VALUE IN HEALTH 20 (2017) 1000–1002

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

# The HIQA's Health Technology Assessment of Breast Screening: Highlighting Some of the Challenges Posed by Evaluations of Screening Programs

*Conor Teljeur, PhD*[1,2,*], *Patrick Moran, MSc*[1], *Patricia Harrington, PhD*[1], *Máirín Ryan, PhD*[1]

[1]*Health Information and Quality Authority, Ireland, Trinity College, Dublin, Ireland;* [2]*Department of Public Health and Primary Care, Trinity College, Dublin, Ireland*

ABSTRACT

A review of the Health Information and Quality Authority's (HIQA) assessment of breast cancer surveillance cancer criticized how the results were presented and interpreted. The health technology assessment (HTA) investigated surveillance options for women aged less than 50 years who were at elevated risk of breast cancer. Surveillance strategies using digital mammography, magnetic resonance imaging, or a combination of the two modalities were modeled on the basis of diagnostic test accuracy. The HTA faced a number of issues, including the use of diagnostic test accuracy as a surrogate for long-term outcomes; modeling interventions that were not considered acceptable to clinicians; and extrapolating for screening intervals and age ranges not directly supported by available evidence. The evaluation of screening programs gives rise to challenges in terms of balancing an adequate exploration of the possible options while also being cognizant of what is appropriately supported by evidence. In this article, the authors of the HIQA report discuss the challenges posed by this particular HTA.

***Keywords:*** breast cancer, diagnostic test efficacy, health technology assessment, screening, surveillance.



## Introduction

A review of a recent national health technology assessment (HTA) of breast cancer surveillance options for women at elevated risk of breast cancer highlighted that cost-effectiveness was reported relative to a common comparator rather than relative to the theoretical next best option [1]. The manner in which cost-effectiveness is reported has implications for how it is interpreted, and it may have consequences for the efficient use of resources.

In seeking to bridge the gap between the evidence and the needs of the decision maker, HTA must cross the sometimes gaping chasm between theory and practice. In the case of the Health Information and Quality Authority's (HIQA) breast cancer surveillance HTA, a number of methodological and practical challenges were faced [2]. This article highlights some of those challenges, and as the authors of the HIQA report, we discuss the context of the HTA and why there is sometimes a divergence between theory and practice in the evaluation of screening interventions.

## Description of the HTA

HIQA is a state agency in Ireland with a statutory remit to carry out HTA to support national-level policy decisions. The breast cancer surveillance HTA was carried out in response to a request from the National Cancer Control Programme [2]. The target population was women aged less than 50 years with an elevated risk of breast cancer due to genetic predisposition. Breast cancer in this population is considered to be somewhat different to that in the general population, with tumors being more aggressive and more difficult to detect. The context of the HTA was that women aged less than 50 years with an elevated risk of breast cancer were being provided with surveillance on an ad hoc basis, with variability in the imaging modality, frequency, and start age observed across individuals with the same level of risk, thereby leading to a mixture of over- and undersurveillance.

The primary concern of the screening service was standardizing surveillance within the given budget while at minimum maintaining current effectiveness. The stated criteria for suitable surveillance strategies were as follows: they should be at least as effective as and no more costly than existing surveillance; be on or close to the cost-effectiveness frontier; and demonstrate a mortality reduction relative to no surveillance. The HTA considered the full spectrum of domains: technology description, clinical effectiveness and safety, economic evaluation, social and organization features, and ethical and legal issues.

The HTA advised that women with BRCA1 and BRCA2 mutations should receive annual magnetic resonance imaging (MRI)

* *Address correspondence to:* Conor Teljeur, Health Information and Quality Authority, George's Court, George's Lane, Dublin 7, Ireland. E-mail: cteljeur@hiqa.ie.

1098-3015$36.00 – see front matter Copyright © 2017, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2016.11.024

beginning at age 30 years and combined MRI and digital mammography annually from age 40 years. Women with other high penetrance genetic mutations should receive annual MRI alone from an age of 30 years, with the exception of women with TP53 mutations, who should be provided with annual MRI alone from age 20 years. Women at high and moderate familial risk of breast cancer should receive annual digital mammography from age 40 years, although limited potential for mortality reduction or quality-adjusted life-year gain relative to no surveillance was noted for both groups. The suggested surveillance options were very similar to those included in the National Institute for Health and Care Excellence guidance in the United Kingdom [3].

Given the paucity of data available on the patient cohort, it is not surprising that there was considerable uncertainty in the results. The uncertainty in effectiveness was driven primarily by uncertainty in the probability of developing cancer and the probability of cancer-related mortality by stage at diagnosis.

## Issues Faced in the HTA

The criticism of O'Mahony and Normand centers on the presentation of average cost-effectiveness ratios (ACERs), or incremental cost-effectiveness relative to a common comparator, in the HIQA report [1]. As stated previously, however, although we also presented the incremental cost-effectiveness ratios relevant to the next best alternative, we were explicit that an a priori stated criterion was for suitable strategies to be at least as effective as and no more costly than the common comparator of existing surveillance. Comparison to the next best alternative assumes that all comparators are equally applicable.

We now consider three issues that introduced considerable complexity into the interpretation of the economic evaluation component of the HTA.

### *Diagnostic Test Accuracy As Outcome*

In the breast cancer surveillance HTA, diagnostic test accuracy was used to model long-term outcomes. This is a common approach when evaluating screening interventions [4]. In Thornbury and Fryback's model of diagnostic efficacy, diagnostic test accuracy sits on the second lowest tier, below diagnostic impact, therapeutic impact, patient outcomes (e.g., mortality or functional outcomes), and societal outcomes [5]. In the case of cancer, earlier diagnosis is usually assumed to result in improved survival. Breast cancer screening has been particularly controversial due to the risks of overtreatment and repeated exposure to x-rays [6].

Clinical diagnosis and treatment often depend on factors other than the results of a single diagnostic test (e.g., clinical history and additional testing) [7]. A model based on diagnostic test accuracy alone entails a number of assumptions about the equivalence of diagnosis. We assumed that the imaging modality would not affect the timing or choice of treatment. Different tests with the same accuracy, however, may detect quite different cases and may result in different patient outcomes [8]. For example, MRI and digital mammography may perform differently in the diagnosis of benign and invasive tumors. Using diagnostic test accuracy as a surrogate for mortality in the breast cancer surveillance HTA created uncertainty that could not be readily quantified due to the absence of long-term outcome data to either calibrate or validate the model.

Ideally, long-term outcomes would be used rather than the surrogate outcome of diagnostic test accuracy [9,10]. Indeed, some HTA agencies do not accept diagnostic test accuracy as evidence of clinical effectiveness due to the uncertain association with outcomes. Although there is evidence regarding mortality reductions associated with mammography-based surveillance of women at elevated risk of breast cancer, there is scant evidence regarding an MRI-based surveillance program [11].

### *Modeling Otherwise Unsupported Interventions*

In theory, all relevant comparators for which there is adequate evidence should be included in an analysis [12]. The reality is that not all options are necessarily supported by equivalent or even adequate quality evidence. The comparators chosen for the breast cancer surveillance HTA were based on a review of the available evidence and expert clinical advice. Clinical and patient input is critical to the interface between the assessors and service provision and provides essential insight into the feasibility and acceptability of different comparators in practice.

The evaluation of breast cancer surveillance included a number of strategies that were not clinically acceptable for all subgroups. For example, although MRI has typically only been considered as an adjunct to mammography, the assessment included a range of MRI-only strategies. Due to the effects of ionizing radiation, MRI-only prior to age 40 years was considered acceptable. With the exception of women at the highest risk of developing breast cancer (i.e., those with a TP53 mutation), however, MRI-only surveillance from age 40 to 49 years was not acceptable to clinicians due to the aforementioned lack of long-term outcome data. MRI-only strategies for women aged 40 years and over could have been identified at the outset as unacceptable and excluded from the analysis. Their inclusion, however, highlighted the fact that, on the basis of a surrogate outcome, they may be superior to a mammography-only approach.

Including alternatives that may not be clinically acceptable created a difficulty for presenting the findings. On the basis of diagnostic test accuracy, MRI-only strategies were deemed to be more efficient than those combined with mammography, even though they were not considered clinically acceptable. A balance must be struck between being overly restrictive in the selection of comparators and providing useful information that may identify alternatives that are theoretically more efficient. Had we excluded those strategies, the cost-effectiveness plane would have been substantially different.

As with all research, the limitations must be considered when interpreting the results. Screening, like vaccination, targets those who are currently well and may never develop the target condition. Because of this, the bar for the evidence base is often set higher due to the differing risk balance. While a decision-analysis model combining data from multiple sources and a number of simplifying assumptions may have indicated that MRI-only screening may be more effective than mammography, this is no substitute for primary studies demonstrating improvements in disease-specific morbidity and mortality.

### *Extrapolation from Available Data*

The evidence for both mammography and MRI was largely based on 12-month screening intervals. The applicability of the data to other screening intervals was dependent on rates of disease progression and whether the diagnostic accuracy of the test is sensitive to the disease spectrum. The HTA used evidence from the Breast Cancer Surveillance Consortium [13] that suggested that digital mammography in women aged less than 50 years who were at average risk had increased test sensitivity with longer intervals between screens whereas specificity remained unchanged.

It is not uncommon for economic evaluations of screening programs to assess a range of screening intervals and age ranges. Using appropriate data, such analyses are beneficial and can lead to screening programs with a better balance between benefits and harms. If the screening frequency modeled is substantially different from that used in the evidence underpinning diagnostic

accuracy or efficacy, however, it would raise serious concerns about the validity of the findings. An assessment that recommends screening strategies for which the only evidence of efficacy or effectiveness has been generated by a modeling exercise lacks external validity and therefore may be of little or no use to a decision maker and undermine the credibility of HTA.

## Discussion

This article describes the issues encountered in the HIQA HTA of breast cancer surveillance and how they highlight the complexities of the assessment of screening programs. The intention of the HTA was to provide evidence to support the standardization of an ad hoc surveillance service and to select strategies that were at least as effective and within the overall budget of existing surveillance.

Although the theoretical framework within which the HTA was carried out provides a basis for conducting empirical analyses, its limitations must be recognized to avoid drawing conclusions that could lead to poor policymaking. In particular, there was a dearth of information on the target population and long-term outcome data for the surveillance strategies modeled. Also of note, some of the strategies modeled were not considered clinically acceptable, creating challenges for the presentation of results and formulation of advice. This is an important issue that has consequences for the interpretation of cost-effectiveness and receives little coverage in the literature.

Without long-term follow-up data, it is not possible to state that screening leads to improved patient-relevant outcomes. There is a lack of consensus as to what level of evidence is required to support decisions regarding screening technologies, with some agencies requiring long-term outcomes as evidence of diagnostic test efficacy. In many cases, only the surrogate outcome of diagnostic test accuracy is available: a stringent approach means that potentially life-saving screening programs may not be considered due to a lack of follow-up data. A less stringent approach, however, could result in the adoption of screening programs for which the only evidence of effectiveness is based on modeled outcomes. In reporting HTAs of screening strategies, the quality and quantity of evidence supporting each strategy should be abundantly clear.

Breast cancer surveillance using MRI as an adjunct to mammography is becoming increasingly common. For women at elevated risk and aged less than 40 years, MRI is often offered as the risks associated with mammography are considered to outweigh the benefits. It is possible that a randomized controlled trial of MRI-based surveillance will never be carried out, in which case, longer-term outcomes, such as survival, will only ever be estimated using observational data that may be at substantial risk of bias.

The use of surrogate outcomes supported by assumptions regarding applicability must necessarily introduce marked uncertainty into economic evaluations. Although there are a number of well-established methods for conveying uncertainty within CEA, the impact of using surrogate outcomes, and the implications this has for the external validity of the results obtained for competing strategies, must be thoroughly explored. The inclusion of a cost-effectiveness acceptability curve or sensitivity analysis in a CEA does not ensure that the decision maker appreciates the magnitude or sources of uncertainty. Value of information analyses may assist in identifying important sources of uncertainty, although it is unclear that policymakers are typically willing to commission further research to reduce decision uncertainty.

Although the underlying issues highlighted in this article are generic, the extent to which they may arise in HTAs of other screening programs will be context specific. For a different evaluation there may, for example, be substantial long-term evidence of reduced morbidity and mortality. Other issues not highlighted here may also affect the accuracy of models of screening programs, such as adherence to surveillance, timely treatment, and adverse events. Failure to account for these issues in economic models can lead to inaccurate estimates of cost-effectiveness.

## Conclusions

Ideally, all interventions included in an HTA are feasible and are based on the same high-quality evidence of patient-relevant outcomes. Overreliance on modeling and extrapolation from surrogate outcomes could lead to the recommendation of unsupported strategies over those for which there is evidence. Adopting a pragmatic balance enables the inclusion of potentially useful interventions for which there is limited evidence while simultaneously acknowledging the limitations that exist to present policymakers with comprehensive, fair, and balanced information to inform decision making.

Given the breadth of domains included in a typical HTA, the recommendations of an HTA are the synthesis of all the evidence targeted to match the decision maker's criteria rather than only interpretation of the cost-effectiveness analysis. The transparency of HTA means that all of the evidence and methodology is typically reported and open to scrutiny, which is to be welcomed.

## Acknowledgments

We would like to thank Prof. Ciaran O'Neill for his helpful comments on this article.

Source of financial support: No financial support or funding was received regarding this article or associated work.

## REFERENCES

[1] O'Mahony J, Normand C. HIQA's CEA of breast screening: pragmatic policy recommendations are welcome, but ACERs reported as ICERs are not. Value Health 2015;18:941–5.
[2] Health Information and Quality Authority. Health Technology Assessment of Surveillance of Women Aged Less than 50 Years at Elevated Risk of Breast Cancer. Dublin, Ireland: Health Information and Quality Authority; 2013.
[3] National Institute for Health and Care Excellence. Familial Breast Cancer: Classification, Care and Managing Breast Cancer and Related Risks in People with a Family History of Breast Cancer (CG164). London, United Kingdom: NICE; 2013.
[4] Staub LP, Dyer S, Lord SJ, Simes RJ. Linking the evidence: intermediate outcomes in medical test assessments. Int J Technol Assess Health Care 2012;28:52–8.
[5] Thornbury JR. Eugene W Caldwell Lecture. Clinical efficacy of diagnostic imaging: love it or leave it. AJR Am J Roentgenol 1994;162:1–8.
[6] Berry DA. Breast cancer screening: controversy of impact. Breast 2013;22(Suppl 2):S73–6.
[7] Broeze KA, Opmeer BC, van der Veen F, et al. Individual patient data meta-analysis: a promising approach for evidence synthesis in reproductive medicine. Hum Reprod Update 2010;16:561–7.
[8] Kuhl CK, Schrading S, Bieling HB, et al. MRI for diagnosis of pure ductal carcinoma in situ: a prospective observational study. Lancet 2007;370:485–92.
[9] Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, et al. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. BMJ 2012;344:e686.
[10] Pavlovic M, Teljeur C, Wieseler B, et al. Endpoints for relative effectiveness assessment (REA) of pharmaceuticals. Int J Technol Assess Health Care 2014;30:508–13.
[11] Bick U. Intensified surveillance for early detection of breast cancer in high-risk patients. Breast Care (Basel) 2015;10:13–20.
[12] Health Information and Quality Authority. Guidelines for the Economic Evaluation of Health Technologies in Ireland. Dublin, Ireland: Health Information and Quality Authority; 2014.
[13] Ballard-Barbash R, Taplin SH, Yankaskas BC, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. AJR Am J Roentgenol 1997;169:1001–8.