrease for N+ cancer [29]. In the fifth Norwegian study, ABCR (regional or distant cancer) rose before the introduction of screening, and fluctuated thereafter at levels that were generally above the last pre-screening level [35]. In a regional Dutch study, ABCR (Stage IIA+ cancer) was described to be stable before and after the introduction of screening [12]. In Ireland, ABCR (Stage 2+) in a region

targeted by screening in 2000 fell by 20% in comparison with a region in which screening was implemented only seven years later [42].

Method check

The right-hand column in Table 2 gives the results of the review of selected papers against the ten-point checklist.

– Insert Table 2 here –

The issue of follow up time (#1) is related to the short time window after prevalence screening where a decrease in ABCR can be observed. Studies with a long time window, most notably seven studies [8, 12, 13, 19, 34, 37, 41] in which the time difference between the year of start of the screening programme and the last year of observation was ≥ 15 years, will not be able to show this decrease. This is particularly problematic when interpreting annual percent changes [13, 41]. If screening is working as anticipated annual percentage changes will be substantial in the first years of a programme, but will be small or absent after the programme has achieved widespread coverage as the new lower incidence will be roughly constant. The related problem of the effect of a dynamic population on exposure time (#2) applies to all studies. Foca et al. excluded women aged 50-54 years but not new immigrants and late attendees [15]. Anttila et al. provided separate data for women aged 50-54 years and 55 years or older [34].

The problem due to pace of implementation (#3) applies especially to the Swedish study [33], the Italian studies [15, 29, 30, 32, 44], the nationwide Norwegian studies [19, 36, 38, 39], the Danish studies [8, 37], and the nationwide Dutch study [14]. In fact, it is rare that a mammography service screening programme is started simultaneously throughout a large geographic area. In two of these

studies, there was explicit adjustment of the analysis to address this issue. In the Swedish study, the first screening years in some counties were omitted from analysis because mammography coverage, or the level of exposure, was still low [33]. In addition, in this study, individual data on screening exposure was available for the nominal screening period. In the study of Foca et al. the years of observation were synchronised at the municipality level, and those municipalities where saturation was not reached within a short (arbitrary) time interval were not taken into consideration [15]. This proved to be a practical but powerful approach to account for gradual programme implementation. In other studies, at least some information was available for the reader to assess the potential size of the problem. The papers reporting the nationwide Dutch study and the Danish study drew the reader's attention to this issue by presenting results for individual years and for regions implementing screening at different times [14, 37]. One of the Italian studies also had individual data on screening exposure during the nominal screening period [30].

The prevalence effect problem (#4) applies virtually to all studies with markedly stepwise implementation of the programme. Of the two problems concerning the reference incidence, the inevitable lack of a verifiable estimate of the underlying background incidence rate (#5) applies to all studies. Outside of a randomised trial, the estimation cannot be performed without assumptions regarding the likely incidence of breast cancer, and specifically late stage breast cancer, in the absence of screening. The problem of its decreasing validity over time (#6) applies especially to those studies, already mentioned above, in which the time interval between the last prescreening year and the last year of observation was ≥ 15 years [8, 12, 13, 19, 34, 37, 41]. However, again, presentation of data for individual years affords the reader a means of assessing the likely extent of underestimation [37].

Difficulties with the definition of advanced cancer (#7) apply to all studies, because all such definitions have pros and cons. Some used the pT information alone [8, 15, 44], others used multiple

advanced stage definitions with separate results [13, 19, 29, 31, 33, 36-39, 43], or a single definition of advanced stage based on the TNM system [12, 14, 30, 32, 34, 35, 40-42].

Of the two problems concerning tumour stage information, the problem of stage migration (#8) applies to all studies except those where the definition of advanced cancer was exclusively based on pT information [8, 15, 44]. More than half of the studies did not take changes in the proportion of unknown stage information (#9) into consideration, providing no trend in missing tumour stage data [12-14, 31, 34, 36, 37, 40-44] or only very partial data [32]. A stable trend was reported by one of the Italian studies [29]. A percent decrease of incident breast cancers with missing stage information was observed in other two Italian studies [15, 30], in the Swedish study [33], in three Norwegian studies [35, 38, 39], and in a study from Denmark [8]. In two of these, the resulting bias was adjusted for in the design [15] and, respectively, in the analysis [33].

Finally, the problem of a lack of standardised statistical approach (#10) applies especially to those studies reporting purely descriptive data [29, 35, 42] or incidence curves without numerical data [12, 19] and those based on the joinpoint analysis [13, 41] and the Poisson regression analysis [8, 40, 43], the results of which are difficult to interpret.

DISCUSSION

The 22 studies included in this review showed considerable variation in results on the estimated effect of the introduction of population-based mammography screening programmes on the ABCR. Of note, there are four circumstantial indications that the overall effect of methodological issues resulted in an underestimation of the impact on ABCR: first, most biases have a conservative direction (#2, #3, #4, #8, and #9); second, most of the largest studies reported a significant decrease in ABCR [14, 15, 33, 44]; third, the decrease was more pronounced after some adjustments for design biases were made [15, 33]; and, fourth, taking the entire series of studies into consideration,

nine of them found a significant, albeit varying, reduction in ABCR. They represent the majority of published studies once those affected by critical limitations are excluded. In our opinion, the report by Buiatti et al. [32], focusing the first 3 years of screening, and the four nationwide Norwegian studies [19, 36, 38, 39], with their conflicting and partly opposite findings, are difficult to interpret. Furthermore, the study by Larsen et al. demonstrated clearly that stage-specific incidence of breast cancer in Norway was influenced by changes in coding and classification practices, which makes it even more challenging to evaluate and compare stage-specific trends and stage migration of breast cancer by age and time [19]

Nonetheless, the conclusions of the available literature still warrant careful interpretation, because not all methodological concerns could be avoided. Also, while the direction of the potential biases can be predicted, it is difficult and sometimes impossible to estimate their magnitude. Some of the problems are unavoidable and apply to all studies (specifically #2, #5, #7), whereas others could potentially be addressed in the design phase. In any case, it would be arbitrary to rank their consequences in terms of relative impact on study results, which may also vary in relation to local contingencies. More realistically, we aimed at summarising the challenges in designing studies on ABCR in order to improve consistency in the reporting of results.

Ideally, the study population should be rapidly saturated by exposure to screening, and this should take less time than that needed for the expected effect on ABCR to become apparent. From this point of view population-based service screening programmes often cannot provide this ideal situation. The dynamic nature of the target populations, together with the phased introduction of most screening programmes and the fact that the prevalence screen will be associated with an increase in ABCR, will lead to an underestimate of the decrease in ABCR, as will the reduction in the proportion of unknown-stage tumours.

In addition, certain statistical analyses, such as the joinpoint analysis (#10), may generate false negative results. Conversely, problems of estimation of underlying incidence in the absence of screening, and particular definitions of advanced stage (#5 and #7) may have been responsible for unpredictable effects in either direction. Many of the problems also arise from the reliability and validity of incidence data, in particular the unavailability of reliable reference incidence rates for advanced cancer, especially in a historical comparison period, together with the sharp decrease in the proportion of unknown-stage cancers following the introduction of screening. Stage migration bias, caused by the implementation of sentinel lymph node biopsy between the mid-1990s and early 2000s [18, 19], will also have had an impact.

Furthermore, the inconsistency in the definition of advanced cancers gives rise to difficulties in interpreting the collected evidence. There is a possibility of a residual improvement within stage categories, but this is more difficult to demonstrate. The consistency between studies in the use of tumour diameter, stage and other parameters was limited. Another limitation in the classification of advanced cancers, especially in studies performed nowadays, is the variation among cancer registries (and within cancer registries over time) in what clinical and pathological data they collect. There is growing interest in the effect of screening, if any, on biological and molecular markers, but it will be some time before sufficient data are generated to answer this question. Incidentally, we believe that deficiencies in staffing, organisation, access, and funding of ongoing mammography service screening programmes warrant much greater consideration in the debate about their effectiveness.

From a scientific point of view, however, the most severe limitations of reviewed studies (#1 to #4) affected the study design. The main departures from the ideal design of a temporal correlation study were the following. First, as shown in the Swedish Two-County trial [2, 15], the time window available to observe an impact (if any) on ABCR closes rapidly. In populations where screening has been ongoing for a longer time [12, 13, 41], analysis should focus on establishing whether incidence

of advanced disease is lower than before, not 'still decreasing'. The misuse of the joinpoint analysis and of the Poisson regression analysis (#10) is itself related to the assumption that the downward incidence trend must continue indefinitely [13]. This cannot be the case, unless a substantial increase of mammography sensitivity occurs over time. Second, the 3-year latency of the effect of screening on ABCR means that, in the dynamic target population of a service screening programme, at any point in time, there is always a subset of women with an exposure time to screening that is too short to have an effect on the risk of advanced breast cancer. Third, and more important, service screening programmes in Europe were introduced very gradually. This inevitably caused the same dilution of effects as that historically described for cervical cancer screening in Denmark and Norway as compared with Finland and Sweden [34].

In fairness, most of the studies reviewed either attempted to control for possible problems by adjustment in statistical analysis or presented data in sufficient detail for the reader to judge the likely presence and direction of potential biases. There have been surprisingly few attempts, on the other hand, to adjust the design to minimise biases. The only previous literature review on ABCR following the introduction of mammography screening programmes did not take into consideration the limitations of published articles, except for the stage migration bias [5, 19]. The authors concluded that trends in advanced breast cancer incidence do not support a role for screening in the decrease in mortality. The present work demonstrates that the available literature cannot support such a conclusion, and indeed supports the opposite.

CONCLUSIONS

In summary, all studies were challenged by multiple issues, although to a varying extent. The trend in most of evaluable results, even though inconsistent, does support a reduction in advanced breast

cancer incidence following the introduction of mammography screening. In view of the impact on ABCR observed in RCTs [1], we conclude that much of the current controversy on mammography service screening programmes is due to observational data that were gathered and/or analysed with methodological approaches which could not capture stage effects in full [27, 28]. Notwithstanding this fact, changes in ABCR remain an important early indicator of effectiveness. Improving the knowledge of limitations in previous studies will help to establish consensus on the correct methodology. The development of more robust and empirically driven techniques should take into account both the practical implementation of cancer screening activities and the evaluation of their effects. This will enable a better fit of the design of studies on ABCR to the particular context of a mammography service screening programme.

LIST OF ABBREVIATIONS

ABCR: advanced breast cancer rate;

APC: annual percent change;

CI: confidence interval;

EAPC: estimated annual percent change;

IRR: incidence rate ratio;

M1: distant spread;

N+: node-positive;

NA: not applicable;

NOS: not otherwise specified;

NR: not reported;

NS: not significant;

O:E: observed:expected;

pT: pathologic tumour size category;

RCT: randomized controlled trial;

RR: relative risk;

S: significant;

SOSSEG: Swedish Organised Service Screening Evaluation Group;

T2+: tumour size >2cm;

TNM: Tumour, Node, Metastasis;

TX: unknown tumour size;

UICC: Unione Internationale Contre le Cancer;

UK: United Kingdom;

W: women

DECLARATIONS

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

All data generated or analysed during this study are included in this published article.

Competing interests

MB is a member of the editorial board (Associate Editor) of BMC Cancer. The other authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

MB conceived of the idea for the study, designed the study, analysed and interpreted the data, and drafted the manuscript. PA coordinated the literature search, and analysed and interpreted the data. SD analysed and interpreted the data and helped to draft the manuscript. SH conceived of the idea of the study, analysed and interpreted the data. IN analysed and interpreted the data. EP contributed to the design of the study, analysed and interpreted the data. SM conceived of the idea for the study, designed the study, analysed and interpreted the data, and helped to draft the manuscript. LB designed the study, analysed and interpreted the data, and drafted the manuscript. All authors critically reviewed the manuscript and provided final approval for submission.

Acknowledgements

We would like to thank Roberta Maroni and Zoheb Shah for their help in updating the literature search.

REFERENCES – updated

1. Autier P, Héry C, Haukka J, Boniol M, Byrnes G. Advanced breast cancer and breast cancer mortality in randomized controlled trials on mammography screening. J Clin Oncol. 2009;27:5915-23.
2. Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. Radiol Clin North Am. 1992;30:187-210.
3. Tabár L, Yen AM, Wu WY, Chen SL, Chiu SY, Fann JC, 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.
4. Nagtegaal ID, Duffy SW. Reduction in rate of node metastases with breast screening: consistency of association with tumor size. Breast Cancer Res Treat. 2013;137:653-63.
5. Autier P, Boniol M, Middleton R, Doré JF, Héry C, Zheng T, et al. Advanced breast cancer incidence following population-based mammographic screening. Ann Oncol. 2011;22:1726-35.
6. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. N Engl J Med. 2012;367:1998-2005.
7. Welch HG, Prorok PC, O'Malley AJ, Kramer BS. Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. N Engl J Med. 2016;375:1438-47.
8. Jørgensen KJ, Gøtzsche PC, Kalager M, Zahl P-H. Breast cancer screening in Denmark. A cohort study of tumor size and overdiagnosis. Ann Intern Med. 2017;166:313-23.
9. Ernst MF, Voogd AC, Coebergh JW, Roukema JA. Breast carcinoma diagnosis, treatment, and prognosis before and after the introduction of mass mammographic screening. Cancer. 2004;100:1337-44.
10. Hofvind S, Lee CI, Elmore JG. Stage-specific breast cancer incidence rates among participants and non-participants of a population-based mammographic screening program. Breast Cancer Res Treat. 2012;135:291-9.
11. Mook S, Van 't Veer LJ, Rutgers EJ, Ravdin PM, Van de Velde AO, Van Leeuwen FE, et al. Independent prognostic value of screen detection in invasive breast cancer. J Natl Cancer Inst. 2011;103:585-97.
12. Nederend J, Duijm LE, Voogd AC, Groenewoud JH, Jansen FH, Louwman MW. Trends in incidence and detection of advanced cancer at biennial screening mammography in the Netherlands: a population based study. Breast Cancer Res. 2012;14:R10.
13. Autier P, Boniol M. The incidence of advanced breast cancer in the West Midlands, United Kingdom. Eur J Cancer Prev. 2012;21:217-21.
14. Fracheboud J, Otto SJ, Van Dijck JAAM, Broeders MJ, Verbeek AL, de Koning HJ, et al. Decreased rates of advanced breast cancer due to mammography screening in the Netherlands. Br J Cancer. 2004;91:861-7.
15. Foca F, Mancini S, Bucchi L, Zappa M, Naldoni C, Falcini F, et al. Decreasing incidence of late-stage breast cancer after the introduction of organized mammographic screening in Italy. Cancer. 2013;119:2022-8.
16. Paci E, EUROSCREEN working group. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. J Med Screen. 2012;19 Suppl 1:5-13.
17. UICC. TNM classification of malignant tumours. 7th ed. New York: Wiley-Blackwell; 2009.
18. Maaskant AJ, van de Poll-Franse LV, Voogd AC, Coebergh JW, Tutein Nolthenius-Puylaert MBCJE, Nieuwenhuijzen GAP. Stage migration due to introduction of the sentinel node procedure: a population-based study. Breast Cancer Res Treat. 2009;113:173-9.
19. Larsen IK, Myklebust TÅ, Johannesen TB, Møller B, Hofvind S. Stage-specific incidence and survival of breast cancer in Norway: the implications of changes in coding and classification practice. Breast. 2018;38:107-13.

20. Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: a randomised screening trial. BMJ. 2014;348:g366.
21. Bucchi L, Barchielli A, Ravaioli A, Frederico M, De Lisi V, Ferretti S, et al. Screen-detected vs clinical breast cancer: the advantage in the relative risk of lymph node metastases decreases with increasing tumor size. Br J Cancer. 2005;92:156-61.
22. Anttila A, Läärä E. Cervix cancer: geographical correlations. In: Sankila R, Démaret E, Hakama M, Lynge E, Schouten LJ, Parkin DM, editors. Evaluation and monitoring of screening programmes. Luxembourg: Office for Official Publications of the European Communities; 2001. p. 77-98.
23. Coburn NG, Chung MA, Fulton J, Cady B. Decreased breast cancer tumor size, stage, and mortality in Rhode Island: an example of a well-screened population. Cancer Control. 2004;11:222-30.
24. Escobedo LG, Zhong Z, Key C. Breast and cervical cancer screening and disease incidence and stage in New Mexico. Cancer Causes Control. 2002;13:137-45.
25. Harmer C, Staples M, Kavanagh AM. Evaluation of breast cancer incidence: is the increase due entirely to mammographic screening? Cancer Causes Control. 1999;10:333-7.
26. Kricker A, Farac K, Smith D, Sweeny A, McCredie M, Armstrong BK. Breast cancer in New South Wales in 1972-1995: tumor size and the impact of mammographic screening. Int J Cancer. 1999;81:877-80.
27. Broeders M, Moss S, Nyström L, Njor S, Jonsson H, Paap E, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. J Med Screen. 2012;19 Suppl 1:14-25.
28. Moss SM, Nyström L, Jonsson H, Paci E, Lynge E, Njor S, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of trend studies. J Med Screen. 2012;19 Suppl 1:26-32.
29. Barchielli A, Paci E. Trends in breast cancer mortality, incidence, and survival, and mammographic screening in Tuscany, Italy. Cancer Causes Control. 2001;12:249-55.
30. Paci E, Duffy SW, Giorgi D, Zappa M, Crocetti E, Vezzosi V, et al. Quantification of the effect of mammographic screening on fatal breast cancers: the Florence Programme 1990-96. Br J Cancer. 2002;87:65-9.
31. Schouten LJ, de Rijke JM, Huveneers JA, Verbeek ALM. Rising incidence of breast cancer after completion of the first prevalent round of the breast cancer screening programme. J Med Screen. 2002;9:120-4.
32. Buiatti E, Barchielli A, Bartolacci S, Federico M, De Lisi V, Bucchi L, et al. The impact of organised screening programmes on the stage-specific incidence of breast cancer in some Italian areas. Eur J Cancer. 2003;39:1776-82.
33. Swedish Organised Service Screening Evaluation Group. Effect of mammographic service screening on stage at presentation of breast cancers in Sweden. Cancer. 2007;109:2205-12.
34. Anttila A, Sarkeala T, Hakulinen T, Heinävaara S. Impacts of the Finnish service screening programme on breast cancer rates. BMC Public Health. 2008;8:38.
35. Hofvind S, Sørum R, Thoresen S. Incidence and tumor characteristics of breast cancer diagnosed before and after implementation of a population-based screening program. Acta Oncol. 2008;47:225-31.
36. Kalager M, Adami HO, Bretthauer M, Tamimi RM. Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. Ann Intern Med. 2012;156:491-9.
37. Christiansen P, Vejborg I, Kroman N, Holten I, Garne JP, Vedsted P, et al. Position paper: breast cancer screening, diagnosis, and treatment in Denmark. Acta Oncol. 2014;53:433-44.

38. Lousdal ML, Kristiansen IS, Møller B, Støvring H. Trends in breast cancer stage distribution before, during and after introduction of a screening programme in Norway. Eur J Public Health. 2014;24:1017-22.
39. Lousdal ML, Kristiansen IS, Møller B, Støvring H. Effect of organised mammography screening on stage-specific incidence in Norway: population study. Br J Cancer. 2016;114:590-6.
40. Simbrich A, Wellmann I, Heidrich J, Heidinger O, Hense HW. Trends in advanced breast cancer incidence rates after implementation of a mammography screening program in a German population. Cancer Epidemiol. 2016;44:44-51.
41. Autier P, Boniol M, Koechlin A, Pizot C, Boniol M. Effectiveness of and overdiagnosis from mammography screening in the Netherlands: population based study. BMJ. 2017;359:j5224.
42. Hanley JA, Hannigan A, O'Brien KM. Mortality reductions due to mammography screening: contemporary population-based data. PLoS One. 2017;12:e0188947.
43. Molinié F, Delacour-Billon S, Tretarre B, Delafosse P, Seradour B, Colonna M. Breast cancer incidence: decreasing trend in large tumours in women aged 50-74. J Med Screen. 2017;24:189-94.
44. Puliti D, Bucchi L, Mancini S, Paci E, Baracco S, Campari C, et al. Advanced breast cancer rates in the epoch of service screening: the 400,000 women cohort study from Italy. Eur J Cancer. 2017;75:109-16.
45. Fracheboud J, De Gelder R, Otto SJ, Van Ineveld BM, Otten JDM, Broeders MJM, et al. National evaluation of breast cancer screening in the Netherlands, 1990-2007. Twelfth evaluation report (in Dutch). XII. Rotterdam: Dept of Public Health, Erasmus University Rotterdam; 2009.
46. Swedish Organised Service Screening Evaluation Group. Reduction in breast cancer mortality from organized service screening with mammography: 1. Further confirmation with extended data. Cancer Epidemiol Biomarkers Prev. 2006;15:45-51.
47. Nagtegaal ID, Allgood PC, Duffy SW, Kearins O, Sullivan EO, Tappenden N, et al. Prognosis and pathology of screen-detected carcinomas: how different are they? Cancer. 2011;117:1360-8.
48. Lawrence G, O'Sullivan E, Kearins O, Tappenden N, Martin K, Wallis M. Screening histories of invasive breast cancers diagnosed 1989-2006 in the West Midlands, UK: variation with time and impact on 10-year survival. J Med Screen. 2009;16:186-92.
49. Broeders M, Nyström L, Ascunce N, Riza E, Becker N, Törnberg S, et al. Epidemiological guidelines for quality assurance in breast cancer screening. In: Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L, editors. European guidelines for quality assurance in breast cancer screening and diagnosis. Luxembourg: Office for Official Publications of the European Communities; 2006. p. 15-56.
50. Day NE, Williams DRR, Khaw KT. Breast cancer screening programmes: the development of a monitoring and evaluation system. Br J Cancer. 1989;59:954-8.

LEGENDS

Tables 1 and 2 and figure 2 in this paper are original for this article.

Figure 1 is reproduced with permission from Foca et al. [15].

Figure 1. Expected effect of mammography service screening on the occurrence of advanced breast cancer, illustrated by Figure 2, right panel, from Foca et al. [15]. Ratios with 95% confidence intervals are illustrated between the observed and expected age-standardised incidence rates of breast cancer per 100,000 women according to 2-year screening period (ages 55 to 74 years). pT indicates pathologic tumour classification.

Figure 2. Flowchart of search strategy and selection of papers

Table 1. Characteristics of the screening programmes, and design and results of studies of the impact of mammography screening on the incidence of advanced breast cancer

Table 2. Ten-point checklist of main methodological problems affecting studies of the effect of mammography screening programmes on the incidence of advanced breast cancer

ADDITIONAL FILES

Additional file 1. Search strategy

Additional file 2. Rebuttal

Table 1. Characteristics of the screening programmes, and design and results of studies of the impact of mammography screening on the incidence of advanced breast cancer

Study generalities					
	First author	Barchielli A	Paci E	Schouten LJ	Buiatti E
	Year of publication	2001	2002	2002	2003
	Country	Italy	Italy	The Netherlands	Italy
	Regional area(s)	Florence area (city of Florence and surrounding municipalities)	City of Florence	Limburg	7 areas in central and northern Italy
The screening programme					
	Target age (years)	50-69	50-69	49-69 (49-75 since 1998)	50-69
	Target population	164,000[b]	60,000	NR	1,033,000
	Screening interval (mos)	24	24	24	24
	Year of start	Some municipalities, early 1970s; city of Florence, 1990; other municipalities, after 1992	1990	1990	Locally varying between 1990-98
	Year of saturation[a]	After the end of the time period of observation (see Remarks)	1993	1994	After the end of the time period of observation
	Response rate	60%	NR	First invitation, annually 25-82%; subsequent invitations, 77-85%	65% [15]
Study design and results					
	Time period of observation	1985-94	1985-96	1987-99	Prescreening years, locally varying between 1988-97; screening years, locally varying between 1990-99
	Design	Study of all-age incidence by stage in 1985-87, 1988-90, and 1991-94, with a focus on W aged 50-69	Study comparing ABCR in 1990-96 (screening period) vs 1985-86 (prescreening period), and in invited W vs noninvited W	Study comparing ABCR in each year 1987-99 (screening years) vs 1987-90 (prescreening period)	Study comparing ABCR in the screening period vs the prescreening period, by area, in W aged 40-79
	Endpoint	% change in incidence rates, by stage, in 1991-94 vs 1985-87	% and absolute reduction in ABCR, and invited:noninvited RR	IRR	IRR
	Tumour staging	Tumour spread (local, regional, distant)	UICC TNM	UICC TNM	UICC TNM
	Definition of advanced stage	None specified	Stage II+	Distinct definitions: T2+, N+	Tumour size >2 cm or N+ or Stage IV

	Frequency of unknown stage cancer	Unknown tumour spread, stable incidence rate	Unknown stage: 1985-86, 14%; 1990-96, 7%	NR	Unknown stage, 6% with a significant reduction in the screening period in one area
	Results	W aged 50-69: regional, -8.7% (significance, NR); distant, NR	% reduction in ABCR, -19; absolute reduction, -3.6 per 10,000; RR, 0.72 (95% CI, 0.59-0.87)	T2+: increase in 1991 (IRR, 1.22; 95% CI, 1.09-1.37), decrease in 1998 (0.86; 0.77-0.97) and 1999 (0.82; 0.73-0.92). N+: increase in 1991 (IRR, 1.28; 1.13-1.45), decrease in 1995 (0.83; 0.73-0.94) and 1999 (0.72; 0.63-0.81)	IRR by area, from 0.91 (p = 0.07) to 1.21 (p = 0.02)
	Remarks	By 1995, only part of municipalities of the Florence area were targeted by screening			The study was limited to ≤3 screening years for 5/7 areas. A moderate reduction in ABCR was observed 4-6 years after the start of the programme in one area

Table 1. CONTINUED

Study generalities					
	First author	Fracheboud J	SOSSEG	Anttila A	Hofvind S
	Year of publication	2004	2007	2008	2008
	Country	The Netherlands	Sweden	Finland	Norway
	Regional area(s)	NA (nationwide study)	13 counties	NA (nationwide study)	Rogaland, Akershus, Hordaland, and Oslo counties
The screening programme					
	Target age (years)	50-69	Locally varying between 40-74	Mainly 50-59	50-69
	Target population	813,000 in 1997 [45]	4,403,000 person-years	NR	NR
	Screening interval (mos)	24	24 in most counties	24	24
	Year of start	Utrecht and Nijmegen regions ("old" regions), mid-1970s; the 7 remaining regions ("new" regions), 1990-91	Locally varying between 1988-96	1987	Rogaland county, 1995; Akershus, Hordaland, and Oslo counties, 1996
	Year of saturation[a]	Approximately 1994	NA	1992	NR
	Response rate	78% [45]	70-90% [46]	NR	First 10 years, national average 76%
Study design and results					
	Time period of observation	1989-97	Prescreening epoch, locally varying between 1968-95; screening epoch, locally varying between 1988-2001 (see Remarks)	1971-2002	1987-2004
	Design	Study of all-age incidence in each year 1990-97 vs 1989, by group of regions, with a focus on W aged 50-69	Study comparing advanced cancer risk in the screening epoch vs the prescreening epoch (W aged 40-69)	Study comparing the observed ABCR in the years 1998-2002 with that expected based on extrapolation of rates from 1971 to 1986 in 5-year age groups between 50-69	Study of ABCR in 1987-95 (prescreening period) and 1996-2004 (screening period)
	Endpoint	EAPC in ABCR	RR of advanced cancer adjusted for the proportion with missing stage data and the increase in underlying incidence	Excess RR in %	ABCR
	Tumour staging	UICC TNM	UICC TNM	Tumour spread (localised, non-	Tumour spread (local, regional, distant)

				localised). Non-localised (or regional/distant) spread mainly based on lymph node status	
	Definition of advanced stage	T2+ and (N+ and/or M1)	Distinct definitions: tumour size >2 cm, N+, Stage II+	Non-localised spread	Regional or distant spread
	Frequency of unknown stage cancer	TX, between 2.1% and 3.2% annually (no time trend data)	Tumour size unknown: prescreening epoch, median among counties 12%; screening epoch, 1%[b]	Unknown stage (NOS), 9.4% (no time trend data)	Unknown regional spread, NR. Unknown distant spread: 1987-95, 2%; 1996-2004, 7%[b]
	Results	EAPC in ABCR: new regions, +3 up to 1994 and -2.14 (95% CI, -3.47 to -0.80) between 1995-97, for a total of -12.1 in 1997 vs 1989 (63.0 vs 71.6/100,000); old regions, -5.5 (-8.52 to -2.37)	RR of tumour size >2 cm, 0.74 (95% CI, 0.69-0.79); RR of N+, 0.89 (0.84-0.95); RR of Stage II+, 0.84 (0.79-0.89)	Excess RR: W aged 50-54, -6% (NS); W aged 55-59, -18% (S); W aged 64-64, -21% (S); W aged 65-69, -16% (S); total, -9% (S)	ABCR, increase from 75 to 86 in 1987-95, 98 and 96 in 1996 and 1997, fluctuation between 84 and 99 in 1998-2005 (significance, NR)
	Remarks		In 5 counties, the prescreening and screening epochs were not contiguous, in order to have a coverage close to zero and 100%, respectively	The implementation of the programme had a stepwise “pseudo-randomized” design for evaluation purposes	

Table 1. CONTINUED

Study generalities					
	First author	Autier P	Kalager M	Nederend J	Foca F
	Year of publication	2012	2012	2012	2013
	Country	England	Norway	The Netherlands	Italy
	Regional area(s)	West Midlands	NA (nationwide study)	Southern region	700 municipalities in 6 administrative regions of central and northern Italy
The screening programme					
	Target age (years)	50-64	50-69	50-75	50-69
	Target population	NR	NR	NR	693,000
	Screening interval (mos)	36	24	24	24
	Year of start	1988	4 counties, 1996; the remaining 15 counties, over the following 9 years	1990-91 [14]	Locally varying between 1991-2005
	Year of saturation[a]	1991	2005	NR	NA (see Design)
	Response rate	1992-94, 70%; 1995-2004, 75%[b]	77%	NR	65%
Study design and results					
	Time period of observation	1989-2004 (no prescreening years)	1986-2005	1980-2008	Locally varying between 1990-2006
	Design	Joinpoint regression analysis of time trend in annual ABCR	Study comparing ABCR in the invited population (1998-2005, i.e. excluding the prevalence round) with the prescreening population (1987-95)	Study of ABCR in each year 1980-2008	Study comparing observed ABCR with expected (prescreening) ABCR, by year of screening (W aged 55-74). For each municipality, the screening years were numbered from 1 to 8
	Endpoint	APC	IRR	No numerical endpoints: curve of ABCRs as a marginal information in a study of the prevalence of advanced cancer among screened W	IRR
	Tumour staging	UICC TNM	UICC TNM	UICC TNM	UICC TNM
	Definition of advanced stage	Distinct definitions: tumour size >50 mm, N+	Stage III+	Stage IIA+	T2+
	Frequency of unknown stage cancer	NX, 20% (no time trend data)	NR	Unknown stage: screen-detected cancers, 0.1% (stable); interval	TX: year 1, 10%; year 2, 9%; thereafter, <5%

				cancers, 0%; others and prescreening, NR	
	Results	>50 mm in size: APC, 0.2 (95% CI, -2.2 to 2.7) in 1989-2004. N+: increase in ABCR in 1989-92, decrease in 1993-95, stable return to the level of 1989 in 1995-2000; APC, -0.7 (-1.8 to 0.3) in 1989-2004, 1.1 (0.1-2.0) in 1992-2004	IRR, 0.76 (95% CI, 0.61-0.91)	Curve interpreted as showing that ABCR was stable between 1980-2008 and did not decline after the introduction of screening	IRR: no significant changes in years 1-2, between 0.81 (95% CI, 0.75-0.88) and 0.71 (0.64-0.79) from years 3-4 onward
	Remarks	ABCRs were calculated with a 33-step procedure using total incidence data (http://ci5.iarc.fr) and published tumour stage data from the screening programme [47, 48]	An IRR of 0.76 (95% CI, 0.61-0.91) was also observed in not-yet invited population (1996-2003 vs 1986-94)		Eligibility was restricted to those municipalities in which the proportion of total incident cancers that were screen-detected (a proxy of saturation) reached the arbitrary level of 30% within year 2

Table 1. CONTINUED

Study generalities					
	First author	Christiansen P	Lousdal ML	Lousdal ML	Simbrich A
	Year of publication	2014	2014	2016	2016
	Country	Denmark	Norway	Norway	Germany
	Regional area(s)	NA (nationwide study)	NA (nationwide study)	NA (nationwide study)	Münster district
The screening programme					
	Target age (years)	50-69	50-69	50-69	50-69
	Target population	NR	NR	NR	NR
	Screening interval (mos)	24	24	24	24
	Year of start	Old regions: Copenhagen municipality, 1991; Funen county, 1993. Late regions: Bornholm municipality, 2001; West Zealand county, 2004; the remaining areas, 2007	One county, 1995; the remaining 18, during the following 9 years	1995	2005
	Year of saturation[a]	2010	2004	2004	2008
	Response rate	First screen: Copenhagen, 71%; Funen, 85%. Subsequent screens: Copenhagen, 62%; Funen, 82%	76% [35]	76% [35]	55%
Study design and results					
	Time period of observation	1990-2011 both for early and late screening regions	1987-2010	1987-2011	2000-13
	Design	Study of all-age incidence by stage, with a focus on W aged 50-69	Study of all-age incidence comparing ABCR in 2005-10 (screening period) vs 1987-95 (prescreening period), with a focus on W aged 50-69	Open cohort study of ABCR in W eligible for screening vs the historic (prescreening) population of W of the same age	Log-linear Poisson regression analysis of time trend in ABCR in 2006-08 (implementation phase) and 2009-13 by 5-year age group between 45-79
	Endpoint	ABCR	IRR	IRR	Slope value from the log-linear Poisson regression model (average annual change), and absolute ABCR difference (2013 vs 2000)

	Tumour staging	UICC TNM	UICC TNM	UICC TNM	UICC TNM
	Definition of advanced stage	Distinct definitions: tumour size >20 mm, N+	Distinct definitions: Stage II, Stage III, Stage IV	Distinct definitions: Stage II, Stage III+	Stage II+
	Frequency of unknown stage cancer	NR	Unknown stage: 1987-95, 9%; 2005-10, 4%	Missing information (NOS): 1987-94, 30%; 1995-2002, 19%; 2003-11, 7%	TX and/or NX, 10% (no time trend data)
	Results	>20 mm in size, transient increase in 2008-09 in old screening regions; N+, significant decline from 117 in 2001-07 to 98 in 2010-2011 in late screening regions	Stage II: IRR, 1.47 (95% CI, 1.40-1.55). Stage III: IRR, 1.32 (1.13-1.55). Stage IV, 0.67 (0.57-0.68). Total advanced: IRR, 1.35 (1.29-1.42)	Stage II: IRR, 1.26 (95% CI, 1.21-1.31). Stage III+: IRR, 0.80 (0.74-0.87)	Average annual change (2009-2013): W aged 50-54, 0.016 (95% CI, -0.024 to 0.056); W aged 55-59, -0.054 (-0.095 to -0.014); W aged 60-64, -0.089 (-0.128 to -0.050); W aged 65-69, -0.113 (-0.153 to -0.073). Absolute ABCR difference (2013 vs 2000): W aged 50-54, -0.002 (-0.191 to 0.187); W aged 55-59, -0.346 (-0.533 to -0.160); W aged 60-64, -0.279 (-0.454 to -0.105); W aged 65-69, -0.320 (-0.515 to -0.126)
	Remarks		IRRs for W aged 20–49, who were presented as a control group for time trends in stage-specific incidence, were similar to those for W aged 50-69	Missing stage values were multiply imputed. We report unadjusted estimates, since the purpose and necessity of adjustment were not clear. For each analysis, the IRRs for the screening vs historic group were also compared with the IRRs for the younger (ineligible) vs younger historic group. The unadjusted relative IRR was 1.14 (95% CI, 1.07-1.22) for Stage II and 1.00 (0.87-1.15) for Stage III+	Missing stage values were multiply imputed

Table 1. CONTINUED

Study generalities					
	First author	Autier P	Hanley JA	Jørgensen KJ	Molinié F
	Year of publication	2017	2017	2017	2017
	Country	The Netherlands	Ireland	Denmark	France
	Regional area(s)	NA (nationwide study)	23 out of 26 counties	NA (nationwide study)	Hérault, Isère, Loire-Atlantique
The screening programme					
	Target age (years)	50-69 (50-75 since 1997)	50-64	50-69	50-74
	Target population	NR	NR	703,289	NR
	Screening interval (mos)	24	24	24	24
	Year of start	1988	11 counties, 2000 (Region 1); 12 counties, 2007 (Region 2)	Locally varying between 1991-2007	Locally varying during the 1990s
	Year of saturation[a]	NR	NR	Coverage still incomplete at the end of the time period of observation	NR
	Response rate	Around 80%	68-76%	62- 82%	NR
Study design and results					
	Time period of observation	1989-2012	2000-13	1980-2010	2000-10
	Design	Multi-objective study, with a joinpoint regression analysis of time trend in annual ABCR 1989-2012 for W aged ≥50	Multi-objective study, with a comparison of annual ABCR between Region 1 and Region2	Multi-objective study, with a Poisson regression analysis of time trend in annual ABCR 1980-2010	Poisson regression analysis of time trend in ABCR in 2000-10 among W aged 20-49, 50-74, and 75 and older
	Endpoint	APC	ABCR in Region 1 minus ABCR in Region 2 as a percentage of the latter	APC, and ABCR ratio before and after the introduction of screening, both in the screening and nonscreening areas	APC
	Tumour staging	UICC TNM	UICC TNM	UICC TNM	UICC TNM
	Definition of advanced stage	Stage 2+	Stage 2+	T2+	Distinct definitions: T2+, Stage II+
	Frequency of unknown stage cancer	Unknown stage: 2009-11, 1% (no time trend data)	NR	TX: 1980-2004, 8-10%; 2004-10, 4-5%	TX, 3% (no time trend data)
	Results	APC, -0.16 (95% CI, -0.36 to 0.04)	ABCR, 20% lower in Region 1 in 2007	Screening areas. APC in ABCR:	W aged 50-74. T2+: APC, - 1.9 (95% CI, -

			(significance, NR) and then narrowing	before screening, -0.5 (95% CI, -1.9 to 0.9); after screening, -1.1 (-1.8 to -0.3). ABCR ratio, 0.96 (0.90 to 1.02)	2.8 to -1.0). Stage II+: APC, -2.0 (-2.7 to -1.3)
	Remarks	Stage 2+ probably indicates T2+, since a tumour size of 20 mm is referred to as the threshold size to distinguish between Stage 1 and 2	It is not clear whether Stage 2+ indicates Stage II+	Nonscreening areas. APC in ABCR: before screening, 1.7% (95% CI, 0.8% to 2.6%); after screening, 3.0% (2.6% to 3.3%). ABCR ratio, 1.46 (1.41 to 1.52)	Overall, a 20.9% linear decrease in T2+ cancer over 11 years in the three screening areas were noted for W aged 50-74. No change in ABCR was found in younger or older W

Table 1. CONTINUED

Study generalities			
	First author	Puliti D	Larsen IK
	Year of publication	2017	2018
	Country	Italy	Norway
	Regional area(s)	Nine health care districts in central and northern Italy	NA (nationwide study)
The screening programme			
	Target age (years)	50-69	50-69
	Target population	413,000	NR
	Screening interval (mos)	24	24
	Year of start	Locally varying between 1991-98	1996
	Year of saturation[a]	Locally varying between 1993-2000	2005
	Response rate	NR	75%
Study design and results			
	Time period of observation	Locally varying between 1991-2011	1980-2015
	Design	Cohort study of ABCR in attenders to screening vs non-attenders, with a comparison of the observed number of ABC among W invited with that expected based on prescreening ABCR	Study of all-age stage-specific incidence based on different staging systems, with a focus on W aged 50-69
	Endpoint	O:E ratio	No numerical endpoints: curve of ABCRs
	Tumour staging	UICC TNM	UICC TNM
	Definition of advanced stage	T2+	Distinct definitions: Stage II, Stage III, Stage IV
	Frequency of unknown stage cancer	TX, 10-29% (no time trend data)	Unknown stage, 40% with an apparent increase in the first half of the time period of observation and an apparent decrease in the second

	Results	T2+: O:E ratio, 0.83 (95% CI, 0.80-0.86)	Curves interpreted as showing an incidence decrease for Stage II and an increase for Stage III
	Remarks	The O:E ABC ratio was estimated for the purposes of assessment of self-selection bias	

[a] Year of saturation: the year by which all women in the initial target population were invited at least once.

[b] Indirectly derived or calculated from numbers, tables, and figures in the paper.

Table 2. Ten-point checklist of main methodological problems affecting studies of the effect of mammography screening programmes on the incidence of advanced breast cancer

Point #	Issue	Problem	Consequence	Potentially affected studies (reference number)
1	Follow up time	The time window available to observe a decrease (if any) in ABCR is narrow and closes rapidly. In the Two-County trial, ABCR in the study group began to decrease 4 years after randomization and stabilized at a lower level on the 8th year [2].	The ABCR is expected to increase with the prevalence screening, it may fall in the years immediately following the prevalence screen, and will likely be stable at the end of screening in a cohort of women. In trend and dynamic population analysis, in the absence of an individual time zero (time at entry), the effect is confounded and the effect of screening on ABCR is underestimated. This is particularly applicable to estimates of annual percent change.	8, 12, 13, 19, 34, 37, 41
2	Exposure time	The target population is a dynamic one (but the same holds true for cohort studies). Because there is a latency for the effect of screening on ABCR to take place, at any point in time there are women (i.e., new quinquagenarians, new immigrants, and late attendees) with insufficient exposure time.	The effect of screening on ABCR is underestimated, due to a disproportionate influence of prevalence screens.	All studies
3	Pace of implementation	Public health screening programmes are implemented gradually, in a markedly stepwise fashion, since large populations are divided in distinct administrative units each targeted by an independent local plan of action.	The effect of screening on ABCR is diluted. Until implementation is completed, there are women who are diagnosed with breast cancer before being invited, and who greatly contribute to ABCR.	8, 14, 15, 19, 29, 30, 32, 33, 36-39, 44
4	Prevalence effect	The prevalence screen may be associated with a transient increase in ABCR [13].	During a stepwise implementation of the programme, when the time elapsed from the start is theoretically sufficient to see a decrease in ABCR, this is counteracted by an opposite effect due to newly enrolled women – especially if invitations increase over time.	8, 14, 15, 19, 29, 30, 32, 33, 36-39, 44
5	Reference incidence (i)	The reference (or underlying) incidence rate, with which to compare the rate observed after the introduction of screening, is not known with precision [49].	The rate can be estimated using the rate observed in the last few years before screening, assuming its stability over time, or by linear extrapolation of a pre-existing trend. The second approach is arguably preferable, but both are dependent on underlying assumptions about trends or absence of trends in incidence, and	All studies

			results can vary depending on these assumptions.	
6	Reference incidence (ii)	Whatever incidence rate is being used as a reference, its validity decreases with increasing number of years of observation due to uncontrollable changes (or in the pace of such changes) in the underlying risk of breast cancer.	Assessing the long-term effect of screening on ABCR is subject to considerable uncertainty and there is potential for inaccuracy in either direction (overestimation or underestimation of effect).	8, 12, 13, 19, 34, 37, 41
7	Definition of advanced cancer	There is no agreed definition of advanced breast cancer [50], even though there is general agreement that large or metastatic cancers are 'late stage'.	The definition is chosen based on differing criteria. The pT information alone, which is the most available one, is direct and relatively unaffected by biases due to confounding. Conversely, multiple-stage data are more meaningful, since the effect of screening may differ across different categories of advanced cancers.	All studies
8	Stage migration	The introduction of sentinel lymph node biopsy between mid-1990s and mid-2000s caused a substantial increase in the registered incidence of node-positive breast cancer (stage migration bias) [18].	The use of pN staging is problematic in studies of trends in ABCR over the last two decades, since changes in the risk of node-positive cancer cannot affected by stage migration. The increase in node-positive disease is likely to be population-specific and will depend on the rate of change of local surgical policy. However, reductions in node-positive disease as a results of screening are likely to be underestimated rather than overestimated due to the stage migration.	12-14, 19, 29-43
9	Missing data on tumour stage	Whatever staging system is being used, the introduction of a screening programme tends to bring an improved quality of breast cancer registration, with a sharp decrease in the proportion of unknown-stage cancers.	Because more cases are increasingly placed in all known-stage categories, an apparent increase in all stage-specific rates occurs – including ABCR.	8, 15, 30, 32, 33, 38, 39,
10	Statistical approach	The statistical approach is not standardised, and includes the provision of purely descriptive information and the use of methods which are difficult to interpret, such as joinpoint analysis.	Descriptive information does not allow evaluation of the magnitude and significance of observed changes in ABCR. Methods like the joinpoint analysis are useful for assessing the points in time when ABCR begins to decrease and when it stabilizes, but may be misleading when used to assess the significance of the trend. Also, the important issue is arguably what happened to ABCR following the screening rather than at what point a change occurred in the direction of a trend, which is affected by both confounding and analytic assumptions.	8, 12, 13, 19, 29, 35, 40-43

Figure 1

Click here to download Figure Figure 1 - 3 July 2018.TIF

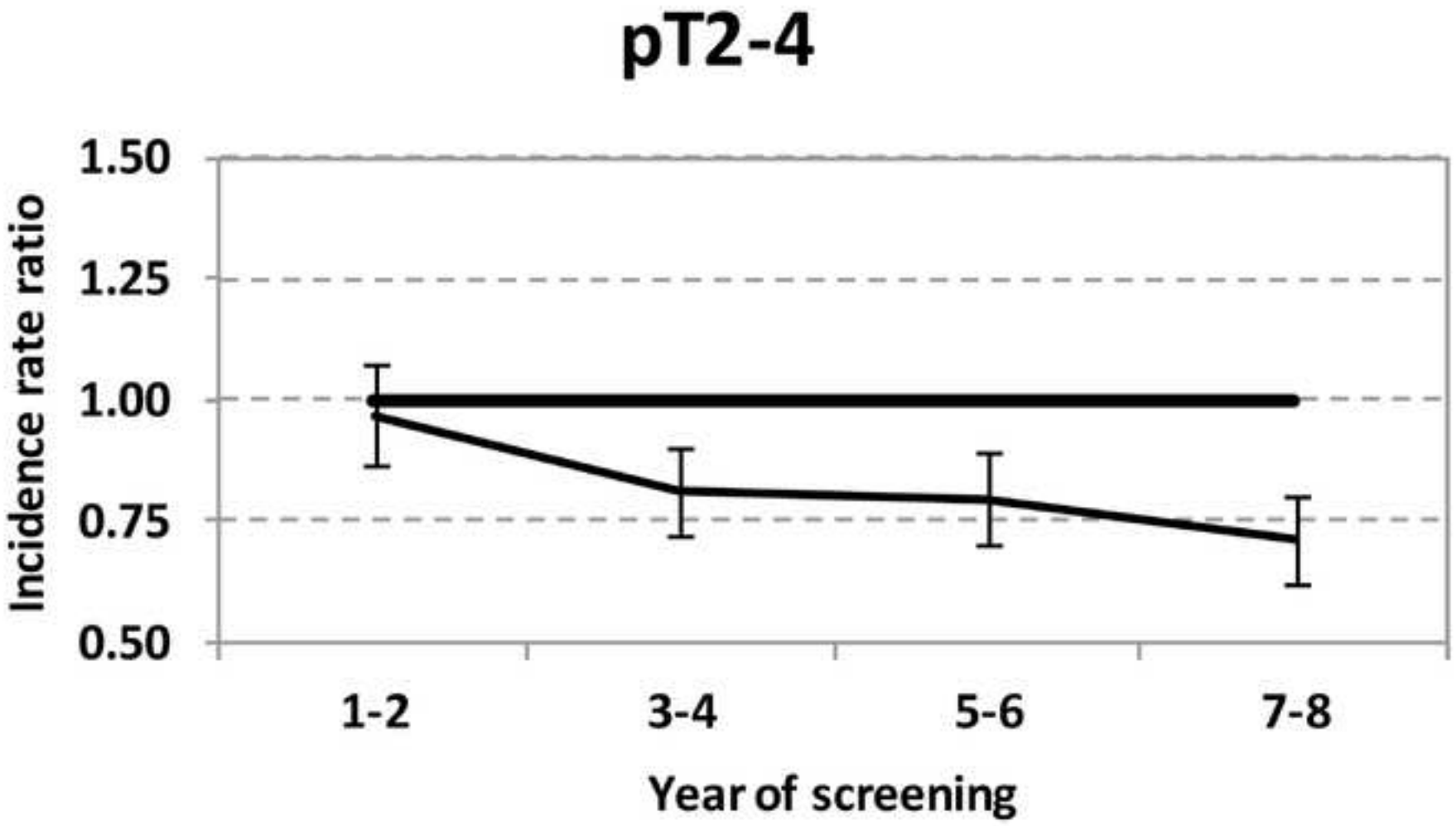
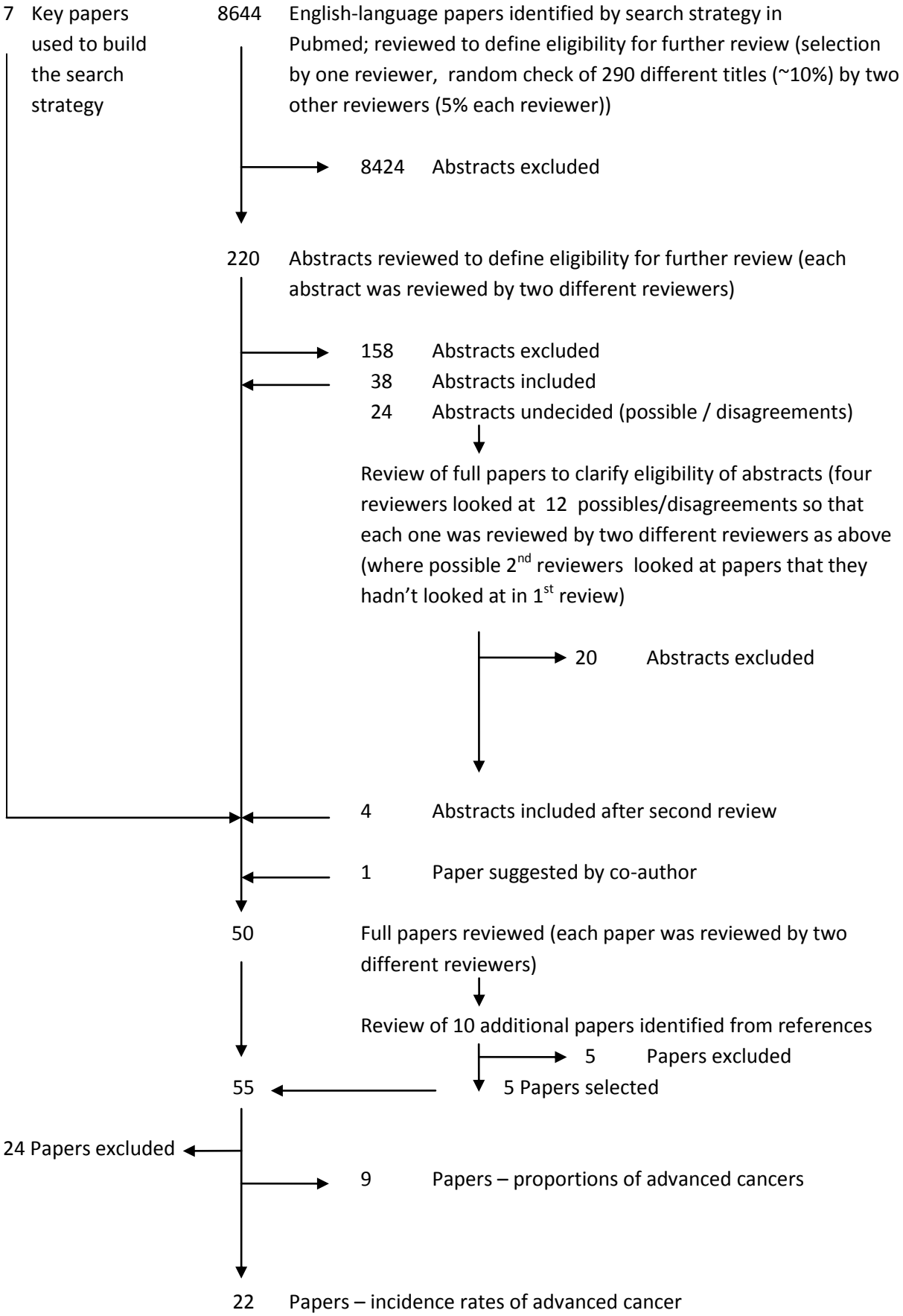


Figure 2

Click here to download Figure Figure 2 - flow chart - 3 July 2018.pdf



Additional file - search strategy