Population-level breast screening: how to salvage the concept?

Ashley K Clift (ORCID)
Department of Surgery and Cancer, Imperial College London, Hammersmith Hospital, London W12 0HS, UK
Corresponding author: Ashley K Clift. Email: ashley.k.clift@gmail.com

Oscar Wilde would have it that all great ideas are dangerous. The 'catch it early, save a life' soundbite does seemingly offer a sensible and logical approach to reducing breast cancer mortality. However, this attractive albeit reductionist aphorism is problematic due to the fact that breast screening harms, and also the questionable power of breast screening to reduce overall mortality.

Evidence from systematic reviews and meta-analyses of randomised trials shows that breast screening does indeed exert a reduction on breast cancer mortality.[1] However, the body of evidence in its totality is fraught with controversy regarding the different methods by which observational and trial data can be interpreted. Here, I argue that the oft-promulgated perspicacity of 'catch it early' is blunt and in its current format harms women unnecessarily. This commentary briefly recapitulates several points of debate regarding the harms posed by breast screening. Thereafter, it anticipates future perspectives on how screening could be improved through modern methods of risk-stratification wherein a diversification of focus must be taken – the way to improve screening is not to focus on detection, detection, detection.

Debate regarding the harms and benefits of screening

Does screening offer additional benefits on top of treatment?

Various study designs and populations have been studied by virtue of 'natural experiments'. These include the staggered roll-out of screening in Denmark, where for 17 years, only 20% of the population were offered screening, forming a large concomitant non-screened control group.[2] Robust, multi-national observational evidence suggests that changes in breast cancer mortality have occurred regardless of the introduction of screening mammography, or irrespective of one attaining requisite maturity for eligibility.[2,3] Downward trends are seen in comparisons of nations that had similar access to treatment but temporally divergent introduction of screening.[3] Combined with advances in treatment strategies seen since trials were conducted, the window in which screening can exert an effect over and above these may well have narrowed.

Overdiagnosis in cancer screening is unavoidable

Data from the most recent report of the European Randomised Screening for Prostate Cancer trial estimated that at the 16-year follow-up, for each 570 men invited to screening, one prostate cancer death would be avoided and 18 cases would be detected.[4] This vast reservoir of non-progressive or very low-risk cellular derangements that is only tapped into by screening is also echoed in cervical screening.[5] Analysis of the National Lung Screening Trial found that any lung cancer detected by screening had an 18.5% (95% CI: 5.4 to 30.6) probability of being overdiagnosed.[6] It is difficult to posit a priori that the extent of overdiagnosis in breast cancer should be significantly different to that seen in other cancers. Indeed, autopsy studies suggest a significant reservoir of ductal carcinoma in situ that may never progress to an invasive, fatal breast cancer.[7]

Methods for studying the phenomenon of overdiagnosis

There is no single uniformly accepted optimal method for assessing overdiagnosis in breast screening. Individual cases of overdiagnosis cannot be determined, but only inferred from observed incidences. Quantification of estimates has used a range of methodologies, from statistical modelling approaches to meta-analysis of trial data. With myriad techniques, it is perhaps unsurprising that estimates of overdiagnosis range between 0 and 54% in published studies.[8] Results from statistical

modelling tend to be towards the lower end of the scale (indeed, those that 'adjust' for lead-time tend to predict <5%), with observational/epidemiological estimates tending to be far higher, depending on which denominator is used, such as the age group in which the incidence/death ratios are calculated. Statistical modelling may be reliant on assumptions that are possibly over-reductionist and may even 'adjust overdiagnosis away' completely.[9]

Mammography screening should increase the incidence of earlier stage breast cancers and reduce the incidence of late stage disease. However, several studies suggest a violation of this assumption. By comparing the huge increase in diagnosis of early cancers and the minimal change in the diagnosis of late stage cancers in the US, only 8 of every 122 excess cancers detected by screening (per 100,000) could be expected to progress, yielding an overdiagnosis rate of 31%.[10] A similar disconnect between the increase in the detection of early breast cancer not being reflected in the decrease of late stage disease has also been seen in studies from the Netherlands, Scandinavia and multi-national studies.[3,11,12] The major point of contention with such studies centres on the assumptions regarding underlying breast cancer incidence trends: it is not possible to definitely calculate what a nation's contemporary breast cancer incidence rate would be without screening and trends in incidence are used. Briefly, if rising incidence trends are the reality, mammography could be shown to have a significant effect, yet if the incidence is stable then the opposite can be deduced.

Divergent 'balance sheets' can be surmised from the same trial data, and this is before one considers that in the many decades that have passed since these trials operated, technology and the expertise of radiologists have advanced, leaving the field with a suboptimal evidence base from which to draw conclusions. The independent review and meta-analysis by Marmot et al. demonstrated that for each breast cancer death avoided with screening, three women are overdiagnosed (rate 11–19%),[1] and in the meta-analyses of the Nordic Cochrane group, 10 women are overdiagnosed per breast cancer death averted.[13] Accepting the Marmot values, a programme in which three women are inappropriately dealt a cancer diagnosis and perhaps pursuant surgery, chemotherapy or radiotherapy for each death averted must be accepted to be imperfect and in dire need of improvement.

What can be done better?

In the absence of a rigorously developed, validated and implementable alternative screening modality, there are several possible avenues in which improved methods may be realised although these are associated with some uncertainties.

Screening strategies influenced by personal risk profiles

Given the indiscriminate bluntness of offering the same screening test across populations where the only precision factor is age, movement towards a function where decision to screen is based on an individual's risk profile is one avenue for minimising harms of those least likely to benefit and maximising benefits to those at highest risk of breast cancer.

This will be reliant on risk prediction models that have high discrimination and calibration demonstrated during derivation and are shown to robustly generalise to different populations. Ideally, these should incorporate demographic, clinical and genetic data to fully capture the manifold determinants of risk, which is yet to be realised. Aside from the highly penetrant archetypal *BRCA* mutations in familial breast cancer, large studies have indeed found myriad other susceptibility variants of low/medium penetrance, although genetic variation may thus far only explain 50% of breast cancer heritability.[14] Furthermore, statistical risk models tend to focus on similar, small sets of clinical variables and the most robust may only explain 29.1% of the variation in time to breast cancer diagnosis.[15] The derivation of multiparametric models has begun[16] but validation studies are needed as are clinical and economic analyses of such models' ramifications on clinical practice. Existing risk prediction models tend not to generalise well, are of moderate quality and may only attain an AUROC of 0.71.[17]

It is unlikely, however, that we will be able to predict incident breast cancer risk with full certainty, and we must be prepared to accept that there is the real possibility of fewer averted breast cancer deaths, as has been shown in some modelling studies.[18]

What is the role of artificial intelligence in breast cancer prediction and diagnosis?

Artificial intelligence and machine learning are often considered as major contenders for the 'next generation' of medical practice, with their skills in making inferences and predictions that physician's brains cannot detect or compute.

A recent newsworthy collaboration led by Google demonstrated an artificial intelligence-based algorithm developed using convolutional neural networks on over 20,000 mammograms from US and UK women.[19] Herein, the algorithm was associated with absolute reductions in false positives and false

negatives of 1.2%–5.7% and 2.7%–9.4%, respectively, as well as an absolute increase in the area under the receiver operating curve of 11.5% as compared to the 'average' radiologist.

The algorithm could predict breast cancer cases with such accuracy within a 2–3-year horizon, which is within the screening interval in the UK. Thus, there is no evidence that this algorithm can impact screening intensity, which is a major shortcoming in 'personalising' screening strategies. The focus of this study was prediction of a diagnosis of breast tumours – this will, by definition, not be able to reduce overdiagnosis, as the algorithm was trained to solve for identification of lesions alone. To reduce the harms from screening, this and similar algorithms must seek to venture towards selectively directing appropriate screening and thus detection of tumours in women that are likeliest to be dangerous and thus require treatment. Other work has integrated imaging data with electronic healthcare data but with a limited range of metrics.[20]

Uncertainties regarding 'artificial intelligence' in breast screening may include the risk of bias in training data, its capacity for changing screening intensity and their ability to make predictions regarding the clinical behaviour/trajectory of detected lesions. Models should be trained on population-representative samples, include a mix of ethnicities and other factors to ensure that they generalise to all women, and the next step should be to diversify the aim from pure detection.

Integrated multiparametric models encompassing clinical, genetic and radiological profiles

There is the possibility to pursue an avenue that builds on the clinical and genetic risk models aforementioned. Quantifying risk in a general practice/community setting or accurately detecting lesions are two steps, but how can imaging data from those that are directed to screening be included?

Risk stratification could be implemented in computing an individual woman's risk pre-diagnosis (either lifetime or nearer horizons), but also on identification of a breast lesion. In the unlikely event that all life-threatening breast cancers can be predicted by a single 'screen vs. don't screen' system, I believe that both aspects of stratification should occur and be implemented into practice. A tantalising hypothesis for reducing overdiagnosis and truly personalising breast cancer care is the stacking/sequencing of algorithms derived using machine learning methods that can not only direct screening to those with appropriately high risk of being adversely affected by breast tumours, but also

decide appropriate treatment strategy or direct modulation of breast cancer risk.

Such algorithms should be assessed in new possible scenarios, such as:

- Diversify the focus from early detection to prevention
  - Multiparametric models may be able to direct personalised lifestyle or pharmacological modifications that can reduce risk of incident breast cancer
  - The concept of breast cancer chemoprevention may be suitable for wider roll-out to women without 'standard' factors such as *BRCA* mutation on the basis of calculated risk score
- Focus on risk of death from breast cancer
  - Models with an explicit focus on directing screening to those with higher mortality risk may aid a reduction in the detection of clinically non-significant lesions
- The possibility of active surveillance of breast changes or lesions deemed as 'low-risk'
  - Models that risk stratify carcinoma in situ or small, indolent tumours in older women may be useful to judiciously direct appropriate treatment

Conclusions

Novel risk stratification and image analysis algorithms have significant potential for 'personalised' screening concepts. However, the focus must be shifted away from purely finding every tumour, as this will not counteract overdiagnosis. Training models to find all pathologically confirmed tumours uses a sub-optimal gold standard, as histology cannot discriminate between indolent and fatal breast cancer. Screening should be made 'smarter' by not just minimising false results but directing expedited diagnosis and treatment to those that have tumours likely to actually cause harm. Such endeavours may comprise calculation of lifetime risk and/or individual lesion risk and will necessitate concerted, multidisciplinary action in terms of dataset linkage, as well as integration of biostatistical, genetic, epidemiological, radiological and health economic expertise.

ORCID iD: Ashley K Clift https://orcid.org/0000-0002-0061-979X

References

1. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012; 380: 1778–1786.

2. Jørgensen KJ, Zahl PH and Gøtzsche PC. Breast cancer mortality in organised mammography screening in Denmark: comparative study. *BMJ* 2010; 340: 797.

3. Autier P, Boniol M, La Vecchia C, Vatten L, Gavin A, Hery C, et al. Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database. *BMJ* 2010; 341: 335.

4. Hugosson J, Roobol MJ, Mansson M, Tammela TLJ, Zappa M, Nelen V, et al. A 16-yr follow-up of the European Randomized study of Screening for Prostate Cancer (figure presented). *Eur Urol* 2019; 76: 43–51.

5. Van Niekerk D, Ogilvie G and Miller D. Rethinking cervical cancer screening. *J Obstet Gynaecol Can* 2012; 34: 411–412.

6. Patz EF, Pinksy P, Gatsonis C, Sicks JD, Kramer BS, Tammenägi MC, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med* 2014; 174: 269–274.

7. Welch HG and Black WC. Using autopsy series to estimate the disease 'reservoir' for ductal carcinoma in situ of the breast: how much more breast cancer can we find? *Ann Intern Med* 1997; 127: 1023–1028.

8. Løberg M, Lousdal ML, Bretthauer M and Kalager M. Benefits and harms of mammography screening. *Breast Cancer Res* 2015; 17: 63.

9. Autier P and Boniol M. Breast cancer screening: evidence of benefit depends on the method used. *BMC Med* 2012; 10: 163.

10. Bleyer A and Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. *N Engl J Med* 2012; 367: 1998–2005.

11. de Glas NA, de Craen AJ, Bastiaannet E, Op' t Land EG, Kiderlen M, van de Water W, et al. Effect of implementation of the mass breast cancer screening programme in older women in the Netherlands: population based study. *BMJ* 2014; g5410.

12. Autier P, Boniol M, Middleton R, Doré J-F, Héry C, Zheng T, et al. Advanced breast cancer incidence following population-based mammographic screening. *Ann Oncol* 2011; 22: 1726–1735.

13. Gøtzsche PC and Jørgensen KJ. Screening for breast cancer with mammography. *Cochr Database Syst Rev* 2013. doi:10.1002/14651858.CD001877.pub5.

14. Skol AD, Sasaki MM and Onel K. The genetics of breast cancer risk in the post-genome era: thoughts on study design to move past BRCA and towards clinical relevance. *Breast Cancer Res* 2016; 18: 99.

15. Hippisley-Cox J and Coupland C. Development and validation of risk prediction algorithms to estimate future risk of common cancers in men and women: prospective cohort study. *BMJ Open* 2015; 5.

16. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med* 2019; 21: 1708–1718.

17. Louro J, Posso M, Hilton Boon M, Román M, Domingo L, Castells X, et al. A systematic review and quality assessment of individualised breast cancer risk prediction models. *Br J Cancer* 2019; 121: 76–85.

18. Pashayan N, Morris S, Gilbert FJ and Pharoah PDP. Cost-effectiveness and benefit-to-harm ratio of risk-stratified screening for breast cancer: a life-table model. *JAMA Oncol* 2018; 4: 1504–1510.

19. McKinney SM, Sienek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020; 577: 89–94.

20. Akselrod-Ballin A, Chorev M, Soshan Y, Spiro A, Hazan A, Melamed R, et al. Predicting breast cancer by applying deep learning to linked health records and mammograms. *Radiology* 2019; 292: 331–342.