

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](https://doi.org/10.1016/j.socscimed.2019.03.028)

License:

Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Document Version

Peer reviewed version

Citation for published version (Harvard):

Bromley, H, Petrie, D, Mann, GB, Nickson, C, Rea, D & Roberts, T 2019, 'Valuing the health states associated with breast cancer screening programmes: a systematic review of economic measures', *Social Science and Medicine*, vol. 228, pp. 142-154. <https://doi.org/10.1016/j.socscimed.2019.03.028>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

Checked for eligibility: 19/03/2019

Bromley, H. et al (2019) Valuing the health states associated with breast cancer screening programmes: A systematic review of economic measures, *Social Science & Medicine*, 228: 142-154; <https://doi.org/10.1016/j.socscimed.2019.03.028>

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

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

3

4 **Abstract**

5

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

14 40 economic evaluations of breast cancer screening programmes and 10 primary studies
15 measuring utilities for health states associated with mammography were reviewed in full. The
16 economic evaluations made different assumptions about the measures used, duration applied and the
17 sequelae included in each health state. 22 evaluations referenced utilities based on assumptions or
18 used measures that were not methodologically appropriate. There was significant heterogeneity in the
19 utilities generated by the 10 primary studies, including the methods and population used to derive
20 them. No study asked women to explicitly consider the risk of overdiagnosis when valuing the health
21 states described.

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

26

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

28

29 **Introduction**

30 Evidence regarding the cost-effectiveness of healthcare technology is increasingly required to inform
31 the decision on whether to fund and implement new treatment (1). Many decision-making bodies
32 require interventions to be assessed using cost per quality-adjusted-life-years (QALYs) (2), a single
33 summary measure combining life expectancy with individuals' relative preferences for health states in
34 terms of quality of life (3, 4). Health state utility values (HSUVs) are cardinal measures of preference
35 rated on a utility scale anchored from dead (0) to perfect health (1). Utilities can be valued directly or
36 indirectly (5). Direct methods ask individuals to value hypothetical health states, and preferences are
37 directly measured onto the utility scale using the standard gamble (SG), time-trade off (TTO) or
38 visual analogue scale (VAS) (6). The TTO and SG elicit individual choices under uncertainty in life
39 expectancy or risk of death and good health (7), whereas the VAS provides an intermediate valuation
40 of health on a graduated rating scale (8). Indirect methods use a generic multi-attribute utility
41 instrument (MAUI), such as the EuroQol-5 Dimensions (EQ-5D) (9). Current or hypothetical health is
42 mapped onto a generic health instrument and indirectly valued using tariffs for the generic health
43 states that have previously been estimated using direct valuation methods from the general population
44 (10).

45
46 Economic evaluations impact health policy decisions and so the methodological quality of the
47 parameters used to inform such analyses must be robust (11). Whilst a growing wealth of literature
48 has explored the importance of the economic approach used, less attention has been given to the
49 methods and quality of the evidence used to inform HSUVs and thus QALYs (12). It is important that
50 the methods used to identify, select and appraise utilities are transparent and systematic to reduce
51 model bias and potential misallocation of resources (13). Several criteria are important for the
52 selection of relevant HSUVs (14). The first relates to the health states, methods, descriptive system
53 and population used to elicit the utilities. Where HSUVs have been measured directly, the validity,
54 reliability and feasibility of the generated values should also be explicitly considered (15). Second, the
55 duration of impact applied must be measured appropriately for both temporary and chronic health
56 states associated with the intervention (16). The third relates to the generalisability of the condition,

57 severity and population characteristics in the utility study to those in the economic evaluation using
58 them (2).

59
60 The quality of the HSUVs applied is particularly pertinent in the appraisal of oncological
61 interventions, where quality of life may have greater influence on QALYs than the modest gains in
62 life expectancy (17). Many studies have evaluated the cost-effectiveness of breast cancer screening
63 (18, 19), with those including quality of life in their evaluation reporting fewer net benefits, yet few
64 have commented on the quality of the utility estimates used to inform them. When deciding on
65 preferred screening policy it is critical to be able to accurately value the options available to women of
66 being able to attend routine screening (20). This means valuing all associated benefits and risks
67 associated with the alternative screening policy in terms of utility (21).

68
69 Screening for breast cancer in women aged 50 to 74 is recommended because of the ability to capture
70 disease earlier and reduce treatment intensity and disease mortality (22, 23). Decision makers must
71 value the risk that screening would lead to a woman having necessary (and perhaps less intense)
72 treatment at an earlier stage than she would have otherwise had, against the risk of the woman having
73 an unnecessary diagnosis and treatment (24). This valuation is made even more challenging because
74 there is limited evidence on the rate of progression for many breast tumour types (25). If policy
75 makers are to interpret cost-effectiveness analyses of mammography screening and balance the
76 benefits and harms of such interventions appropriately, the utilities used in such evaluations must
77 reflect the health states and effect on those experiencing the sequelae, including overdiagnosis and
78 overtreatment (20).

79
80 **The challenges associated with valuing health states for breast cancer screening**

81 There are several challenges relating to the identification and assessment of HSUVs for use in the
82 economic evaluation of breast cancer screening programmes specifically. First, the natural history of
83 breast cancer is poorly understood (26, 27). Not all valuation methods for deriving utility may account
84 for the uncertainty in disease progression in the valuation process (28). Second, overdiagnosis and

85 overtreatment from screening create a “paradoxical popularity” because individual women may value
86 unnecessary treatment inappropriately if screening and intervention for benign disease is misconstrued
87 as life-saving (29). Qualitative evidence suggests that both population and patient understanding of
88 overdiagnosis is poor (30), with most perceiving the prognosis of pre-cancerous disease equal to that
89 of an invasive breast cancer (31). Third, the sequelae associated with breast screening last for different
90 durations (32). The long-term implications of a mastectomy are permanent (33), but the anxiety or
91 reassurance associated with mammography screening or diagnostic investigation may only be
92 temporary (34). Temporary health states require modification of conventional valuation methodology
93 and economic evaluations must consider how both temporary and chronic health states are valued
94 simultaneously within a single model (35). Fourth, it is unclear whose preferences would be best
95 placed to assess the benefits and harms of breast screening (36). The National Institute for Healthcare
96 and Clinical Excellence (NICE) advocate the use of general population preferences in a publicly
97 funded healthcare system, (2) yet given the complexity involved in valuing screening it may be
98 difficult for the lay person to quantify using conventional utility instruments. The preferences and
99 disease characteristics of individual women and breast cancers also vary significantly by demographic
100 (37) and so the generalisability of the population in the primary and economic studies may influence
101 the generated QALYs (38). Such challenges may impact utility instruments, therefore an assessment
102 of the methodology used to overcome these issues is critical in the appraisal of appropriate HSUVs.

103

104 **Objectives**

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

109

110 **Methods**

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

113

114 **Eligibility Criteria**

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

125

126 **Search strategy**

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

132

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

139

140 **Study selection**

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

149

150 **Data extraction and analysis**

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

160

161

162 **Results**

163

164 The database search retrieved 9,447 studies, of which 3,562 were removed as duplicates. A further 3
165 studies were identified through hand-searching the reference lists of relevant studies. A flow diagram
166 of the studies selected or excluded at each stage, with reasons, is provided in Figure 1. Data were
167 extracted for the 50 relevant studies included in the narrative review: 40 economic evaluations using

168 cost per QALYs and 10 primary studies that measured utilities for health states associated with
169 mammography screening.

170

171

172 **Economic evaluations using Cost per QALY**

173

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

183

184 The utilities associated with breast cancer screening are difficult to compare because each study made
185 different assumptions about the value used, the duration over which they were applied and the
186 sequelae included in each of the health states. The values used for screening attendance varied
187 significantly (0.100-0.994) and were applied for a duration of between 2 hours and 7 days. There were
188 similar issues with heterogeneity between HSUVs for diagnosis (0.100-0.895) and treatment (0.100-
189 0.990). Utilities were applied for between 5 days and 6 months for a positive mammogram and a
190 duration of one month to the rest of the woman's life for treatment depending on classification by
191 intervention or disease stage. The duration of utilities or disutilities when applied in economic models
192 can be a key driver in influencing results using QALYs (85), yet few studies justified the duration
193 enforced (29, 80, 81) or considered whether the utilities for temporary health states had used an
194 appropriate chaining adaptation (86).

195

196 Most economic evaluations used the same two sources (32, 62) for their utilities, although there was
197 variation in the actual value used and the generalisability of the population for which they were
198 applied. The first of these sources by Stout et al. (62) applied tariffs based on assumptions for
199 screening, diagnosis and treatment of breast cancer by stage at diagnosis to adjust US healthy
200 population EQ-5D estimates (87). Although the generated utilities were deemed consistent with those
201 reported in other studies (17, 88), it is not clear how the assumed adjustment for each health state was
202 determined, and yet nine economic evaluations (29, 47, 49, 63-65, 78, 83, 84) applied this method.
203 For screening disutility, almost half of the economic evaluations (29, 48, 51, 54, 56, 58, 60, 63, 65,
204 66, 68, 73, 74, 78, 84) used expert VAS utilities derived from a second study in the Netherlands (32),
205 but only three economic evaluations (51, 54, 73) considered the generalisability of the expert sample
206 to the general population in the model to which this was applied. Other economic evaluations made
207 their own adjustments to local population EQ-5D or SF-6D data (47, 49, 50, 59, 69, 79, 81) or used
208 HSUVs elicited directly (VAS, TTO, SG) from samples of women with comparable demographics to
209 try and improve the relevance of the applied utilities to the population in their economic model (52,
210 73, 77). The remaining evaluations cited utilities from another economic model (53, 67), systematic
211 review (46) or made their own assumption of an appropriate value (57, 70-72) but did not provide a
212 detailed critique of how these were derived.

213

214 Sensitivity analyses were used to analyse the uncertainty around HSUVs in the majority of the 40
215 economic evaluations, with at least half reporting quality of life as having a significant effect on cost-
216 effectiveness results. However, not all economic evaluations included all relevant phases of the
217 mammography screening pathway in their analysis and therefore implicitly assumed they had no
218 impact on quality of life. 22 studies (29, 46, 47, 52, 53, 55-59, 64, 67-69, 71, 72, 75, 77, 79-81) did
219 not integrate the potential reassurance or disutility of screening anxiety and diagnostic follow-up in
220 their analyses and a further 27 did not explicitly capture the disutility relating to the risk of
221 overdiagnosis (29, 46, 48, 49, 51-59, 61, 62, 64, 67-71, 73-77). Consequently, no utility loss was
222 applied to reflect this uncertainty in more than half of the economic evaluations, which may bias
223 results (QALYs) toward more frequent screening (29, 49, 52, 54, 59). This limitation was justified in

224 five studies due to the lack of robust HSUVs for mammography screening. For the 13 studies (47, 50,
225 60, 63, 65, 66, 78-84) which did attempt to value overdiagnosis in their analysis, an assumption was
226 made that this was captured in the QALYs across screening strategies by including the temporary
227 disutility of diagnosis and treatment without a corresponding gain in life years. However, the utilities
228 applied used sources which had not highlighted that there was a risk the treatment was unnecessary
229 during the valuation process and therefore is unlikely to fully capture the impact of the risk of
230 overdiagnosis on quality of life.

231

232 **Primary studies**

233

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

242

243 **Methodology**

244 The main method for valuing health states include direct and indirect empirical measurement or
245 expert opinion (95). Multiple approaches were taken to elicit utilities for breast screening health states
246 identified by this review, with more than four different valuation techniques reported within the
247 primary studies. These four primary approaches included the VAS (96), which was anchored from
248 worst to best imaginable health, standard gamble which compared the health state against a gamble of
249 death and perfect health (97), time trade-off which trades years lived in full health against living
250 longer in the health state being valued (98) and EQ-5D which asked trial participants to report their
251 own or hypothetical health on a generic scale and applied general population tariffs to estimate final

252 utility scores (99). More than one technique was used to value screening health states in six studies
253 (15, 16, 34, 90, 93, 94), whilst the remainder used a single technique (32, 89, 91, 92). The standard
254 gamble, initially presented by Neumann and Morgenstern (100) in (101) is the gold standard method
255 for valuing conditions of uncertainty (102), yet only two studies (90, 93) used this technique to
256 capture the potential benefit and risks associated with screening. An alternative choice-based method
257 (TTO) was justified by five studies to reduce cognitive burden associated with the standard gamble
258 (15, 16, 90-92). De Haes et al. (32) did not use a choice-based method but mapped visual analogue
259 scale scores into utilities using a power function (VAS): $TTO = 1 - (1 - VAS)^{1.82}$. (103), although
260 there are reported issues with the reliability of conversion formulas (104). A combination of both
261 direct and indirect methods was used by the remaining studies (34, 94) using tariffs from the US (87)
262 and Dutch general population (105) for the EQ-5D descriptive instrument before and after screening.
263 The Short Form-36 questionnaire was also used by Rijnsburger (94), but the values were never
264 mapped into SF-6D utilities (106). Only half of the studies considered whether the chosen method
265 was appropriate for overcoming the methodological challenges associated with screening health states
266 (15, 16, 32, 91, 92).

267

268 **Duration**

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

277

278 A combination of direct and indirect assumptions (108) were used to specify duration in the remaining
279 studies. Four studies (90-93) specified a single duration of impact for both temporary and chronic

280 health states and applied the same method (TTO or SG) to ensure consistency. The same technique
281 (VAS) was used in two studies (32, 89) to specify the time within the vignettes, although the durations
282 applied varied depending on the timeframe assumed. Other studies (34, 94) did not specify the health
283 state duration per se but indirectly measured utility at discrete time points during the screening
284 process. However, due to variation in follow-up time some women were aware of their results a priori
285 which may have inadvertently biased results.

286

287 **Descriptive system**

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

302

303 No primary study explicitly considered the impact of the risk of overdiagnosis or unnecessary
304 treatment in any of the health states described. Only Gerard (91) and Hall (92) introduced the notion
305 of dying of causes other than breast cancer in their vignettes, although they did not explicitly include
306 the risk of unnecessary follow up and treatment. Kim et al. (93) explicitly included risk in their health

307 state descriptions of surgery and radiotherapy but only provided estimates for recurrence and survival,
308 assuming all treatment was necessary for non-invasive disease.

309

310 **Population**

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

322

323 **Quality assessment**

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

334

335 **Discussion**

336

337 **Principal findings**

338 Population based mammography screening for breast cancer is a major public health investment and
339 significant time investment for women and therefore warrants rigorous scrutiny (20). This systematic
340 review provides the first synthesis of economic measures and health states used to value
341 mammography screening, explicitly including overdiagnosis, and summarises the evidence base
342 informing the population screening debate. The identified evaluations found that quality of life had a
343 significant effect on cost-effectiveness results in sensitivity analyses . Determining whether the
344 associated benefits and harms have been captured appropriately is therefore not only of clinical
345 importance, but may impact how screening policy is determined or overdiagnosis is conceptualised
346 (20).

347

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

359

360 The heterogeneity in utility values raises the question of what economic measure should be used, or
361 whether health related quality of life is suitable for measuring outcomes associated with screening and

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

377

378 The clinical outcomes associated with breast cancer screening programmes are widely contested, yet
379 the benefits and harms of mammography are inadequately appraised in the economic literature
380 informing the debate. Few studies identified by this review integrated all relevant phases of care
381 associated with breast cancer screening into the assessment of quality of life, and the values used were
382 limited in their ability to truly capture the disutility. Thirteen studies included overdiagnosis in their
383 evaluation but applied the same utilities for diagnosis and treatment as a non-overdiagnosed
384 cancer (118-120), even if the costs and quality of life losses were ultimately not necessary or entirely
385 representative. Estimates of screen detected overdiagnosis vary significantly from 0 to 54% (23, 27).
386 Whilst several economic evaluations cite the lack of published utilities as a justification for not
387 including screening or overdiagnosis in their analysis (29, 49, 52, 54), ignoring this harm may
388 inadvertently lead to inappropriate advice to women, decisions on the value of screening programmes
389 and potential misallocation of resources. Similarly, none of the primary studies explicitly considered

390 the impact of unnecessary treatment in their vignettes. The inclusion of overdiagnosis in qualitative
391 descriptions has been shown to change general population preferences toward more conservative
392 management or surveillance strategies(121, 122). The limitations of the economic measures and
393 health states outlined in this review (123) raises concerns about information asymmetry, and whether
394 women can make an informed decision about screening without information on the full benefits and
395 risks. Any potential advantages and risks should be explicitly listed within the descriptions of relevant
396 health states. These findings are not limited to breast cancer; appraising the impact of unnecessary
397 treatment may be relevant in other public interventions such as prostate cancer (120), cervical
398 screening (124) or the management of cardiovascular disease, where treatments reduce the risk of
399 future morbidity and mortality but have side-effects (125). Indeed, a number of cancer screening
400 initiatives have reported varying outcomes when different sets of utilities are assumed (119, 126).

401

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

414

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

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

423

424 **Strengths and limitations**

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

437

438 **Implications**

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

446 personalised, risk-stratified utilities for active monitoring strategies will be required in breast cancer
447 screening models.

448

449 **Recommendations for future research**

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

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

473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500

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 sequelae, 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.

501 **References**

- 502 1. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the
503 economic evaluation of health care programmes: Oxford university press; 2015.
- 504 2. Excellence NfHaC. Guide to the methods of technology appraisal. London: NICE; 2013.
- 505 3. Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation
506 in health care. *Annu Rev Public Health*. 2000;21:587-611.
- 507 4. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities.
508 *British medical bulletin*. 2010;96(1):5-21.
- 509 5. Tolley K. What are health utilities. Hayward Medical Communications, London. 2009.
- 510 6. Torrance GW. Measurement of health state utilities for economic appraisal: a review. *Journal*
511 *of health economics*. 1986;5(1):1-30.
- 512 7. Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods.
513 *Journal of health economics*. 1996;15(2):209-31.
- 514 8. Bass EB, Steinberg EP, Pitt HA, Griffiths RI, Lillemoe KD, Saba GP, et al. Comparison of
515 the rating scale and the standard gamble in measuring patient preferences for outcomes of gallstone
516 disease. *Medical Decision Making*. 1994;14(4):307-14.
- 517 9. Dolan P. Modeling valuations for EuroQol health states. *Medical care*. 1997:1095-108.
- 518 10. Kind P, Hardman G, Macran S. UK population norms for EQ-5D. 1999.
- 519 11. Papaioannou D, Brazier J, Paisley S. Systematic Searching and Selection of Health State
520 Utility Values from the Literature. *Value in Health*. 2013;16(4):686-95.
- 521 12. Ara R, Brazier JE. Populating an economic model with health state utility values: moving
522 toward better practice. *Value in Health*. 2010;13(5):509-18.
- 523 13. Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD. Model
524 parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good
525 Research Practices Task Force Working Group–6. *Medical decision making*. 2012;32(5):722-32.
- 526 14. Jackson LJ, Auguste P, Low N, Roberts TE. Valuing the health states associated with
527 Chlamydia trachomatis infections and their sequelae: a systematic review of economic evaluations
528 and primary studies. *Value in Health*. 2014;17(1):116-30.

- 529 15. Gerard K, Johnston K, Brown J. The role of a pre-scored multi-attribute health
530 classification measure in validating condition-specific health state descriptions. *Health economics*.
531 1999;8(8):685-99.
- 532 16. Johnston K, Brown J, Gerard K, O'Hanlon M, Morton A. Valuing temporary and chronic
533 health states associated with breast screening. *Social Science & Medicine*. 1998;47(2):213-22.
- 534 17. Earle C, Chapman R, Baker C, Bell C, Stone P, Sandberg E, et al. Systematic overview of
535 cost-utility assessments in oncology. *Journal of Clinical Oncology*. 2000;18(18):3302-17.
- 536 18. Brown ML, Finton L. Cost-effectiveness of breast cancer screening: preliminary results of a
537 systematic review of the literature. *Breast cancer research and treatment*. 1993;25(2):113-8.
- 538 19. Schiller-Fruhworth IC, Jahn B, Arvandi M, Siebert U. Cost-Effectiveness Models in Breast
539 Cancer Screening in the General Population: A Systematic Review. *Applied Health Economics &*
540 *Health Policy*. 2017;15(3):333-51.
- 541 20. Marmot MG, Altman D, Cameron D, Dewar J, Thompson S, Wilcox M. The benefits and
542 harms of breast cancer screening: an independent review. *British journal of cancer*.
543 2013;108(11):2205.
- 544 21. Myers ER, Moorman P, Gierisch JM, Havrilesky LJ, Grimm LJ, Ghate S, et al. Benefits and
545 harms of breast cancer screening: a systematic review. *Jama*. 2015;314(15):1615-34.
- 546 22. Løberg M, Lousdal ML, Bretthauer M, Kalager M. Benefits and harms of mammography
547 screening. *Breast Cancer Research : BCR*. 2015;17(1):63.
- 548 23. Nelson HD, Fu R, Cantor A, Pappas M, Daeges M, Humphrey L. Effectiveness of breast
549 cancer screening: systematic review and meta-analysis to update the 2009 US Preventive Services
550 Task Force Recommendation. *Annals of internal medicine*. 2016;164(4):244-55.
- 551 24. Welch HG, Passow HJ. Quantifying the benefits and harms of screening mammography.
552 *JAMA internal medicine*. 2014;174(3):448-54.
- 553 25. Francis A, Thomas J, Fallowfield L, Wallis M, Bartlett JM, Brookes C, et al. Addressing
554 overtreatment of screen detected DCIS; the LORIS trial. *European journal of cancer*.
555 2015;51(16):2296-303.

- 556 26. Groen EJ, Elshof LE, Visser LL, Emiel JT, Winter-Warnars HA, Lips EH, et al. Finding the
557 balance between over-and under-treatment of ductal carcinoma in situ (DCIS). *The Breast*.
558 2017;31:274-83.
- 559 27. Puliti D, Duffy SW, Miccinesi G, De Koning H, Lynge E, Zappa M, et al. Overdiagnosis in
560 mammographic screening for breast cancer in Europe: a literature review. *Journal of medical*
561 *screening*. 2012;19(1_suppl):42-56.
- 562 28. Badia X, Herdman M, Dipstat MR, Ohinmaa A. Feasibility and validity of the VAS and TTO
563 for eliciting general population values for temporary health states: a comparative study. *Health*
564 *Services and Outcomes Research Methodology*. 2001;2(1):51-65.
- 565 29. Raftery J, Chorozoglou M. Possible net harms of breast cancer screening: updated modelling
566 of Forrest report. *Bmj*. 2011;343:d7627.
- 567 30. Hersch J, Jansen J, Barratt A, Irwig L, Houssami N, Howard K, et al. Women's views on
568 overdiagnosis in breast cancer screening: a qualitative study. *Bmj*. 2013;346:f158.
- 569 31. Partridge A, Adloff K, Blood E, Dees EC, Kaelin C, Golshan M, et al. Risk perceptions and
570 psychosocial outcomes of women with ductal carcinoma in situ: longitudinal results from a cohort
571 study. *Journal of the National Cancer Institute*. 2008;100(4):243-51.
- 572 32. De Haes J, de Koning HJ, van Oortmarssen GJ, van Agt HM, de Bruyn AE, van der Maas PJ.
573 The impact of a breast cancer screening programme on quality-adjusted life-years. *International*
574 *journal of cancer*. 1991;49(4):538-44.
- 575 33. Weitzner MA, Meyers CA, Stuebing KK, Saleeba AK. Relationship between quality of life
576 and mood in long-term survivors of breast cancer treated with mastectomy. *Support Care Cancer*.
577 1997;5(3):241-8.
- 578 34. Tosteson AN, Fryback DG, Hammond CS, Hanna LG, Grove MR, Brown M, et al.
579 Consequences of false-positive screening mammograms. *JAMA internal medicine*. 2014;174(6):954-
580 61.
- 581 35. Wright DR, Wittenberg E, Swan JS, Miksad RA, Prosser LA. Methods for measuring
582 temporary health states for cost-utility analyses. *Pharmacoeconomics*. 2009;27(9):713-23.

- 583 36. Brazier JE, Dixon S, Ratcliffe J. The role of patient preferences in cost-effectiveness analysis.
584 Pharmacoeconomics. 2009;27(9):705-12.
- 585 37. Smith RA. Expert Group: IARC Handbooks of Cancer Prevention. Vol. 7: Breast Cancer
586 Screening. Lyon, France: IARC; 2002. 248pp. Breast Cancer Research. 2003;5(4):216.
- 587 38. Brazier J, Deverill M. A checklist for judging preference-based measures of health related
588 quality of life: learning from psychometrics. Health economics. 1999;8(1):41-51.
- 589 39. Dissemination C. CRD's Guidance for Undertaking Reviews in Health Care. Layerthorpe:
590 University of York, Centre for Reviews and Dissemination. 2009.
- 591 40. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA
592 statement for reporting systematic reviews and meta-analyses of studies that evaluate health care
593 interventions: explanation and elaboration. PLoS medicine. 2009;6(7):e1000100.
- 594 41. Peasgood T, Ward SE, Brazier J. Health-state utility values in breast cancer. Expert Rev
595 Pharmacoecon Outcomes Res. 2010;10(5):553-66.
- 596 42. Endnote X. Referencing software. In: Analytics C, editor. Boston, MA 2018.
- 597 43. Higgins JP, Altman DG. Assessing risk of bias in included studies. Cochrane handbook for
598 systematic reviews of interventions: Cochrane book series. 2008:187-241.
- 599 44. McDonagh M, Peterson K, Raina P, Chang S, Shekelle P. Avoiding bias in selecting studies.
600 2013.
- 601 45. Ogwulu CB, Jackson LJ, Kinghorn P, Roberts TE. A systematic review of the techniques
602 used to value temporary health states. Value in Health. 2017;20(8):1180-97.
- 603 46. Ahern CH, Shen Y. Cost-effectiveness analysis of mammography and clinical breast
604 examination strategies: a comparison with current guidelines. Cancer Epidemiology and Prevention
605 Biomarkers. 2009;18(3):718-25.
- 606 47. Arrospide A, Rue M, van Ravesteyn NT, Comas M, Soto-Gordoa M, Sarriguarte G, et al.
607 Economic evaluation of the breast cancer screening programme in the Basque Country: retrospective
608 cost-effectiveness and budget impact analysis. BMC cancer. 2016;16(1):344.

- 609 48. Beemsterboer PM, De Koning HJ, Warmerdam PG, Boer R, Swart E, Dierks ML, et al.
610 Prediction of the effects and costs of breast cancer screening in Germany. *International journal of*
611 *cancer*. 1994;58(5):623-8.
- 612 49. Carles M, Vilapriño E, Cots F, Gregori A, Pla R, Román R, et al. Cost-effectiveness of early
613 detection of breast cancer in Catalonia (Spain). *Bmc Cancer*. 2011;11(1):192.
- 614 50. Christensen MK, Niclasen BV, Iburg KM. Societal costs and effects of implementing
615 population-based mammography screening in Greenland. *International Journal of Circumpolar*
616 *Health*. 2017;76(1):1373580.
- 617 51. de Koning HJ, Martin van Ineveld B, van Oortmarsen GJ, De Haes J, Collette HJ, Hendriks
618 JH, et al. Breast cancer screening and cost-effectiveness; policy alternatives, quality of life
619 considerations and the possible impact of uncertain factors. *International Journal of Cancer*.
620 1991;49(4):531-7.
- 621 52. Forrest AP. Breast cancer screening : report to the Health Ministers of England, Wales,
622 Scotland & Northern Ireland. In: Forrest P, editor. London :: H.M.S.O.; 1986.
- 623 53. Haghghat S, Akbari ME, Yavari P, Javanbakht M, Ghaffari S. Cost-effectiveness of three
624 rounds of mammography breast cancer screening in Iranian women. *Iranian journal of cancer*
625 *prevention*. 2016;9(1).
- 626 54. Hristova L, Hakama M. Effect of screening for cancer in the Nordic countries on deaths, cost
627 and quality of life up to the year 2017. *Acta Oncologica (Stockholm, Sweden)*. 1997;36:1-60.
- 628 55. BreastScreen Australia Evaluation. Economic evaluation and modelling study. Australia:
629 Australian Government Department of Health and Ageing; 2009.
- 630 56. Mittmann N, Stout NK, Tosteson ANA, Trentham-Dietz A, Alagoz O, Yaffe MJ. Cost-
631 effectiveness of mammography from a publicly funded health care system perspective. *CMAJ Open*.
632 2018;6(1):E77-E86.
- 633 57. Morton R, Sayma M, Sura MS. Economic analysis of the breast cancer screening program
634 used by the UK NHS: should the program be maintained? *Breast Cancer (Dove Med Press)*.
635 2017;9:217-25.

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

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

641 60. Rojnik K, Naveršnik K, Mateović R, Rojnik T, Primic Žakelj M. Probabilistic cost-effectiveness
642 effectiveness modeling of different breast cancer screening policies in Slovenia. *Value in health*.
643 2008;11(2):139-48.

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

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

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

652 64. Tosteson AN, Stout NK, Fryback DG, Acharyya S, Herman BA, Hannah LG, et al. Cost-
653 effectiveness of digital mammography breast cancer screening. *Annals of internal medicine*.
654 2008;148(1):1-10.

655 65. Trentham-Dietz A, Kerlikowske K, Stout NK, Miglioretti DL, Schechter CB, Ergun MA, et
656 al. Tailoring breast cancer screening intervals by breast density and risk for women aged 50 years or
657 older: collaborative modeling of screening outcomes. *Annals of internal medicine*. 2016;165(10):700-
658 12.

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

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

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

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

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

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

676 72. Mandelblatt JS, Schechter CB, Yabroff KR, Lawrence W, Dignam J, Extermann M, et al.
677 Toward optimal screening strategies for older women. *Journal of general internal medicine*.
678 2005;20(6):487-96.

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

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

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

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

688 77. Salzmann P, Kerlikowske K, Phillips K. Cost-effectiveness of extending screening
689 mammography guidelines to include women 40 to 49 years of age. *Annals of Internal Medicine*.
690 1997;127(11):955-65.

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

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

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

699 81. Vilapriyo E, Forne C, Carles M, Sala M, Pla R, Castells X, et al. Cost-effectiveness and
700 harm-benefit analyses of risk-based screening strategies for breast cancer. PloS one.
701 2014;9(2):e86858.

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

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

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

711 85. Oliver A. Should we maximise QALYs? Applied health economics and health policy.
712 2004;3(2):61-6.

713 86. McNamee P, Glendinning S, Shenfine J, Steen N, Griffin SM, Bond J. Chained time trade-off
714 and standard gamble methods. The European Journal of Health Economics, formerly: HEPAC.
715 2004;5(1):81-6.

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

- 719 88. Tengs TO, Wallace A. One thousand health-related quality-of-life estimates. *Medical care*.
720 2000;583-637.
- 721 89. Bonomi AE, Boudreau DM, Fishman PA, Ludman E, Mohelnitzky A, Cannon EA, et al.
722 Quality of life valuations of mammography screening. *Quality of Life Research*. 2008;17(5):801-14.
- 723 90. Chie W-C, Huang C-S, Chen J-H, Chang K-J. Utility Assessment for Different Clinical
724 Phases of Breast Cancer in Taiwan. *Journal of the Formosan Medical Association*. 2000;99(9):677-83.
- 725 91. Gerard K, Dobson M, Hall J. Framing and labelling effects in health descriptions: quality
726 adjusted life years for treatment of breast cancer. *Journal of Clinical Epidemiology*. 1993;46(1):77-84.
- 727 92. Hall J, Gerard K, Salkeld G, Richardson J. A cost utility analysis of mammography screening
728 in Australia. *Social Science & Medicine*. 1992;34(9):993-1004.
- 729 93. Kim SH, Jo MW, Ock M, Lee HJ, Lee JW. Estimation of health state utilities in breast cancer.
730 *Patient Preference and Adherence*. 2017;11:531-6.
- 731 94. Rijnsburger A, Essink-Bot M-L, van Dooren S, Borsboom G, Seynaeve C, Bartels C, et al.
732 Impact of screening for breast cancer in high-risk women on health-related quality of life. *British*
733 *journal of cancer*. 2004;91(1):69.
- 734 95. Dolan P. Developing methods that really do value the 'Q' in the QALY. *Health economics,*
735 *policy and law*. 2008;3(1):69-77.
- 736 96. Lamers LM, Stalmeier PF, Krabbe PF, Busschbach JJ. Inconsistencies in TTO and VAS
737 values for EQ-5D health states. *Medical Decision Making*. 2006;26(2):173-81.
- 738 97. Furlong W, Feeny D, Torrance G, Barr R, Horsman J. Guide to design and development of
739 health-state utility instrumentation. Centre for Health Economics and Policy Analysis (CHEPA),
740 McMaster University, Hamilton, Canada; 1992.
- 741 98. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. Measuring and valuing health benefits for
742 economic evaluation: OXFORD university press; 2017.
- 743 99. Longworth L, Rowen D. Mapping to obtain EQ-5D utility values for use in NICE health
744 technology assessments. *Value in Health*. 2013;16(1):202-10.
- 745 100. Von Neuman J, Morgenstern O. *Theory of games and economic behavior*. 1953.
- 746 101. Tolley KH. No title. What are health utilities? What is series, 2009. 2012.

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

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

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

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

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

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

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

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

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

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

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

774 113. Brazier J. Valuing health states for use in cost-effectiveness analysis. *Pharmacoeconomics*.
775 2008;26(9):769-79.

776 114. Torrance GW, Feeny D, Furlong W. Visual analog scales: do they have a role in the
777 measurement of preferences for health states? : Sage Publications Sage CA: Thousand Oaks, CA;
778 2001.

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

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

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

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

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

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

792 121. McCaffery K, Nickel B, Moynihan R, Hersch J, Teixeira-Pinto A, Irwig L, et al. How
793 different terminology for ductal carcinoma in situ impacts women's concern and treatment
794 preferences: a randomised comparison within a national community survey. *BMJ open*.
795 2015;5(11):e008094.

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

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

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

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

806 126. Keller A, Gericke C, Whitty JA, Yaxley J, Kua B, Coughlin G. A Cost-Utility Analysis of
807 Prostate Cancer Screening in Australia. *Applied Health Economics and Health Policy*. 2017;15(1):95-
808 111.

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

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

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

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

819 131. Hayman JA, Kabeto MU, Schipper MJ, Bennett JE, Vicini FA, Pierce LJ. Assessing the
820 benefit of radiation therapy after breast-conserving surgery for ductal carcinoma-in-situ. *J Clin Oncol*.
821 2005;23(22):5171-7.

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

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

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

828 135. Ara R, Brazier JE. Using health state utility values from the general population to
829 approximate baselines in decision analytic models when condition-specific data are not available.
830 Value in Health. 2011;14(4):539-45.

831 136. Lidgren M, Wilking N, Jönsson B, Rehnberg C. Health related quality of life in different
832 states of breast cancer. Quality of life research. 2007;16(6):1073-81.

833 137. Souza FH. Mamografia digital em comparação com mamografia convencional no
834 rastreamento de câncer de mama no Brasil: revisão sistemática, custo da doença e análise de custo-
835 efetividade no Sistema Único de Saúde. 2012.

836 138. Cruz LN. Medidas de qualidade de vida e utilidade em uma amostra da população de Porto
837 Alegre. 2010.

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854 **Figures and Tables**

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

856

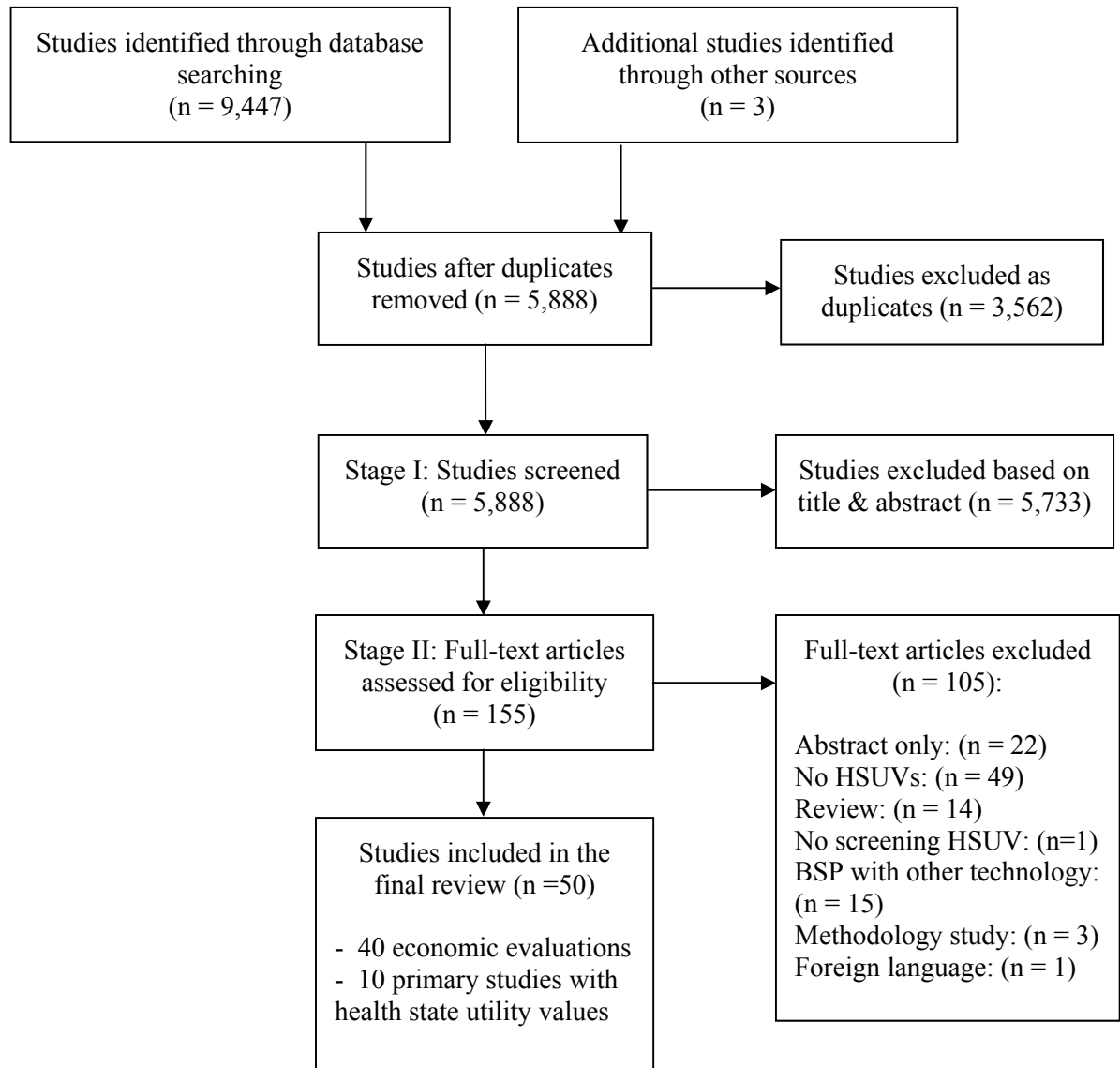


Table 1: Characteristics of the 40 economic evaluations

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 expectancy	(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)
Messeccar (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 Luijck (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)
Vilapriyo (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

