Valuing the health states associated with breast cancer screening programmes

Bromley, Hannah; Petrie, Dennis ; Mann, G Bruce ; Nickson, Carolyn; Rea, Daniel; Roberts, Tracy

DOI:
10.1016/j.socscimed.2019.03.028

Valuing the health states associated with breast cancer screening programmes: a systematic review of economic measures

Abstract

Policy decisions regarding breast cancer screening and treatment programmes may be misplaced unless the decision process includes the appropriate utilities and disutilities of mammography screening and its sequelae. The objectives of this study were to critically review how economic evaluations have valued the health states associated with breast cancer screening, and appraise the primary evidence informing health state utility values (cardinal measures of quality of life). A systematic review was conducted up to September 2018 of studies that elicited or used utilities relevant to mammography screening. The methods used to elicit utilities and the quality of the reported values were tabulated and analysed narratively.

40 economic evaluations of breast cancer screening programmes and 10 primary studies measuring utilities for health states associated with mammography were reviewed in full. The economic evaluations made different assumptions about the measures used, duration applied and the sequalae included in each health state. 22 evaluations referenced utilities based on assumptions or used measures that were not methodologically appropriate. There was significant heterogeneity in the utilities generated by the 10 primary studies, including the methods and population used to derive them. No study asked women to explicitly consider the risk of overdiagnosis when valuing the health states described.

Utilities informing breast screening policy are restricted in their ability to reflect the full benefits and harms. Evaluating the true cost-effectiveness of breast cancer screening will remain problematic, unless the methodological challenges associated with valuing the disutilities of screening are adequately addressed.

Keywords: health state utility value, quality of life, QALY, breast cancer, mammography, screening

Introduction

Evidence regarding the cost-effectiveness of healthcare technology is increasingly required to inform the decision on whether to fund and implement new treatment (1). Many decision-making bodies require interventions to be assessed using cost per quality-adjusted-life-years (QALYs) (2), a single summary measure combining life expectancy with individuals' relative preferences for health states in terms of quality of life (3, 4). Health state utility values (HSUVs) are cardinal measures of preference rated on a utility scale anchored from dead (0) to perfect health (1). Utilities can be valued directly or indirectly (5). Direct methods ask individuals to value hypothetical health states, and preferences are directly measured onto the utility scale using the standard gamble (SG), time-trade off (TTO) or visual analogue scale (VAS) (6). The TTO and SG elicit individual choices under uncertainty in life expectancy or risk of death and good health (7), whereas the VAS provides an intermediate valuation of health on a graduated rating scale (8). Indirect methods use a generic multi-attribute utility instrument (MAUI), such as the EuroQol-5 Dimensions (EQ-5D) (9). Current or hypothetical health is mapped onto a generic health instrument and indirectly valued using tariffs for the generic health states that have previously been estimated using direct valuation methods from the general population (10).

Economic evaluations impact health policy decisions and so the methodological quality of the parameters used to inform such analyses must be robust (11). Whilst a growing wealth of literature has explored the importance of the economic approach used, less attention has been given to the methods and quality of the evidence used to inform HSUVs and thus QALYs (12). It is important that the methods used to identify, select and appraise utilities are transparent and systematic to reduce model bias and potential misallocation of resources (13). Several criteria are important for the selection of relevant HSUVs (14). The first relates to the health states, methods, descriptive system and population used to elicit the utilities. Where HSUVs have been measured directly, the validity, reliability and feasibility of the generated values should also be explicitly considered (15). Second, the duration of impact applied must be measured appropriately for both temporary and chronic health states associated with the intervention (16). The third relates to the generalisability of the condition, severity and population characteristics in the utility study to those in the economic evaluation using them (2).

The quality of the HSUVs applied is particularly pertinent in the appraisal of oncological interventions, where quality of life may have greater influence on QALYs than the modest gains in life expectancy (17). Many studies have evaluated the cost-effectiveness of breast cancer screening (18, 19), with those including quality of life in their evaluation reporting fewer net benefits, yet few have commented on the quality of the utility estimates used to inform them. When deciding on preferred screening policy it is critical to be able to accurately value the options available to women of being able to attend routine screening (20). This means valuing all associated benefits and risks associated with the alternative screening policy in terms of utility (21).

Screening for breast cancer in women aged 50 to 74 is recommended because of the ability to capture disease earlier and reduce treatment intensity and disease mortality (22, 23). Decision makers must value the risk that screening would lead to a woman having necessary (and perhaps less intense) treatment at an earlier stage than she would have otherwise had, against the risk of the woman having an unnecessary diagnosis and treatment (24). This valuation is made even more challenging because there is limited evidence on the rate of progression for many breast tumour types (25). If policy makers are to interpret cost-effectiveness analyses of mammography screening and balance the benefits and harms of such interventions appropriately, the utilities used in such evaluations must reflect the health states and effect on those experiencing the sequalae, including overdiagnosis and overtreatment (20).

The challenges associated with valuing health states for breast cancer screening

There are several challenges relating to the identification and assessment of HSUVs for use in the economic evaluation of breast cancer screening programmes specifically. First, the natural history of breast cancer is poorly understood (26, 27). Not all valuation methods for deriving utility may account for the uncertainty in disease progression in the valuation process (28). Second, overdiagnosis and overtreatment from screening create a “paradoxical popularity” because individual women may value unnecessary treatment inappropriately if screening and intervention for benign disease is misconstrued as life-saving (29). Qualitative evidence suggests that both population and patient understanding of overdiagnosis is poor (30), with most perceiving the prognosis of pre-cancerous disease equal to that of an invasive breast cancer (31). Third, the sequelae associated with breast screening last for different durations (32). The long-term implications of a mastectomy are permanent (33), but the anxiety or reassurance associated with mammography screening or diagnostic investigation may only be temporary (34). Temporary health states require modification of conventional valuation methodology and economic evaluations must consider how both temporary and chronic health states are valued simultaneously within a single model (35). Fourth, it is unclear whose preferences would be best placed to assess the benefits and harms of breast screening (36). The National Institute for Healthcare and Clinical Excellence (NICE) advocate the use of general population preferences in a publicly funded healthcare system, (2) yet given the complexity involved in valuing screening it may be difficult for the lay person to quantify using conventional utility instruments. The preferences and disease characteristics of individual women and breast cancers also vary significantly by demographic (37) and so the generalisability of the population in the primary and economic studies may influence the generated QALYs (38). Such challenges may impact utility instruments, therefore an assessment of the methodology used to overcome these issues is critical in the appraisal of appropriate HSUVs.

Objectives

The objectives of this study were to critically appraise and assess how economic evaluations have captured the health states and utilities associated with mammography screening. Primary studies that have measured HSUVs for relevant health states were also evaluated to examine the quality of the evidence informing cost-effectiveness studies of breast cancer screening and its sequalae.

Methods

The review followed the UK Centre for Review and Dissemination (39) guidelines and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (40).

Eligibility Criteria

The systematic review included studies published in English whereby utilities were either used or elicited for health states relating to mammography screening specifically. Studies were included if they met the following criteria: the participants were women in the general population at risk of breast cancer, the intervention (for economic evaluations) was population mammography screening, the comparator was no screening and the outcomes were cost per QALY (for economic evaluations) or the measurement of health state utility values associated with breast cancer screening and its sequalae. Studies were excluded if they were reviews, editorials, interventions to improve screening participation or breast screening programmes using technologies other than routine mammography (e.g. ultrasound or magnetic resonance imaging), which are not routinely used for screening the general population.

Search strategy

Eleven electronic databases were searched for studies published up to 1 September 2018: MEDLINE, EMBASE, PsycInfo, CINAHL, Econlit, Social Citation Index, Social Sciences Citation Index, Cochrane library, NHS Economic Evaluation Database, Database of Abstracts of Reviews and Effects, and Health Technology Assessment. The reference lists of relevant studies were hand-searched to identify any further relevant studies for inclusion.

The search strategy was developed using the terms published in other systematic reviews of breast cancer screening programmes (20, 41) and Cochrane review guidelines (39). Both Medical Subject Heading and keyword searches were used relating to the term ‘mammography’, ‘breast cancer’, ‘screening’, ‘overdiagnosis’, ‘economic evaluation’ and ‘utility’, with truncation used where appropriate (see Supplementary File). There was no restriction placed on publication year to ensure all relevant studies to date were included in the review.

Study selection

Information retrieved by the database search was managed via Endnote referencing software (42). A two-stage process (43) was used to identify relevant studies for inclusion in the final review. In the first stage, the title and abstract of retrieved studies were checked against the pre-specified eligibility criteria. Relevant studies or those where a decision could not be made based on the title and abstract proceeded to the second stage. In stage two, the full text was assessed for relevance and the reference lists of key articles were hand-searched to identify other potentially relevant studies. Studies citing or reporting utilities which met the required eligibility criteria were included in the final review. A second reviewer screened and checked a sub-sample of studies to negate any selection bias (44).

Data extraction and analysis

An electronic template was used to extract data on the characteristics of the included studies. For each included study, one reviewer extracted data about the study characteristics, the health states and utilities reported and the methodology, population, instruments and duration for which utilities were applied. The data were tabulated and analysed by narrative description as the retrieved HSUVs were too heterogeneous to usefully combine in a meta-analysis. A formal quality appraisal was not performed as there is no agreed quality assessment checklist for assessing studies of this nature (11). However, the methods suggested by Brazier (36) and Papaioannou (11) for the systematic identification, selection and assessment of HSUVs from the literature were used to assess the validity, reliability and robustness of the identified utilities to inform the narrative review (28, 45).

Results

The database search retrieved 9,447 studies, of which 3,562 were removed as duplicates. A further 3 studies were identified through hand-searching the reference lists of relevant studies. A flow diagram of the studies selected or excluded at each stage, with reasons, is provided in Figure 1. Data were extracted for the 50 relevant studies included in the narrative review: 40 economic evaluations using cost per QALYs and 10 primary studies that measured utilities for health states associated with mammography screening.

Economic evaluations using Cost per QALY

Table 1 summarises the characteristics of the 40 economic evaluations using QALYs in their analysis of breast screening programmes. Most evaluated alternative breast screening strategies (46-67), although five studies (68-72) only presented results using cost per QALY in the sensitivity analysis because of the uncertainty around published HSUVs for mammography. Seven studies (68, 70, 72-76) explored the cost-effectiveness of screening elderly women, whereas two studies (77, 78) evaluated extending the lower age limit of screening. Four studies (71, 79-81) assessed the benefits of risk stratified mammography screening and one study (82) appraised opportunistic versus organised mammography screening. The remaining three studies (29, 83, 84) evaluated the benefits and harms of breast cancer screening but reported QALYs without costs in their main analysis.

The utilities associated with breast cancer screening are difficult to compare because each study made different assumptions about the value used, the duration over which they were applied and the sequalae included in each of the health states. The values used for screening attendance varied significantly (0.100-0.994) and were applied for a duration of between 2 hours and 7 days. There were similar issues with heterogeneity between HSUVs for diagnosis (0.100-0.895) and treatment (0.100-0.990). Utilities were applied for between 5 days and 6 months for a positive mammogram and a duration of one month to the rest of the woman's life for treatment depending on classification by intervention or disease stage. The duration of utilities or disutilities when applied in economic models can be a key driver in influencing results using QALYs (85), yet few studies justified the duration enforced (29, 80, 81) or considered whether the utilities for temporary health states had used an appropriate chaining adaptation (86).

Most economic evaluations used the same two sources (32, 62) for their utilities, although there was variation in the actual value used and the generalisability of the population for which they were applied. The first of these sources by Stout et al. (62) applied tariffs based on assumptions for screening, diagnosis and treatment of breast cancer by stage at diagnosis to adjust US healthy population EQ-5D estimates (87). Although the generated utilities were deemed consistent with those reported in other studies (17, 88), it is not clear how the assumed adjustment for each health state was determined, and yet nine economic evaluations (29, 47, 49, 63-65, 78, 83, 84) applied this method. For screening disutility, almost half of the economic evaluations (29, 48, 51, 54, 56, 58, 60, 63, 65, 66, 68, 73, 74, 78, 84) used expert VAS utilities derived from a second study in the Netherlands (32), but only three economic evaluations (51, 54, 73) considered the generalisability of the expert sample to the general population in the model to which this was applied. Other economic evaluations made their own adjustments to local population EQ-5D or SF-6D data (47, 49, 50, 59, 69, 79, 81) or used HSUVs elicited directly (VAS, TTO, SG) from samples of women with comparable demographics to try and improve the relevance of the applied utilities to the population in their economic model (52, 73, 77). The remaining evaluations cited utilities from another economic model (53, 67), systematic review (46) or made their own assumption of an appropriate value (57, 70-72) but did not provide a detailed critique of how these were derived.

Sensitivity analyses were used to analyse the uncertainty around HSUVs in the majority of the 40 economic evaluations, with at least half reporting quality of life as having a significant effect on cost-effectiveness results. However, not all economic evaluations included all relevant phases of the mammography screening pathway in their analysis and therefore implicitly assumed they had no impact on quality of life. 22 studies (29, 46, 47, 52, 53, 55-59, 64, 67-69, 71, 72, 75, 77, 79-81) did not integrate the potential reassurance or disutility of screening anxiety and diagnostic follow-up in their analyses and a further 27 did not explicitly capture the disutility relating to the risk of overdiagnosis (29, 46, 48, 49, 51-59, 61, 62, 64, 67-71, 73-77). Consequently, no utility loss was applied to reflect this uncertainty in more than half of the economic evaluations, which may bias results (QALYs) toward more frequent screening (29, 49, 52, 54, 59). This limitation was justified in five studies due to the lack of robust HSUVs for mammography screening. For the 13 studies (47, 50, 60, 63, 65, 66, 78-84) which did attempt to value overdiagnosis in their analysis, an assumption was made that this was captured in the QALYs across screening strategies by including the temporary disutility of diagnosis and treatment without a corresponding gain in life years. However, the utilities applied used sources which had not highlighted that there was a risk the treatment was unnecessary during the valuation process and therefore is unlikely to fully capture the impact of the risk of overdiagnosis on quality of life.

Primary studies

Ten primary studies (15, 16, 32, 34, 89-94) valued utilities for health states relevant to mammography screening. A summary of the study characteristics and methodology is shown in Table 2. The studies' aims were diverse and measured utilities for a range of relevant health states, including: screening attendance and anxiety, mammography result (true positive, false positive, true negative and false negative), diagnostic investigation of a positive mammogram, treatment of a screen detected breast cancer, breast cancer recurrence and terminal care. The risk of overdiagnosis was not valued independently or explicitly captured within the descriptions of treatment health states in any of the primary studies of breast screening.

Methodology

The main method for valuing health states include direct and indirect empirical measurement or expert opinion (95). Multiple approaches were taken to elicit utilities for breast screening health states identified by this review, with more than four different valuation techniques reported within the primary studies. These four primary approaches included the VAS (96), which was anchored from worst to best imaginable health, standard gamble which compared the health state against a gamble of death and perfect health (97), time trade-off which trades years lived in full health against living longer in the health state being valued (98) and EQ-5D which asked trial participants to report their own or hypothetical health on a generic scale and applied general population tariffs to estimate final utility scores (99). More than one technique was used to value screening health states in six studies (15, 16, 34, 90, 93, 94), whilst the remainder used a single technique (32, 89, 91, 92). The standard gamble, initially presented by Neumann and Morgenstern (100) in (101) is the gold standard method for valuing conditions of uncertainty (102), yet only two studies (90, 93) used this technique to capture the potential benefit and risks associated with screening. An alternative choice-based method (TTO) was justified by five studies to reduce cognitive burden associated with the standard gamble (15, 16, 90-92). De Haes et al. (32) did not use a choice-based method but mapped visual analogue scale scores into utilities using a power function (VAS): TTO = 1-(1-VAS)^1.*82*. (103), although there are reported issues with the reliability of conversion formulas (104). A combination of both direct and indirect methods was used by the remaining studies (34, 94) using tariffs from the US (87) and Dutch general population (105) for the EQ-5D descriptive instrument before and after screening. The Short Form-36 questionnaire was also used by Rijnsburger (94), but the values were never mapped into SF-6D utilities (106). Only half of the studies considered whether the chosen method was appropriate for overcoming the methodological challenges associated with screening health states (15, 16, 32, 91, 92).

Duration

Traditional methods such as the standard gamble, TTO and VAS are targeted towards chronic health states (5, 6, 102). For valuing temporary health states, a two-stage technique known as ‘cascading’ or ‘chaining’ is recommended and can be applied to modify the traditional TTO or SG approach (107). For chaining, the worst temporary health state is known as the anchor health-state because it is used as the lower anchor instead of dead (35). The anchor state is subsequently valued against full health and dead to realign values onto the traditional utility scale (6). Only two studies (15, 16) used a chaining adaptation of the conventional TTO to appropriately value temporary health states for screening attendance and diagnostic investigation.

A combination of direct and indirect assumptions (108) were used to specify duration in the remaining studies. Four studies (90-93) specified a single duration of impact for both temporary and chronic health states and applied the same method (TTO or SG) to ensure consistency. The same technique (VAS) was used in two studies (32, 89) to specify the time within the vignettes, although the durations applied varied depending on the timeframe assumed. Other studies (34, 94) did not specify the health state duration per se but indirectly measured utility at discrete time points during the screening process. However, due to variation in follow-up time some women were aware of their results a priori which may have inadvertently biased results.

Descriptive system

The validity of the health state and utility elicited is dependent on the accuracy of the vignette and should be informed by a thorough review of the literature or input from those well acquainted with the condition (104, 109). HSUVs were generated using health state descriptions in eight (80%) of the primary studies (15, 16, 32, 89-93). Although the vignettes in all eight studies were informed by clinical guidelines and expert input, only five studies (15, 16, 32, 89, 92) validated the clinical scenarios through patient piloting or focus group discussion. Similarly, the framing and labelling of health descriptions can systematically bias choices and perceived quality of life due to the negative connotations associated with cancer and dying (110, 111), yet only two studies (91, 92) explicitly considered the impact of this on their results. The remaining two studies (34, 94) did not use vignettes but indirectly measured the disutility associated with screening by asking women enrolled in a clinical trial of tailored mammography to value their own health ex-ante and ex-post screening using validated health instruments (EQ-5D). Interestingly, both studies commented on the limitations of the sensitivity of the EQ-5D domains in capturing changes in utility for the short-term duration of screening.

No primary study explicitly considered the impact of the risk of overdiagnosis or unnecessary treatment in any of the health states described. Only Gerard (91) and Hall (92) introduced the notion of dying of causes other than breast cancer in their vignettes, although they did not explicitly include the risk of unnecessary follow up and treatment. Kim et al. (93) explicitly included risk in their health state descriptions of surgery and radiotherapy but only provided estimates for recurrence and survival, assuming all treatment was necessary for non-invasive disease.

Population

Health states relevant to breast screening can be valued by three populations groups; the general population, patients and clinical experts (36). Seven (15, 16, 34, 89, 91, 93, 94) of the ten primary studies used general population values, which are preferred by most publicly funded healthcare departments (2, 99), although there was some selection bias toward women of breast screening age. One study (92) collected a mixed sample of public and patient preferences and reported significant differences between the HSUVs measured by those with and without experience of breast cancer. Patient preferences are typically higher than those elicited from the public due to adaptation to the condition or a feeling of necessary intervention (36, 112), but Hall (92) justified their approach as they felt patients were best placed to value the complex side-effects associated with breast surgery. The remaining two studies (32, 90) used an expert sample to overcome the cognitive difficulties experienced in their feasibility piloting of TTO health states.

Quality assessment

Most studies did not explicitly comment on the quality of the reported HSUVs in terms of the validity, reliability and feasibility of the methods used. Among the four studies (15, 16, 32, 91) that reported on reliability, four assessed ranking order and only one (15) tested test-retest consistency. None of the primary studies commented on the time taken to complete the task, although this is routinely recommended for appraising participant comprehensibility (28, 97). At least half of the authors commented on comprehensibility issues relating to the SG and TTO techniques, although only one study (15) provided quantitative evidence to measure the reported cognitive burden using a Likert scale. Whilst most studies justified the VAS based on task acceptability, only three studies (15, 16, 32) considered the theoretical validity of this approach in capturing the temporary or uncertain benefits and risks associated with breast screening specifically.

Discussion

Principal findings

Population based mammography screening for breast cancer is a major public health investment and significant time investment for women and therefore warrants rigorous scrutiny (20). This systematic review provides the first synthesis of economic measures and health states used to value mammography screening, explicitly including overdiagnosis, and summarises the evidence base informing the population screening debate. The identified evaluations found that quality of life had a significant effect on cost-effectiveness results in sensitivity analyses . Determining whether the associated benefits and harms have been captured appropriately is therefore not only of clinical importance, but may impact how screening policy is determined or overdiagnosis is conceptualised (20).

Deciding how breast screening utilities should be captured is fraught with challenges (24, 41). There is no consensus on the most appropriate economic measure and population to use when valuing outcomes in cancer screening programmes. Half of the identified studies in this review used the same two sources to value quality of life (32, 62), but the remainder used values that were based on assumption, used out of context or were not methodologically sound. Unlike prostate and cervical cancer, the natural history of in situ breast disease is not well understood (26), yet the way in which the utilities were assigned to represent the associated health states for screening and its sequalae were not described in detail in any of the studies. Balancing the availability and quality of published HSUVs to inform economic evaluations can be problematic where primary evidence is limited (7), but it is imperative that such limitations are made explicit so that decision makers may consider the implications upon cost-effectiveness results (113).

The heterogeneity in utility values raises the question of what economic measure should be used, or whether health related quality of life is suitable for measuring outcomes associated with screening and overdiagnosis. The commonest approach used in the empirical studies was the VAS, despite this technique being considered methodologically inferior to other choice-based techniques (114, 115). Ideally, the measure chosen should reflect the underlying decision within the valuation process, in line with traditional axioms of utility theory (116). When trading length of life against quality, TTO is more appropriate (86), whereas in a situation in which there is also risk (such as screening and treatment uncertainty), the standard gamble may be more suitable (97). A systematic review of metastatic breast cancer utilities (41) found that the SG was the most frequently used technique for capturing uncertainty in survival and preferred for valuing risk-based utility (112), although there are concerns it may inappropriately conflate health with risk aversion (124). Conversely, preference-based instruments (EQ-5D) are considered the method of choice by NICE (117), but it is unclear whether an indirect approach would be sufficiently sensitive to detect minor changes in utility (34) or reflect the true risks involved, unless respondents are adequately informed about the benefits and harms during the valuation process or vignette. With the majority of identified studies using HSUVs based on author assumption, new empirical evidence to reliably inform such analyses is clearly necessary.

The clinical outcomes associated with breast cancer screening programmes are widely contested, yet the benefits and harms of mammography are inadequately appraised in the economic literature informing the debate. Few studies identified by this review integrated all relevant phases of care associated with breast cancer screening into the assessment of quality of life, and the values used were limited in their ability to truly capture the disutility Thirteen studies included overdiagnosis in their evaluation but applied the same utilities for diagnosis and treatment as a non-overdiagnosed cancer(118-120), even if the costs and quality of life losses were ultimately not necessary or entirely representative. Estimates of screen detected overdiagnosis vary significantly from 0 to 54% (23, 27). Whilst several economic evaluations cite the lack of published utilities as a justification for not including screening or overdiagnosis in their analysis (29, 49, 52, 54), ignoring this harm may inadvertently lead to inappropriate advice to women, decisions on the value of screening programmes and potential misallocation of resources. Similarly, none of the primary studies explicitly considered the impact of unnecessary treatment in their vignettes. The inclusion of overdiagnosis in qualitative descriptions has been shown to change general population preferences toward more conservative management or surveillance strategies(121, 122). The limitations of the economic measures and health states outlined in this review (123) raises concerns about information asymmetry, and whether women can make an informed decision about screening without information on the full benefits and risks. Any potential advantages and risks should be explicitly listed within the descriptions of relevant health states. These findings are not limited to breast cancer; appraising the impact of unnecessary treatment may be relevant in other public interventions such as prostate cancer (120), cervical screening (124) or the management of cardiovascular disease, where treatments reduce the risk of future morbidity and mortality but have side-effects (125). Indeed, a number of cancer screening initiatives have reported varying outcomes when different sets of utilities are assumed (119, 126).

The literature is similarly heterogeneous in the duration and methods used to apply HSUVs. A difficulty in valuing screening interventions is that the process encompasses both temporary and chronic health states (16). The intensity and duration of the utilities associated with screening (15) and diagnostic anxiety (127, 128) vary significantly to the long-term sequelae associated with treatment (129), depending on whether this is classified by intervention or disease stage. There is ongoing debate about how best to overcome such issues (130), including the adaptation of conventional direct approaches (112, 131) or clinical guidelines on the duration of impact for each of the health states (20). Whether such adjustments are practical for screening interventions is debated and there are limitations of QALYs in interventions such as breast screening which may only have a transient impact on utility yet may be highly valued. Thorough sensitivity analysis of the durations applied to QALYs should be undertaken in any economic evaluation of population mammography screening to ascertain the effect of key drivers on cost-effectiveness (35, 38).

Two systematic reviews (19, 132) have previously explored the outcomes of economic evaluations relating to breast screening programmes. Schiller-Fruhwirth et al. (19) reported on the lack of breast screening specific utilities and insufficient reporting of validation in their review of economic models.

A second review (132) reported similar findings relating to a paucity of methodologically appropriate utilities relevant to mammography screening. Other systematic reviews (17, 41) of economic outcomes in the broader breast cancer literature have been equally unable to combine screening values in meta-analyses due to insufficient numbers and inconsistencies in the approach and population used to derive them.

Strengths and limitations

The value of this review is that it provides a critical appraisal of the HSUVs used in economic evaluations of breast screening programmes, alongside a wider appreciation of the methodological issues and challenges associated with the empirical valuation of mammography and its sequelae. It offers new insight into the methodological issues informing the screening and overdiagnosis debate, and recommendations on where to direct future research to improve the appraisal of population screening services. Nonetheless, this review also has limitations. Some studies were not explicit in stating that the condition under study was relevant to mammography screening. Therefore, a subjective judgment had to be made by the reviewers about the health states measured and their relevance for inclusion. Second, the review only included studies published in English and may have excluded relevant HSUVs in other publications. Finally, a summary statistic for the health states associated with mammography screening could not be determined due to the heterogeneity between studies and the methods used to derive reported HSUVs.

Implications

Utilities informing breast screening policy are restricted in their ability to reflect the full benefits and harms. Primary health state estimation, incorporating the potential benefits and risks in the valuation process, should be pursued to provide methodologically robust empirical data for the economic appraisal of mammography screening policy. To exclude such harm from the evaluation process is negligent. . As screening evolves in line with technological advancements and improvements in genetic understanding (future risk), quality of life values should also be adjusted. Similarly, as the identification of low risk disease from screening becomes more prevalent (25), it is likely that more personalised, risk-stratified utilities for active monitoring strategies will be required in breast cancer screening models.

Recommendations for future research

The economic evaluation of mammography screening remains problematic due to uncertainties in the natural history of the disease, duration of sequalae and risk of potential unnecessary treatment. The following methodological recommendations are highlighted for researchers planning economic evaluations of population breast cancer screening:

- Economic evaluations should explicitly include all relevant utilities and disutilities associated with mammography screening and its sequelae. Overdiagnosis should be explicitly captured in the evaluation of population screening policy, alongside extensive uncertainty analysis where there is debate on the extent of unnecessary treatment.
- New empirical evidence based on adequately informed utility data is needed to inform breast cancer screening decisions. The findings suggest the standard gamble and EQ-5D as the most appropriate economic measures to value screening health states, but vignettes should explicitly describe the advantages and risks of screening during the valuation process.
- Groups at high or low risk for breast cancer should be considered in sub-group analysis, and quality of life values risk-stratified accordingly. It is likely that the management and prognosis for ductal carcinoma in situ will have markedly different implications than the disutilities associated with high risk, invasive disease.
- Consistency in the duration for which the penalties are applied to screening, diagnosis and treatment related health states should be standardised by a panel of experts, clinicians and patients to prevent study heterogeneity driving cost-effectiveness results.
- Breast cancer screening evaluations assume perfect compliance with treatment which may not be reflective of clinical practice. The utility of active surveillance or non-invasive management, included in other population cancer screening evaluations, may be adopted by some women with low risk disease and should be considered in the breast cancer setting.

Conclusion

Breast cancer screening programmes are deemed cost-effective for women aged 50-74 in the general population. Nonetheless, the evidence informing breast cancer screening policy have several limitations that must be addressed to determine what would be the most cost-effective approach. This review highlights the methodological challenges associated with valuing the utilities and disutilities associated with breast cancer screening, and suggests economic measures are unlikely to adequately capture the outcomes of screening in terms of quality of life.

There is no single recommended approach for valuing the health states associated with breast cancer screening and its sequalae, but women should be properly informed about the benefits and risks during the valuation process or vignettes. Overdiagnosis is not appropriately accounted for in the appraisal of mammography screening and undervaluation may lead to inappropriate decisions on the value of screening programmes. The measurement of health state utility values derived from adequately informed individuals, as well as sub-group analysis by risk group, is necessary if the debate on population screening programmes is to be adequately addressed.

References

1. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes: Oxford university press; 2015.

2. Excellence NIfHaC. Guide to the methods of technology appraisal. London: NICE; 2013.

3. Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. Annu Rev Public Health. 2000;21:587-611.

4. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. British medical bulletin. 2010;96(1):5-21.

5. Tolley K. What are health utilities. Hayward Medical Communications, London. 2009.

6. Torrance GW. Measurement of health state utilities for economic appraisal: a review. Journal of health economics. 1986;5(1):1-30.

7. Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. Journal of health economics. 1996;15(2):209-31.

8. Bass EB, Steinberg EP, Pitt HA, Griffiths RI, Lillemoe KD, Saba GP, et al. Comparison of the rating scale and the standard gamble in measuring patient preferences for outcomes of gallstone disease. Medical Decision Making. 1994;14(4):307-14.

9. Dolan P. Modeling valuations for EuroQol health states. Medical care. 1997:1095-108.

10. Kind P, Hardman G, Macran S. UK population norms for EQ-5D. 1999.

11. Papaioannou D, Brazier J, Paisley S. Systematic Searching and Selection of Health State Utility Values from the Literature. Value in Health. 2013;16(4):686-95.

12. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. Value in Health. 2010;13(5):509-18.

13. Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group–6. Medical decision making. 2012;32(5):722-32.

14. Jackson LJ, Auguste P, Low N, Roberts TE. Valuing the health states associated with Chlamydia trachomatis infections and their sequelae: a systematic review of economic evaluations and primary studies. Value in Health. 2014;17(1):116-30.

15. Gerard K, Johnston K, Brown J. The role of a pre□scored multi□attribute health classification measure in validating condition□specific health state descriptions. Health economics. 1999;8(8):685-99.

16. Johnston K, Brown J, Gerard K, O'Hanlon M, Morton A. Valuing temporary and chronic health states associated with breast screening. Social Science & Medicine. 1998;47(2):213-22.

17. Earle C, Chapman R, Baker C, Bell C, Stone P, Sandberg E, et al. Systematic overview of cost-utility assessments in oncology. Journal of Clinical Oncology. 2000;18(18):3302-17.

18. Brown ML, Fintor L. Cost-effectiveness of breast cancer screening: preliminary results of a systematic review of the literature. Breast cancer research and treatment. 1993;25(2):113-8.

19. Schiller-Fruhwirth IC, Jahn B, Arvandi M, Siebert U. Cost-Effectiveness Models in Breast Cancer Screening in the General Population: A Systematic Review. Applied Health Economics & Health Policy. 2017;15(3):333-51.

20. Marmot MG, Altman D, Cameron D, Dewar J, Thompson S, Wilcox M. The benefits and harms of breast cancer screening: an independent review. British journal of cancer. 2013;108(11):2205.

21. Myers ER, Moorman P, Gierisch JM, Havrilesky LJ, Grimm LJ, Ghate S, et al. Benefits and harms of breast cancer screening: a systematic review. Jama. 2015;314(15):1615-34.

22. Løberg M, Lousdal ML, Bretthauer M, Kalager M. Benefits and harms of mammography screening. Breast Cancer Research : BCR. 2015;17(1):63.

23. Nelson HD, Fu R, Cantor A, Pappas M, Daeges M, Humphrey L. Effectiveness of breast cancer screening: systematic review and meta-analysis to update the 2009 US Preventive Services Task Force Recommendation. Annals of internal medicine. 2016;164(4):244-55.

24. Welch HG, Passow HJ. Quantifying the benefits and harms of screening mammography. JAMA internal medicine. 2014;174(3):448-54.

25. Francis A, Thomas J, Fallowfield L, Wallis M, Bartlett JM, Brookes C, et al. Addressing overtreatment of screen detected DCIS; the LORIS trial. European journal of cancer. 2015;51(16):2296-303.

26. Groen EJ, Elshof LE, Visser LL, Emiel JT, Winter-Warnars HA, Lips EH, et al. Finding the balance between over-and under-treatment of ductal carcinoma in situ (DCIS). The Breast. 2017;31:274-83.

27. Puliti D, Duffy SW, Miccinesi G, De Koning H, Lynge E, Zappa M, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. Journal of medical screening. 2012;19(1_suppl):42-56.

28. Badia X, Herdman M, Dipstat MR, Ohinmaa A. Feasibility and validity of the VAS and TTO for eliciting general population values for temporary health states: a comparative study. Health Services and Outcomes Research Methodology. 2001;2(1):51-65.

29. Raftery J, Chorozoglou M. Possible net harms of breast cancer screening: updated modelling of Forrest report. Bmj. 2011;343:d7627.

30. Hersch J, Jansen J, Barratt A, Irwig L, Houssami N, Howard K, et al. Women’s views on overdiagnosis in breast cancer screening: a qualitative study. Bmj. 2013;346:f158.

31. Partridge A, Adloff K, Blood E, Dees EC, Kaelin C, Golshan M, et al. Risk perceptions and psychosocial outcomes of women with ductal carcinoma in situ: longitudinal results from a cohort study. Journal of the National Cancer Institute. 2008;100(4):243-51.

32. De Haes J, de Koning HJ, van Oortmarssen GJ, van Agt HM, de Bruyn AE, van der Maas PJ. The impact of a breast cancer screening programme on quality‐adjusted life‐years. International journal of cancer. 1991;49(4):538-44.

33. Weitzner MA, Meyers CA, Stuebing KK, Saleeba AK. Relationship between quality of life and mood in long-term survivors of breast cancer treated with mastectomy. Support Care Cancer. 1997;5(3):241-8.

34. Tosteson AN, Fryback DG, Hammond CS, Hanna LG, Grove MR, Brown M, et al. Consequences of false-positive screening mammograms. JAMA internal medicine. 2014;174(6):954-61.

35. Wright DR, Wittenberg E, Swan JS, Miksad RA, Prosser LA. Methods for measuring temporary health states for cost-utility analyses. Pharmacoeconomics. 2009;27(9):713-23.

36. Brazier JE, Dixon S, Ratcliffe J. The role of patient preferences in cost-effectiveness analysis. Pharmacoeconomics. 2009;27(9):705-12.

37. Smith RA. Expert Group: IARC Handbooks of Cancer Prevention. Vol. 7: Breast Cancer Screening. Lyon, France: IARC; 2002. 248pp. Breast Cancer Research. 2003;5(4):216.

38. Brazier J, Deverill M. A checklist for judging preference‐based measures of health related quality of life: learning from psychometrics. Health economics. 1999;8(1):41-51.

39. Dissemination C. CRD's Guidance for Undertaking Reviews in Health Care. Layerthorpe: University of York, Centre for Reviews and Dissemination. 2009.

40. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS medicine. 2009;6(7):e1000100.

41. Peasgood T, Ward SE, Brazier J. Health-state utility values in breast cancer. Expert Rev Pharmacoecon Outcomes Res. 2010;10(5):553-66.

42. Endnote X. Referencing software. In: Analytics C, editor. Boston, MA 2018.

43. Higgins JP, Altman DG. Assessing risk of bias in included studies. Cochrane handbook for systematic reviews of interventions: Cochrane book series. 2008:187-241.

44. McDonagh M, Peterson K, Raina P, Chang S, Shekelle P. Avoiding bias in selecting studies. 2013.

45. Ogwulu CB, Jackson LJ, Kinghorn P, Roberts TE. A systematic review of the techniques used to value temporary health states. Value in Health. 2017;20(8):1180-97.

46. Ahern CH, Shen Y. Cost-effectiveness analysis of mammography and clinical breast examination strategies: a comparison with current guidelines. Cancer Epidemiology and Prevention Biomarkers. 2009;18(3):718-25.

47. Arrospide A, Rue M, van Ravesteyn NT, Comas M, Soto-Gordoa M, Sarriugarte G, et al. Economic evaluation of the breast cancer screening programme in the Basque Country: retrospective cost-effectiveness and budget impact analysis. BMC cancer. 2016;16(1):344.

48. Beemsterboer PM, De Koning HJ, Warmerdam PG, Boer R, Swart E, Dierks ML, et al. Prediction of the effects and costs of breast‐cancer screening in Germany. International journal of cancer. 1994;58(5):623-8.

49. Carles M, Vilaprinyo E, Cots F, Gregori A, Pla R, Román R, et al. Cost-effectiveness of early detection of breast cancer in Catalonia (Spain). Bmc Cancer. 2011;11(1):192.

50. Christensen MK, Niclasen BV, Iburg KM. Societal costs and effects of implementing population-based mammography screening in Greenland. International Journal of Circumpolar Health. 2017;76(1):1373580.

51. de Koning HJ, Martin van Ineveld B, van Oortmarssen GJ, De Haes J, Collette HJ, Hendriks JH, et al. Breast cancer screening and cost‐effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. International Journal of Cancer. 1991;49(4):531-7.

52. Forrest AP. Breast cancer screening : report to the Health Ministers of England, Wales, Scotland & Northern Ireland. In: Forrest P, editor. London :: H.M.S.O.; 1986.

53. Haghighat S, Akbari ME, Yavari P, Javanbakht M, Ghaffari S. Cost-effectiveness of three rounds of mammography breast cancer screening in Iranian women. Iranian journal of cancer prevention. 2016;9(1).

54. Hristova L, Hakama M. Effect of screening for cancer in the Nordic countries on deaths, cost and quality of life up to the year 2017. Acta Oncologica (Stockholm, Sweden). 1997;36:1-60.

55. BreastScreen Australia Evaluation. Economic evaluation and modelling study. Australia: Australian Government Department of Health and Ageing; 2009.

56. Mittmann N, Stout NK, Tosteson ANA, Trentham-Dietz A, Alagoz O, Yaffe MJ. Cost-effectiveness of mammography from a publicly funded health care system perspective. CMAJ Open. 2018;6(1):E77-E86.

57. Morton R, Sayma M, Sura MS. Economic analysis of the breast cancer screening program used by the UK NHS: should the program be maintained? Breast Cancer (Dove Med Press). 2017;9:217-25.

58. Pataky R, Ismail Z, Coldman AJ, Elwood M, Gelmon K, Hedden L, et al. Cost-effectiveness of annual versus biennial screening mammography for women with high mammographic breast density. Journal of medical screening. 2014;21(4):180-8.

59. Pharoah PD, Sewell B, Fitzsimmons D, Bennett HS, Pashayan N. Cost effectiveness of the NHS breast screening programme: life table model. Bmj. 2013;346:f2618.

60. Rojnik K, Naveršnik K, Mateović□Rojnik T, PrimicŽakelj M. Probabilistic cost□effectiveness modeling of different breast cancer screening policies in Slovenia. Value in health. 2008;11(2):139-48.

61. Souza FH, Polanczyk CA. Is Age-targeted full-field digital mammography screening cost-effective in emerging countries? A micro simulation model. Springerplus. 2013;2(1):366.

62. Stout NK, Rosenberg MA, Trentham-Dietz A, Smith MA, Robinson SM, Fryback DG. Retrospective cost-effectiveness analysis of screening mammography. Journal of the National Cancer Institute. 2006;98(11):774-82.

63. Stout NK, Lee SJ, Schechter CB, Kerlikowske K, Alagoz O, Berry D, et al. Benefits, harms, and costs for breast cancer screening after US implementation of digital mammography. JNCI: Journal of the National Cancer Institute. 2014;106(6).

64. Tosteson AN, Stout NK, Fryback DG, Acharyya S, Herman BA, Hannah LG, et al. Cost-effectiveness of digital mammography breast cancer screening. Annals of internal medicine. 2008;148(1):1-10.

65. Trentham-Dietz A, Kerlikowske K, Stout NK, Miglioretti DL, Schechter CB, Ergun MA, et al. Tailoring breast cancer screening intervals by breast density and risk for women aged 50 years or older: collaborative modeling of screening outcomes. Annals of internal medicine. 2016;165(10):700-12.

66. van Luijt PA, Heijnsdijk EAM, de Koning HJ. Cost-effectiveness of the Norwegian breast cancer screening program. International Journal of Cancer. 2017;140(4):833-40.

67. Wong IO, Kuntz KM, Cowling BJ, Lam CL, Leung GM. Cost effectiveness of mammography screening for Chinese women. Cancer: Interdisciplinary International Journal of the American Cancer Society. 2007;110(4):885-95.

68. Kerlikowske K, Salzmann P, Phillips KA, Cauley JA, Cummings SR. Continuing screening mammography in women aged 70 to 79 years: impact on life expectancy and cost-effectiveness. Jama. 1999;282(22):2156-63.

69. Madan J, Rawdin A, Stevenson M, Tappenden P. A rapid-response economic evaluation of the UK NHS Cancer Reform Strategy breast cancer screening program extension via a plausible bounds approach. Value Health. 2010;13(2):215-21.

70. Mandelblatt JS, Wheat ME, Monane M, Moshief RD, Hollenberg JP, Tang J. Breast cancer screening for elderly women with and without comorbid conditions: a decision analysis model. Annals of internal medicine. 1992;116(9):722-30.

71. Mandelblatt JS, Schechter CB, Yabroff KR, Lawrence W, Dignam J, Muennig P, et al. Benefits and costs of interventions to improve breast cancer outcomes in African American women. Journal of Clinical Oncology. 2004;22(13):2554-66.

72. Mandelblatt JS, Schechter CB, Yabroff KR, Lawrence W, Dignam J, Extermann M, et al. Toward optimal screening strategies for older women. Journal of general internal medicine. 2005;20(6):487-96.

73. AUSTRALIA B. Benefits, harms and costs of screening mammography in women 70 years and over: a systematic review. The Medical Journal of Australia. 2002;176(6):266-71.

74. Boer R, De Koning H, Van Oortmarssen G, Van der Maas P. In search of the best upper age limit for breast cancer screening. European Journal of Cancer. 1995;31(12):2040-3.

75. Messecar DC. Mammography screening for older women with and without cognitive impairment. Journal of Gerontological Nursing. 2000;26(4):14-24.

76. Rafia R, Brennan A, Madan J, Collins K, Reed MW, Lawrence G, et al. Modeling the cost-effectiveness of alternative upper age limits for breast cancer screening in England and Wales. Value in Health. 2016;19(4):404-12.

77. Salzmann P, Kerlikowske K, Phillips K. Cost-effectiveness of extending screening mammography guidelines to include women 40 to 49 years of age. Annals of Internal Medicine. 1997;127(11):955-65.

78. Sankatsing VD, Heijnsdijk EA, van Luijt PA, van Ravesteyn NT, Fracheboud J, de Koning HJ. Cost□effectiveness of digital mammography screening before the age of 50 in T he N etherlands. International journal of cancer. 2015;137(8):1990-9.

79. Pashayan N, Morris S, Gilbert FJ, Pharoah PP. Cost-effectiveness and benefit-to-harm ratio of risk-stratified screening for breast cancer: A life-table model. JAMA Oncology. 2018.

80. Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. Annals of internal medicine. 2011;155(1):10-20.

81. Vilaprinyo E, Forne C, Carles M, Sala M, Pla R, Castells X, et al. Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer. PloS one. 2014;9(2):e86858.

82. de Gelder R, Bulliard J-L, de Wolf C, Fracheboud J, Draisma G, Schopper D, et al. Cost-effectiveness of opportunistic versus organised mammography screening in Switzerland. European Journal of cancer. 2009;45(1):127-38.

83. Mandelblatt JS, Stout NK, Schechter CB, Van Den Broek JJ, Miglioretti DL, Krapcho M, et al. Collaborative modeling of the benefits and harms associated with different US breast cancer screening strategies. Annals of internal medicine. 2016;164(4):215-25.

84. Van Ravesteyn NT, Stout NK, Schechter CB, Heijnsdijk EA, Alagoz O, Trentham-Dietz A, et al. Benefits and harms of mammography screening after age 74 years: model estimates of overdiagnosis. JNCI: Journal of the National Cancer Institute. 2015;107(7).

85. Oliver A. Should we maximise QALYs? Applied health economics and health policy. 2004;3(2):61-6.

86. McNamee P, Glendinning S, Shenfine J, Steen N, Griffin SM, Bond J. Chained time trade-off and standard gamble methods. The European Journal of Health Economics, formerly: HEPAC. 2004;5(1):81-6.

87. Hoover DR, Crystal S, Kumar R, Sambamoorthi U, Cantor JC. Medical expenditures during the last year of life: findings from the 1992–1996 Medicare current beneficiary survey. Health services research. 2002;37(6):1625-42.

88. Tengs TO, Wallace A. One thousand health-related quality-of-life estimates. Medical care. 2000:583-637.

89. Bonomi AE, Boudreau DM, Fishman PA, Ludman E, Mohelnitzky A, Cannon EA, et al. Quality of life valuations of mammography screening. Quality of Life Research. 2008;17(5):801-14.

90. Chie W-C, Huang C-S, Chen J-H, Chang K-J. Utility Assessment for Different Clinical Phases of Breast Cancer in Taiwan. Journal of the Formosan Medical Association. 2000;99(9):677-83.

91. Gerard K, Dobson M, Hall J. Framing and labelling effects in health descriptions: quality adjusted life years for treatment of breast cancer. Journal of Clinical Epidemiology. 1993;46(1):77-84.

92. Hall J, Gerard K, Salkeld G, Richardson J. A cost utility analysis of mammography screening in Australia. Social Science & Medicine. 1992;34(9):993-1004.

93. Kim SH, Jo MW, Ock M, Lee HJ, Lee JW. Estimation of health state utilities in breast cancer. Patient Preference and Adherence. 2017;11:531-6.

94. Rijnsburger A, Essink-Bot M-L, van Dooren S, Borsboom G, Seynaeve C, Bartels C, et al. Impact of screening for breast cancer in high-risk women on health-related quality of life. British journal of cancer. 2004;91(1):69.

95. Dolan P. Developing methods that really do value the ‘Q’in the QALY. Health economics, policy and law. 2008;3(1):69-77.

96. Lamers LM, Stalmeier PF, Krabbe PF, Busschbach JJ. Inconsistencies in TTO and VAS values for EQ-5D health states. Medical Decision Making. 2006;26(2):173-81.

97. Furlong W, Feeny D, Torrance G, Barr R, Horsman J. Guide to design and development of health-state utility instrumentation. Centre for Health Economics and Policy Analysis (CHEPA), McMaster University, Hamilton, Canada; 1992.

98. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. Measuring and valuing health benefits for economic evaluation: OXFORD university press; 2017.

99. Longworth L, Rowen D. Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. Value in Health. 2013;16(1):202-10.

100. Von Neuman J, Morgenstern O. Theory of games and economic behavior. 1953.

101. Tolley KH. No title. What are health utilities? What is series, 2009. 2012.

102. Shepard DS. Cost‐effectiveness in Health and Medicine. By MR Gold, JE Siegel, LB Russell, and MC Weinstein (eds). New York: Oxford University Press, 1996. The journal of mental health policy and economics. 1999;2(2):91-2.

103. Bombardier C, Wolfson A, Sinclair A, McGreer A. Comparison of three preference measurement methodologies in the evaluation of a functional status index. Choices in Health Care: University of Toronto, Toronto; 1982.

104. Brazier J, Rowen D, Mavranezouli I, Tsuchiya A, Young T, Yang Y, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). 2012.

105. Aaronson NK, Muller M, Cohen PD, Essink-Bot M-L, Fekkes M, Sanderman R, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. Journal of clinical epidemiology. 1998;51(11):1055-68.

106. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of health economics. 2002;21(2):271-92.

107. Torrance GW, Furlong W, Feeny D. Health utility estimation. Expert Review of Pharmacoeconomics & Outcomes Research. 2002;2(2):99-108.

108. Attema AE, Brouwer WB. The value of correcting values: influence and importance of correcting TTO scores for time preference. Value in Health. 2010;13(8):879-84.

109. Hughes R, Huby M. The construction and interpretation of vignettes in social research. Social Work and Social Sciences Review. 2012;11(1):36-51.

110. Hershey JC, Kunreuther HC, Schoemaker PJ. Sources of bias in assessment procedures for utility functions. Management science. 1982;28(8):936-54.

111. O'Connor AM. Effects of framing and level of probability on patients' preferences for cancer chemotherapy. Journal of Clinical Epidemiology. 1989;42(2):119-26.

112. Jansen SJ, Stiggelbout AM, Wakker PP, Vliet Vlieland TP, Leer J-WH, Nooy MA, et al. Patients' utilities for cancer treatments: a study of the chained procedure for the standard gamble and time tradeoff. Medical Decision Making. 1998;18(4):391-9.

113. Brazier J. Valuing health states for use in cost-effectiveness analysis. Pharmacoeconomics. 2008;26(9):769-79.

114. Torrance GW, Feeny D, Furlong W. Visual analog scales: do they have a role in the measurement of preferences for health states? : Sage Publications Sage CA: Thousand Oaks, CA; 2001.

115. Brazier J, Green C, McCabe C, Stevens K. Use of visual analog scales in economic evaluation. Expert review of pharmacoeconomics & outcomes research. 2003;3(3):293-302.

116. Von Neumann J, Morgenstern O. Theory of games and economic behavior, 2nd rev. 1947.

117. Tosh JC, Longworth LJ, George E. Utility values in National Institute for Health and Clinical Excellence (NICE) technology appraisals. Value in Health. 2011;14(1):102-9.

118. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. The Lancet. 2007;370(9590):890-907.

119. Lew J-B, Simms K, Smith M, Lewis H, Neal H, Canfell K. Effectiveness Modelling and Economic Evaluation of Primary HPV Screening for Cervical Cancer Prevention in New Zealand. PLOS ONE. 2016;11(5):e0151619.

120. Hayes JH, Ollendorf DA, Pearson SD, Barry MJ, Kantoff PW, Stewart ST, et al. Active surveillance compared with initial treatment for men with low-risk prostate cancer: a decision analysis. Jama. 2010;304(21):2373-80.

121. McCaffery K, Nickel B, Moynihan R, Hersch J, Teixeira-Pinto A, Irwig L, et al. How different terminology for ductal carcinoma in situ impacts women's concern and treatment preferences: a randomised comparison within a national community survey. BMJ open. 2015;5(11):e008094.

122. Nickel B, Barratt A, Copp T, Moynihan R, McCaffery K. Words do matter: a systematic review on how different terminology for the same condition influences management preferences. BMJ open. 2017;7(7):e014129.

123. Campbell B, Yang JC, Gonzalez J, Reed S, Havrilesky L, Johnson FR, et al. Patient preferences for DCIS treatment strategies: A discrete-choice experiment. Annals of Surgical Oncology. 2018;25 (2 Supplement 1):38-9.

124. Esserman LJ, Thompson IM, Reid B. Overdiagnosis and overtreatment in cancer: an opportunity for improvement. Jama. 2013;310(8):797-8.

125. Bulliard J-L, Chiolero A. Screening and overdiagnosis: public health implications. Public health reviews. 2015;36:8-.

126. Keller A, Gericke C, Whitty JA, Yaxley J, Kua B, Coughlin G. A Cost-Utility Analysis of Prostate Cancer Screening in Australia. Applied Health Economics and Health Policy. 2017;15(1):95-111.

127. Gram I, Lund E, Slenker S. Quality of life following a false positive mammogram. British journal of cancer. 1990;62(6):1018.

128. Lindfors KK, O'Connor J, Parker RA. False-positive screening mammograms: effect of immediate versus later work-up on patient stress. Radiology. 2001;218(1):247-53.

129. Parker PA, Youssef A, Walker S, Basen-Engquist K, Cohen L, Gritz ER, et al. Short-term and long-term psychosocial adjustment and quality of life in women undergoing different surgical procedures for breast cancer. Annals of surgical oncology. 2007;14(11):3078-89.

130. Stiggelbout AM, Kiebert GM, Kievit J, Leer J-WH, Stoter G, De Haes J. Utility assessment in cancer patients: adjustment of time tradeoff scores for the utility of life years and comparison with standard gamble scores. Medical Decision Making. 1994;14(1):82-90.

131. Hayman JA, Kabeto MU, Schipper MJ, Bennett JE, Vicini FA, Pierce LJ. Assessing the benefit of radiation therapy after breast-conserving surgery for ductal carcinoma-in-situ. J Clin Oncol. 2005;23(22):5171-7.

132. Arnold M. Simulation modeling for stratified breast cancer screening - a systematic review of cost and quality of life assumptions. BMC Health Services Research. 2017;17(1):802.

133. Parmigiani G. Modeling in medical decision making,. West Sussex, England.: John Wiley and Sons, Ltd,; 2002.

134. Rosser R, Kind P. A scale of valuations of states of illness: is there a social consensus? International journal of epidemiology. 1978;7(4):347-58.

135. Ara R, Brazier JE. Using health state utility values from the general population to approximate baselines in decision analytic models when condition-specific data are not available. Value in Health. 2011;14(4):539-45.
136. Lidgren M, Wilking N, Jönsson B, Rehnberg C. Health related quality of life in different states of breast cancer. Quality of life research. 2007;16(6):1073-81.
137. Souza FH. Mamografia digital em comparação com mamografia convencional no rastreamento de câncer de mama no Brasil: revisão sistemática, custo da doença e analise de custo-efetividade no Sistema Único de Saúde. 2012.
138. Cruz LN. Medidas de qualidade de vida e utilidade em uma amostra da população de Porto Alegre. 2010.

Figures and Tables

Figure 1: A PRISMA flow chart of studies included and excluded at each stage

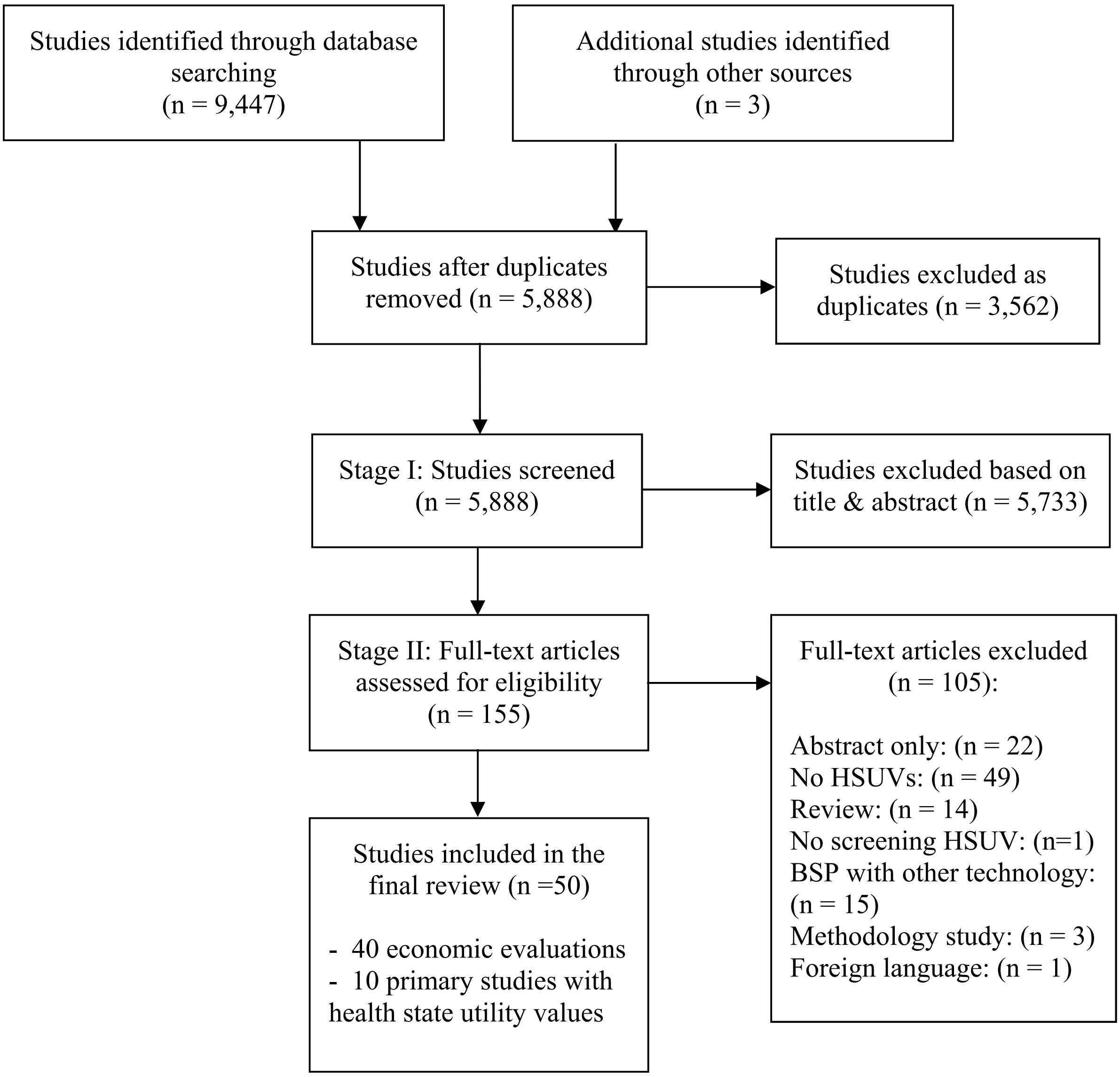

Table 1: Characteristics of the 40 economic evaluations

Lead Author	Study objective	Country; Population; Study type	Health states included	Method used by cited HSUVs	Information presented about HSUVs	HSUVs in study	Duration	Cited sources for HSUVs
Ahern (46)	Assess the cost-effectiveness of mammography screening and breast examination	USA; Women aged 40-79; 10 MM strategies (1-2y +/- CBE); MSM	Treatment (intervention)	VAS	Utilities from another model and systematic review (expert VAS transformed to SG using SG=1-(1-VAS)$^{2.29}$	0.590-1.000	6 months, 1 year, lifelong	(17, 133)
Arrospide (47)	Retrospective economic evaluation of Basque BSP	Spain; Women aged 50-69; 2y MM, MSM	Diagnosis Treatment (disease stage)	EQ-5D assumption	Applied assumptions from another model: tariff for the disutility of breast cancer applied to healthy population EQ-5D data (Spain).	0.338-0.824	1 year/ life expectancy	(62)
Barratt (73)	Assess the cost-effectiveness of extending BSP for women over 70	Australia; Women aged over 70, 2y MM, MSM	Screening Diagnosis Treatment (intervention)	VAS	Extrapolated the QALYs from another model which used expert VAS (systematic review).	0.288-0.994	Unclear	(32, 74)
Beemsterboer (48)	Economic evaluation of different screening strategies in Germany	Germany; Women aged 50-69, 2y MM, MSM	Screening Diagnosis Treatment (intervention)	VAS	Expert VAS utilities and durations (transformed to TTO)	0.288-0.994	1 week-lifelong	(32)
Boer (74)	Economic evaluation of extending the upper age limit of BSP	Netherlands; Women aged 50-69 and >70; 2y MM; MSM	Screening Diagnosis Treatment (intervention)	VAS	Expert VAS utilities and durations (transformed to TTO)	0.288-0.994	1 week-lifelong	(32)
Carles (49)	Economic evaluation of breast screening strategies in Catalonia	Spain; Women aged 50-79; 1-2y MM, Probabilistic model	Screening Diagnosis Treatment (disease stage)	EQ-5D assumption	Used the assumptions for duration and loss from healthy population EQ-5D in another model (US)	0.657-0.994	7 days-lifelong	(62)
Christensen (50)	Evaluate the cost-effectiveness of mammography screening in Greenland.	Greenland; Women aged 50-69; 2y MM; CEA	Screening Diagnosis Treatment (intervention)	Systematic review, assumption	Population QoL (Greenland) adjusted using values from a systematic review. Methods not reported.	0.480-0.810	6 months	(41)
De Gelder (82)	Economic evaluation of opportunistic and organised population mammography screening	Switzerland; Women aged 50-69; 2y MM, MSM	Screening Diagnosis Treatment (intervention)	VAS	Expert VAS utilities and durations (transformed to TTO) used in another model	0.288-0.994	1 week- lifelong	(32, 51)
De Koning (51)	Evaluate the cost-effectiveness of different BSP strategies	Netherlands; Women aged 40-75; 5 variants 1.3-3y MM, MSM	Screening Diagnosis Treatment (intervention)	VAS	HSUVs and durations based on 27 experts VAS (transformed to TTO)	0.289-0.994	1 week-lifelong	(32)
Forrest (52)	Cost-effectiveness of implementing a national BSP in the UK.	UK; Women aged 50-65; 3y MM; CUA	Treatment (intervention)	Rosser scale	Rosser ratio rating scale values used for the disutility of surgery	0.920	Lifelong	(134)
Haghighat (53)	Economic evaluation of mammography screening in Iran	Iran; Women aged 40-70, 3y MM, Markov model	Treatment (disease stage)	Assumption	Used the assumptions in another economic model of BSP	0.300-0.950	Unclear	(67)
Hakama (54)	Economic evaluation of Nordic breast screening strategies.	Nordic region; Women aged 50-69, CUA	Screening Diagnosis Treatment (intervention)	VAS	Expert VAS utilities and durations (transformed to TTO)	0.288-0.994	1 week-lifelong	(32)
IMS Health (55)	Economic evaluation of BSP in Australia	Australia; strategies for women aged 40-79, 2y MM, MSM	Treatment (disease stage)	VAS	Expert VAS ratings, authors adjusted weighting and duration using local treatment data	0.774-0.864	Unclear	(51)

Kerlikowske (68)	Economic evaluation of mammography screening in elderly women.	USA; Women aged 65-79; 2y MM, Markov model	Treatment (disease stage)	Assumption	Authors made assumptions of plausible estimates based on published HSUVs (TTO/VAS).	0.300-0.900	Lifelong	(32, 92)
Madan (69)	Cost-effectiveness of extending the lower age limit of BSP	UK; Women 47-49 years; 3y MM; MSM	Diagnosis	EQ-5D Assumption	Baseline UK healthy general population EQ-5D scores adjusted by assumption in sensitivity analysis	Not reported	Unclear	(10)
Mandelblatt (70)	Evaluate the cost-effectiveness of BSP in elderly women with and without comorbid disease	USA; Women aged 65-85 years; 1-2y MM; Decision model	Screening Diagnosis Treatment (disease stage)	Assumption	Assumption of plausible HSUVs based on expert VAS in the literature for similar health states	0.100-0.900	5 days, 30 days, life expectance	(32)
Mandelblatt (71)	Economic evaluation of targeted mammography screening in African American women	USA; African American women aged ≥40,; 1-2y MM; MSM	Treatment (disease stage)	Assumption	Assumption of HSUVs by disease stage, no description of how values were determined is provided.	0.500-1.000	Unclear	No cited source
Mandelblatt (72)	Evaluate the cost-effectiveness of a BSP in older women	USA; Women aged 50+; 2y MM; MSM	Treatment (disease stage)	Assumption	Assumption of HSUVs by disease stage, no description of how values were determined is provided.	0.550-0.950	1 year	No cited source
Mandelblatt (83)	Partial evaluation of mammography strategies considering screening and treatment advances	USA; Women aged 40-74; 1-2y MM; MSM	Screening Diagnosis Treatment (disease stage)	EQ-5D Assumption, VAS	Expert VAS values (screening and diagnosis) and assumptions from another model (treatment) for US population EQ-5D tariffs	0.354-0.856	1 weeks, 5 weeks, 2 years	(32, 62)
Messecar (75)	Economic evaluation of BSP for older women with and without cognitive impairment	USA; Women aged 75-85y; 2y MM; Decision model	Treatment (disease stage)	TTO	Used general population TTO preferences for treatment	0.260-0.800	Lifelong	(91)
Mittmann (56)	Updated cost-effectiveness of BSP in Canada.	Canada; Women aged 40-74; 1-3y MM; MSM	Screening Diagnosis	EQ-5D Assumption, VAS	Expert VAS values (screening and diagnosis) and assumptions from another model (treatment) for US population EQ-5D tariffs	0.895-0.994	1 week, 5 weeks, 2 years	(32)
Morton (57)	Economic analysis of the BSP in the UK	UK; Women aged 50-70; 3y MM; CUA	Screening Treatment (intervention)	Assumption	Used QALYs from another economic model of the UK BSP	Not reported	Unclear	(29, 69)
Pashayan (79)	Cost-effectiveness or risk-stratified screening for breast cancer	UK; Women aged 50-69; 3y MM (risk); Lifetable	Treatment (intervention)	EQ-5D Assumption	Used adjusted population EQ-5D utilities from another economic model (systematic review)	Not reported	1 year, lifelong	(41, 59, 135)
Pataky (58)	Cost-effectiveness of population BSP by age & frequency	Canada; Women aged 40-74; 1-2y MM; MSM	Diagnosis, Treatment (disease stage)	VAS, SG	Systematic review (expert and population VAS to SG)	0.389-1.000	2 weeks-lifelong	(17, 32, 89, 131)
Pharoah (59)	Economic evaluation of the National Health Service BSP	UK; Women aged 50-70; 3y MM; Life-table	Treatment (intervention)	EQ-5D Assumption	UK general population EQ-5D adjusted by a 0.9 relative reduction	Not reported	Lifelong	(41, 135)
Rafia (76)	Cost-effectiveness of extending the upper age limit of the UK BSP.	UK; Women aged 50-90; 3y MM; MSM	Screening Diagnosis Treatment (disease stage)	EQ-5D Assumption modified by VAS, SG	Expert VAS, population SG and expert opinion used to adjust baseline UK population EQ-5D	0.360-0.910	2 hours, 3 weeks, 1-3 years, lifetime	(10, 89, 94)
Raftery (29)	Assess the benefit and harms of the UK BSP (partial evaluation)	UK; Women aged 50-70; 3y MM, life-table	Diagnosis Treatment (intervention)	Assumption	Systematic review, other models (expert VAS, population EQ-5D)	Not reported	0.2 years-Lifelong	(32, 62) (41)
Rojnik (60)	Economic evaluation of alternative breast screening strategies in Slovenia	Slovenia; Women aged 40-80 years; 1-3y MM; MSM	Screening Diagnosis Treatment (intervention)	VAS, SG	Expert VAS and SG utilities (oncology nurses), literature review	0.515-0.994	1 month-lifelong	(32)
Salzmann (77)	Cost-effectiveness of extending mammography screening to women aged 40 to 49 years.	USA; Women aged 40-49 years and 50-69 years; 1.5-2y MM, Markov model	Treatment (disease stage)	TTO	Australian patient TTO utilities in sensitivity analysis	0.300-0.800	Unclear	(92)

Sankatsing (78)	Cost-effectiveness of mammography screening before the age of 50.	Netherlands; Women aged 40-74 years; 2y MM; MSM	Screening Diagnosis Treatment (disease stage)	EQ-5D, Assumption VAS	Expert VAS utilities (screening, diagnosis) Decrements in US healthy general population EQ-5D (treatment) from another model.	Unclear	1 week, lifelong	(32, 62)
Schousboe (80)	Cost-effectiveness of mammography screening by risk factors.	USA; Women aged 40-79; 1-2y MM, MSM	Treatment (disease stage)	EQ-5D	Swedish breast cancer patient EQ-5D applied to Swedish general female population EQ-5D	0.620-1.000	5 days- lifelong	(136)
Souza (61)	Economic evaluation of implementing a national BSP in Brazil	Brazil; Women aged 40-69; 1-2y MM; MSM	Diagnosis Treatment (disease stage and intervention)	SF-6D assumption	Author assumption of plausible estimate for false positive MM. HSUVs were estimated based on patient SF-6D scores (Brazil)	0.686-0.800	2 months-lifelong	(137, 138)
Stout (62)	Economic evaluation comparing alternative screening strategies.	USA; Women aged 40-80; 1-5y MM; DESM	Screening Diagnosis Treatment (disease stage)	EQ-5D assumption	Age-sex specific EQ-5D for healthy women (US) adjusted for negative effects of breast cancer diagnosis and treatment.	0.354-0.856	1 week- lifelong	(62)
Stout (63)	Assess the benefit, harms and costs of digital mammography screening	USA; Women aged 40-74; 1-2y MM; MSM	Screening Diagnosis Treatment (disease stage)	EQ-5D assumption, VAS	Population EQ-5D (US) adjusted using assumptions from another model. Expert VAS utilities included in sensitivity analysis.	0.354-0.586	1 week-lifelong	(32, 62)
Tosteson (64)	Evaluate the cost-effectiveness of digital mammography screening	USA; Women aged ≥ 40;1y MM; MSM	Treatment (disease stage)	EQ-5D assumptions	Applied the duration and weighting assumptions from another economic model of BSP to healthy population EQ-5D data (USA)	0.430-0.860	Unclear	(62)
Trentham-Dietz (65)	Economic evaluation of tailored mammography screening for women over 50 years	USA; Women aged 50-74; 1-3y MM (risk); MSM	Screening Diagnosis Treatment (disease stage)	EQ-5D assumption, VAS	Expert VAS (screening and diagnosis). US population EQ-5D (treatment) adjusted using assumptions from another model	Unclear	1 week-lifelong	(32, 62)
Van Luijit (2017) (80)	Economic evaluation of the Norwegian BSP	Norway; Women aged 50-69; 2y MM; MSM	Screening Diagnosis Treatment (disease stage)	VAS	Expert VAS utilities (transformed to TTO) from the literature	0.288-0.994	1 week-lifelong	(32)
Van Ravesteyn (84)	Assess the benefits and harms of mammography after age 74 years (partial evaluation)	USA; Women aged 50-94; 2y MM; MSM	Screening Diagnosis Treatment (disease stage)	EQ-5D assumption, VAS	Expert VAS for (screening and diagnosis). US population EQ-5D assumptions adopted from another economic model.	0.600-0.994	1 week, 5 weeks, 2 years, life expectancy	(32, 62)
Vilaprinyo (81)	Cost-effectiveness of risk-based breast screening strategies for breast cancer	Spain; Women aged 40-74; 1-5y MM, Probabilistic model	Diagnosis Treatment (disease stage)	EQ-5D	Patient EQ-5D (Sweden) for treatment extrapolated using the methods from another model.	0.655-0.859	2 months-5 years	(80, 136)
Wong (67)	Economic evaluation of biennial mammography screening in Hong Kong.	China; Women aged 40-79; 2y MM, Markov model	Treatment (disease stage)	Assumption	HSUVs from another economic model (but values do not match those cited)	0.300-0.950	Lifelong	(71)

Legend

BSP: breast screening programme, EQ-5D: Euroqol-5D, HSUV: health state utility value, LYG: life years gained, MM: mammogram, MSM: microsimulation model, QALY: quality adjusted life year, QoL: quality of life, , TTO: Time trade off, SG: standard gamble, VAS: visual analogue scale

Table 2: Characteristics of the 10 primary studies which had elicited HSUVs

Author; Country	Study aim	Participants	Health states valued	Utility range	Duration range	Technique	Further information methods
Bonomi USA (89)	Obtain QoL values for mammography screening and breast cancer treatment	131 women sampled from a population breast screening programme (aged 50-79)	Screening attendance Screening result (FP, TN) Diagnostic mammogram Treatment (intervention) Disease free at 1 year Recurrence at 1 year Terminal care	0.804 0.457-0.891 0.553 0.397-0.530 0.768 0.330 0.358	2 hours 2 weeks 2 weeks 4 months-5 years Lifelong 4 months 3 months	VAS	14 vignettes via in-person or telephone interview. VAS anchored death-perfect health
Chie Taiwan (90)	Utility in different clinical phases of breast cancer.	21 clinical and public health experts	Screening attendance Diagnosis Initial treatment (intervention) Post-treatment (intervention) Recurrence at 1 year Terminal care	0.900-1.000 0.700-0.900 0.500-0.800 0.600-0.800 0.250-0.300 0.100-0.150	20 years for all	VAS, TTO SG	17 vignettes via face-to-face interview (visual aids). VAS anchored death-perfect health.
De Haes Netherlands (32)	Elicit utilities for use in an economic model of BSP	27 clinical and public health experts	Screening attendance Diagnosis Initial treatment (intervention) Post-treatment (intervention) Disease free >1 year Terminal care	0.994 0.895 0.717-0.820 0.844-0.914 0.947-0.960 0.288	1 week 5 weeks 2 months-2 years 10 months Lifelong 1 month	VAS	15 vignettes via face-to-face interview. VAS anchored worst-best imaginable health. VAS scores transformed to TTO using the formula: TTO=1-(1-VAS) ^1.82
Gerard Australia (91)	Explore framing and labelling effects on breast cancer values.	180 women from the local general population (aged 45-69)	Treatment (intervention) of screen detected breast cancer with and without breast cancer death	0.150-0.750	10-30 years (age dependent)	TTO	9 different presentations of two breast cancer vignettes (varied cancer terminology and pronoun).
Gerard UK (15)	Determine the feasibility of mapping EQ-5D to TTO for validating breast cancer descriptions.	440 women from the general population eligible for breast screening (aged 40-64)	True negative False positive True negative False positive	0.910-0.940 0.210-0.790 0.480-0.660 0.450-0.660	12 months Lifelong	TTO (chain) EQ-5D	Two-stage chaining used to adjust temporary health states onto death-full health scale. EQ-5D mapped onto TTO using 3/5 dimensions.
Hall Australia (92)	Derive utilities for use in an economic evaluation of BSP in Australia.	44 women from the general population and 60 breast cancer patients (aged 45-69)	Treatment (intervention) of a screen detected breast cancer	0.270-0.800	10-30 years (age dependent)	TTO	6 vignettes via face-to-face interview
Johnston; UK (16)	Derive QoL values for key breast screening outcomes	440 women from the general population eligible for breast screening (aged 40-64)	True negative False positive True negative False positive	0.91 0.66 0.66 0.66	12 months Lifelong	VAS TTO (chain)	Two-stage chaining method used to adjust temporary health states onto death-full health scale
Kim Korea (93)	Determine the utility of breast cancer health states in Korean population.	509 general population men and women (aged >19)	Treatment (intervention) of screen detected non-invasive, invasive or advanced breast cancer, recurrence, terminal care	VAS: 0.170-0.681 SG: 0.352-0.804	Lifelong	VAS, SG	8 vignettes via face-to-face interview. VAS anchored worst-best health (readjusted to dead).
Rijnsburger Netherlands (94)	Assess the QoL of screening high-risk women for breast cancer.	334 women in a high-risk breast screening trial (mean age 40.9).	Screening attendance	VAS: 0.807-0.819 EQ-5D: 0.880	Unclear	VAS, EQ-5D, SF-36	Direct measurement at time points of 2 months prior, during and 1-4 weeks after attending screening.
Tosteson USA (34)	Measure the QoL impact of false-positive mammograms	1028 women in digital breast screening trial: 534 = negative, 494 = false positive	Screening (negative mammogram) Diagnosis (positive mammogram)	VAS: 0.830-0.860 EQ-5D: 0.900-0.910	Unclear	EQ-5D VAS	Direct measurement at baseline and up to 1 year after screening. VAS anchored worst-best imaginable health