ECONOMIC EVALUATION OF HER2 TARGETED BREAST CANCER THERAPY

AN ECONOMIC EVALUATION OF ALTERNATIVE TEST-TREAT STRATEGIES TO DIRECT HER2 TARGETED BREAST CANCER TREATMENT BASED ON CANADIAN PRACTICE PATTERNS

By

ILIA LIN FERRUSI, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Doctor of Philosophy

DESCRIPTIVE NOTE

McMaster University DOCTOR OF PHILOSOPY (2014) Hamilton, Ontario

Faculty of Health Sciences, Department of Clinical Epidemiology and Biostatistics: Health Research Methodology – Health Technology Assessment

TITLE: An economic evaluation of alternative test-treat strategies to direct HER2 targeted breast cancer treatment based on Canadian practice patterns.

AUTHOR: Ilia Lin Ferrusi B.Sc. (McMaster University)

SUPERVISOR: Associate Professor Deborah A Marshall, B.Sc., MHSA, Ph.D.

NUMBER OF PAGES: xv, 281

ABSTRACT

Background and Objectives: Economic evaluation and decision analysis provide a framework to evaluate incremental costs and effects associated with alternative health interventions. These methods can also be used as a tool to evaluate alternative clinical behaviours or practice patterns. The objective of this thesis was to investigate the impact of current Canadian practices in human epidermal growth factor receptor-2 (HER2) testing to target trastuzumab in early-stage breast cancer (BC).

Methods:

Project 1**:** A systematic review of previous trastuzumab and HER2 testing economic analyses was conducted to identify methodological gaps and key lessons.

Project 2**:** A population-level, retrospective cohort was studied to determine HER2 testing and trastuzumab treatment patterns in Ontario early-stage BC patients.

Project 3**:** A cost-utility analysis of alternative test-treat strategies was conducted using a Markov model of BC calibrated to the Canadian setting, and incorporating Project 2 findings.

Results:

Project 1**:** Previous economic evaluations demonstrated that HER2 test accuracy and sequencing were key considerations when modelling the cost-effectiveness of trastuzumab treatment. Consideration of local testing and treatment practices was lacking.

Project 2**:** HER2 testing and treatment practice differed from guidelines, where documentation was available. Only 88% of equivocal results were confirmed, while 57% of HER2 positive patients received trastuzumab.

Project 3**:** Calibration of the BC model minimised gaps between trial-based survival and expected Canadian survival patterns. Deviations from guidelines in practice suggest that primary testing with fluorescence *in situ* hybridization (FISH) would produce greater health gains at a reduced cost vs. primary immunohistochemistry with FISH confirmation. This finding was more apparent as the prevalence of HER2 positive disease increased. Introduction of newer *in situ* hybridisation tests may be cost-effective as well.

Conclusions: Practice deviations from guidelines are an important consideration when modelling the cost-effectiveness of trastuzumab therapy. Underlying local disease progression and prevalence can also significantly impact outcomes.

ACKNOWLEDGEMENTS

Though it may have seemed that this day would never come, it is here at long last. There are many people to whom I will be forever indebted for this achievement, individuals who have provided support, critical guidance and mentorship beyond expectation.

I must start by thanking my supervisor, Dr. Deborah Marshall. Thank you for taking a chance on me, for your patience with me, for knowing when to nudge me and when to provide support. You have mentored me through challenges that we never expected to face, and I am a better researcher for it. You have exemplified the very definition of perseverance. I have learned so much more than health outcomes research methods from your tutelage! I hope that you have learned as much from me as I have from you.

I also owe significant gratitude to my committee members, Prof. Ron Goeree, Dr. Natasha Leighl, and, Dr. Eleanor Pullenayegum. Each of you provided me with critical intellectual insight through various stages of this work. And, each of you provided encouragement in your own unique way, and those encouragements helped to keep me going. I have had the great fortune to work with many other accomplished outcomes researchers throughout this thesis, and those exposures enlightened my education. Thank you to Drs. Craig Earle, Jeffrey Hoch, Kathryn

Phillips, and Maureen Trudeau for your mentorship. I hope that I can work with you all again in the future.

To my dear friends – thank you for helping to keep me sane! I actually have several friendships that are a direct result of this degree. To my colleagues, Nathalie Kulin and Ruthanne Cameron, I simply couldn't have made it without you! You both taught me that "this too, shall pass." To my fellow HRM students, Natalia Diaz-Granados, Morgan Lim, Sameer Parpia, and Nathan Souza, thank you for the late nights, the laughs, and the reassurance. I have incredible admiration for each of you, and I look forward to seeing the research that you will produce in the years to come.

My fiancé, Dr. Arkan Abadi, is a gentleman and a scholar. Thank you, my dear, for always believing in me, for picking me up when I thought I couldn't do anymore, for providing me with perspective, and for being my shoulder. You are my inspiration!

Lastly, and certainly not least of all, I thank my extraordinary parents. I am everything I am today because of them and their sacrifices.

Thank you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

CHAPTER 4

LIST OF ABBREVIATIONS

+	positive (in reference to status or test result)
-	negative (in reference to status or test result)
AC-T	doxorubicin + cyclophosphamide followed by taxol (chemotherapy regimen)
AJCC	American Joint Committee on Cancer
ALK	anaplastic lymphoma kinase (oncogene)
BC	breast cancer
BCS	breast conserving surgery
BCIRG	Breast Cancer International Research Group
BRAF	v-Raf murine sarcoma viral oncogene homolog B1
CAP	College of American Pathologists
CCO	Cancer Care Ontario
CEA	cost-effectiveness analysis
CEAC	cost-effectiveness acceptability curve
CHF	congestive heart failure
CI	confidence interval
CISH	chromogenic *in situ* hybridisation
CT	computed tomography (scan)
CTX	chemotherapy
DAD	Discharge Abstract Database
DC	early therapy discontinuation
DF	degrees of freedom
DFS	disease-free survival
DR	distant recurrence
EBCTCG	Early Breast Cancer Trialists' Collaborative Group
ECHO	echocardiography
EGFR	epidermal growth factor receptor
ER	estrogen receptor
FISH	fluorescence *in situ* hybridisation
FinHer	Finnish Herceptin trial

FP	family practitioner
GoF	goodness of fit
GP	general practitioner
HER2	human epidermal growth factor receptor-2
HER2+	HER2 positive
HER2-	HER2 negative
HERA	Herceptin Adjuvant trial
HR	hazard ratio
HTA	health technology assessment
ICER	incremental cost effectiveness ratio
ICUR	incremental cost utility ratio
ICES	Institute for Clinical Evaluative Sciences
IF	Ilia Ferrusi
IHC	immunohistochemistry
IKN	ICES key number
ITT	intention-to-treat
KRAS	Kirsten rat sarcoma (viral oncogene)
LHIN	Local Health Integration Network
LL	log likelihood
LN	lymph node
LR	likelihood ratio
LRR	locoregional recurrence
LVEF	left ventricular ejection fraction
LY	life year
MOHLTC	Ministry of Health and Long-Term Care
MT	Dr. Maureen Trudeau
MUGA	multi-gated acquisition scanning
NACRS	National Ambulatory Care Reporting System
NCCTG	North Central Cancer Treatment Group
NDFP	New Drug Funding Program
NED	no evidence of disease
NIH	National Institutes of Health
NL	Dr. Natasha Leighl
NSABP	National Surgical Adjuvant Breast & Bowel Project

NYHA	New York Heart Association
OCR	Ontario Cancer Registry
OHIP	Ontario Health Insurance Program
OICR	Ontario Institute for Cancer Research
OLIS	Ontario Laboratory Information System
OR	odds ratio
OS	overall survival
p(DR|LRR)	probability of having a distant recurrence given that the patient hashed a locoregional recurrence
p(DR|recur)	probability of a recurrence being distant from the no evidence of disease state
p(LRR)	post-locoregional recurrence
p(recur|NED)	probability of having a local or distant recurrence from the no evidence of disease state
PIMS	Pathology Information Management System
PM	personalized medicine
PR	progesterone receptor
QALY	quality adjusted life year
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
SISH	silver *in situ* hybridisation
SDS	Same Day Surgery
SE	standard error
TNM	Tumour Node Metastasis (staging system)
μ	mean probability

PREFACE

This thesis is a ‘sandwich thesis, combining three individual projects with contributions that are already published or are in submission to peer-reviewed journals. The first three papers are published, and the final paper is under review. As such, the reference lists for these papers follow different formatting from the remaining unpublished content. Similarly, all papers are published or submitted to US journals requiring American English spelling of some words compared to the Canadian English spelling used throughout the unpublished work.

DECLARATION OF ACADEMIC ACHIEVEMENT

Ilia Lin Ferrusi contributed to all papers and unpublished work in this thesis, including: development of research questions, development of methodological approaches, performing analyses, interpretation of results, writing of all manuscripts, and, submitting manuscripts to peer-reviewed journals.

CHAPTER 1

INTRODUCTION

1.1DISEASE BACKGROUND

Breast cancer (BC) is a disease of the ductal or lobular breast tissues that is more common in women than men. Cancerous cells are characterised by unregulated growth, leading to the formation of malignant tumours that withdraw vital resources from adjacent tissues. Cancer cells can spread from the incident tumour to nearby or distant tissues via the lymphatic or circulatory systems. The brain, liver, bone and soft tissues are the most common sites of distant metastasis in BC.[1] This characteristic unchecked growth is largely attributed to mutation(s) in genes (termed oncogenes) related to the regulation of cell growth and death.[2,3] Although the exact cause of all BC cases may not be known, dietary, environmental and hereditary factors are believed to be the major causes of genetic mutation leading to tumourigenesis.[4]

Breast cancer is the most common cancer among Canadian women, representing 26% of all new cancer cases in 2012.[5] This trend is consistent across the country, with one in nine Canadian women expected to develop BC in her lifetime.[5] Incidence rates in Canada rose from the early 1980s through the 1990s; this increase was attributed to increased mammography screening and, to a lesser extent, hormonal replacement therapy for peri- and post-menopausal women.[5] Incidence rates have been relatively stable since the mid-2000s.[5]

Increased mammography screening is also credited with earlier diagnosis and a subsequent reduction in mortality, although it is also associated with in-

creased over- diagnosis and over- treatment.[6] In total, BC mortality has decreased by almost 40% from its peak in 1986.[5] Mortality due to BC is the second most common cause of cancer death in Canadian women and represents 14% of all cancer deaths.[5] However, mortality rates are largely dependent on age and disease severity at diagnosis.[5] Disease severity is classified using a system which accounts for tumour size, degree of lymph node involvement and presence of metastasis to distant tissue(s), known as the TNM system.[7] Indeed, the five-year survival among Canadian BC patients diagnosed between 1994 and 1997 was 96%, 86%, 59% and 26% for those diagnosed at stage I, II, III and IV, respectively.[8]

1.2 DIAGNOSIS & TREATMENT OF BREAST CANCER

An overview of the diagnosis and treatment of early-stage BC is provided in Figure 1. The standard of BC screening in Canada is mammography, recommended every two to three years for women with an average risk of disease aged 50 to 74.[9] This recommendation was based on significant reductions in the relative risk of death observed with mammographic screening vs. no screening.[9] Women at higher risk of developing BC, including those with a recent family history, are eligible for more frequent screening and genetic testing. Current guidelines do not endorse screening by magnetic resonance imaging, breast self-examination or clinical examination due poor quality evidence.[9] Lumps identified by mammography are then biopsied for pathological examination to determine malignancy.

Primary treatment of invasive early-stage BC (stage I, II, III) consists of surgery to remove the malignancy and any local metastases to the lymph nodes.[10] Surgical treatment entails radical or modified radical mastectomy, which remove all breast tissue, or breast conserving procedures such as lumpectomy, which remove only the affected portion of the breast tissue.[10] Although early studies suggested a survival advantage for mastectomy over breast conserving options, more recent evidence suggests no difference in survival with breast conserving treatment when followed by radiotherapy to the conserved breast. [10,11] Tumour specimens removed during surgery undergo pathological workup to confirm invasive disease, determine the presence of any metastases to local lymph nodes, and the molecular characteristics of the primary tumour. These molecular features, such as tumour grade and hormone receptor status, are used in conjunction with the TNM staging system to to estimate the patient's risk of recurrence. Patients are categorised as being either at low, intermediate or high risk of experiencing recurrence within 10 years of initial diagnosis (Figure 1). [12]

Adjuvant treatment following primary surgery seeks to eliminate any cancerous cells that were not physically removed. Treatment recommendations are based on risk of recurrence. Local radiation therapy may be advised to shrink large tumours prior to surgical resection, or postoperatively in low or intermedicate risk patients treated with lumpectomy.[13] Systemic pharmaceutical intervention with individual or combinations of cytotoxic compounds, termed chemother-

apy (CTX), is often recommended for intermediate and high risk tumours. Treatment with such pharmaceutical agents, including anthracyclines (e.g. doxorubicin, epirubicin), taxanes (e.g. paclitaxel, docetaxel), alkylating agents (e.g. cyclophosphamide, carboplatin) and anti-metabolites (e.g. fluorouracil, methotrexate) seeks to disrupt rapidly reproducing cancer cells, inducing cell death.[14] Tumours that overexpress estrogen or progesterone receptors can also be treated with endocrine therapy, either receptor blocking agents (e.g. tamoxifen) or aromatase inhibitors (e.g. exemestane, anastrazole).[14–16] Systemic chemotherapy is typically recommended before surgery (neoadjuvant) to improve the resectability of large tumours of those attached to adjacent tissues, or following surgery as an adjuvant treatment for intermediate and high-risk patients.[12,17] Combination chemotherapy significantly improves rates of recurrence and mortality at 15 years follow-up for women age 50-69 years, producing an absolute reduction of 4.1% and 3.0% vs. no adjuvant chemotherapy, respectively.[18] Although these improvements are significant, a large gap in treatment efficacy remains, with approximately 50% of patients experiencing a recurrence in their lifetime. Moreover, the indiscriminate nature of cytotoxic therapy also results in the death of healthy cells undergoing rapid reproduction, such as bone marrow. Chemotherapy and endocrine therapy can cause an array of mild to severe side effects (e.g. alopecia, asthenia, cardiac dysfunction, cognitive dysfunction, myelosuppression, nausea, neuropathy and neutropenia).[14] Reductions in quality of life are well-documented for both

acute and long-term treatment-related side effects.[19,20] Thus, clinicians and patients must make treatment decisions by balancing the ratio of potential benefits and risks of therapy without certainty that the patient's tumour will respond. This variability in response drives the search for new BC treatments that prolong life while minimising reductions in quality of life.

1.3 PERSONALISED MEDICINE: A NEW TREATMENT PARADIGM

Recent advances in molecular and genomic research methods have facilitated a rapid growth in the scientific community's understanding of the genetic variability between tumours. This has shed light on the unpredictability of tumour response to standard therapy, and contributed to the development of the modern personalised medicine (PM) treatment paradigm. Personalised medicine is defined by the National Institutes of Health (NIH) as 'an emerging practice of medicine that uses an individual's genetic profile to guide decisions made in regard to the prevention, diagnosis, and treatment of disease.'[21] As such, it is our understanding of the genomic markers of an individual, or their tumour, that can be used to determine whether a patient might respond to therapy or the appropriate dosage. This paradigm offers great potential by increasing the likelihood that patients will be treated with effective medication, thus minimising unnecessary exposure to ineffective therapy and related side effects. Healthcare system productivity could be improved with PM by increasing the health outcomes obtained per

dollar spent on treatment, while also reducing the costs associated with negative side effects or ineffective therapy. However, PMs can only be effective if provided to patients with the correct phenotype or genotype, and can therefore only be successful in the clinical setting if accompanied by an appropriate diagnostic test.

Trastuzumab (Herceptin) is among the first successful examples of PM in cancer. A 1987 retrospective study identified a subset of BC patients whose tumours overexpressed the human epidermal growth factor recpetor-2 gene (HER2/*neu)* and tended to have higher rates of relapse and poorer survival outcomes.[22] Subsequent studies confirmed that HER2 protein overexpression was also linked to poorer outcomes in BC, including higher tumour grade and earlier metastasis.[23] The aggressive nature of HER2-positive BC is thought to be related to the overexpression of HER2 protein receptors on the surface of the tumour cell, which initiates a signalling cascade promoting cell proliferation, invasion, angiogenesis and inhibition of apoptosis.[24] Moreover, HER2-positive tumours have demonstrated insensitivity to standard chemotherapy, particularly hormone-targeted therapies.[25] The trastuzumab monoclonal antibody was subsequently developed to specifically target and block signalling via the HER2 protein in BC tumour cells.[26]

Early trials in metastatic BC demonstrated significant improvements in pathological response and overall survival vs. standard chemotherapy alone.[27] Trastuzumab was incorporated into routine Canadian clinical practice following

market authorisation and public reimbursement in 1998. Subsequent studies in the adjuvant setting demonstrated substantial improvements in disease-free survival (DFS) and overall survival (OS) sequential or concurrent to traditional chemotherapy. After a median follow-up of 36 months, adjuvant trastuzumab was associated with a hazard ratio (HR) of 0.66 (CI: 0.57, 0.77; $p<0.00001$) and 0.60 (CI: 0.50, 0.71; $p<0.00001$) for DFS and OS outcomes with one year of therapy, respectively.[28] A small proportion (2.6%) of patients experienced symptomatic congestive heart failure. This adverse effect was expected based on metastatic studies, but is important to minimise given the long life expectancy of early-stage BC patients.

As the first commercially successful PM, HER2-targeted therapy with trastuzumab necessitates molecular testing for the HER2/*neu* gene or HER2 protein to correctly identify treatment candidates. Two testing modalities, fluorescence *in situ* hybridisation (FISH) for HER2/*neu* detection, and immunohistochemistry (IHC) for HER2 protein detection, are the dominant methods of HER2 status determination in Canadian practice. Testing by IHC experienced greater uptake owing to its simpler methodology, readily available equipment in pathology laboratories, and lower costs.[29] However, IHC testing is also associated with poorer sensitivity and specificity relative to FISH. Canadian guidelines recommend that patients whose tumours are weakly positive for HER2 amplification by IHC, termed equivocal, should be retested with FISH to clarify HER2 status.[30]

Laboratory investigations have demonstrated significant variability in test sensitivity and specificity across lab settings [31], prompting the development of guidelines to encourage quality assurance programs, appropriate tissue fixation, testing practice and use of confirmatory testing.[30,31] Testing uptake was also inconsistent in the early days of trastuzumab therapy. Studies in the metastatic setting suggested that as many as 52% of eligible patients did not receive HER2 testing.[32] These challenges in optimising test accuracy, sequencing and access suggest that trastuzumab may not be used to optimally in the clinical setting. Elkin *et al.* [33] demonstrated the importance of test sequencing to direct trastuzumab in the metastatic setting using a decision-analytic framework. For example, limited access to testing can lead to under treatment and unnecessary morbidity and mortality for HER2 positive patients, while inappropriate testing can lead to inaccurate diagnoses resulting in both under and over treatment. Such challenges undermine the potential clinical effectiveness and incremental cost-effectiveness of trastuzumab within the PM paradigm.

Other personalised cancer medicines have followed trastuzumab's clinical success with biomarker-targeted therapy [34]:

- cetuximab and panitumab, antibodies that target epidermal growth factor receptor (EGFR) in colon cancer tumours free of the KRAS genetic mutation;

- crizotinib, an antibody that inhibits tyrosine kinase in non-small cell lung cancers with an ALK mutation;
- erlotinib and gefitinib, antibodies that inhibit EGFR tyrosine kinase in non-small cell lung cancer;
- vemurafenib, an antibody that inhibits mutataed BRAF in melanoma;

The clinical and cost-effectiveness of these PMs are directly related to the accuracy of companion genomic biomarker test(s). Although these medicines are in various stages of clinical development, their evaluation and implementation in practice will benefit from the lessons learned from HER2-targeted trastuzumab therapy in BC.

1.4 PERSONALISED MEDICINE IN THE ERA OF 'REAL WORLD DATA'-DRIVEN POLICY

Health policy and research agendas have recently shifted towards a pragmatic approach that more closely reflects actual practice.[35–40] Many terms are used to describe this approach including, including, coverage with evidence development [38], comparative effectiveness research[40], field evaluation [39,41] and 'real world' data or trials [35,40,42,43]. Some jurisdictions have the ability to link health technology reimbursement policy to the collection of additional pragmatic data, including access or coverage with evidence development [38,44], knowledge-

brokering [45], and risk-sharing [40,44] programs. The shift was urged by decision-makers who struggled to interpret health technology assessments (HTA) and develop health policy using clinical and cost-effectiveness evidence based on placebo or single comparator trials in highly selected patient populations. What decision-makers lacked was evidence of how new health technologies might function in their patient population and healthcare setting.[35,40,46] Thus, observational and experimental clinical research methods that emphasize the diversity of the patient population under the conditions of actual practice were prioritised by policy makers and research agencies.

The PM approach to therapy can reduce some clinical and policy uncertainty when a population of known responders is identified by genomic testing. However there is still much to be learned about the effectiveness of PMs using pragmatic research methods. Important questions about the accuracy of testing, test sequencing and treatment decisions based on test results need to be answered in the context of actual practice. These factors can influence rates of over and under treatment. In the context of cost-effectiveness analysis, over and under treatment tend to increase the estimated incremental cost per life year (LY) or quality-adjusted LY (QALY) gained. Pragmatic research approaches can therefore facilitate our understanding of the joint effects of test accuracy, test sequencing and treatment decisions in conjunction with clinical trial estimates of drug efficacy within decision-analytic framework.

1.5 THESIS AIM & RESEARCH QUESTION

Although several health economic evaluations of trastuzumab treatment in BC are published, none reflect local practice patterns. This research makes a unique contribution to the literature by documenting local patterns of practice, and by estimating the cost-effectiveness of HER2 testing and adjuvant trastuzumab treatment in Canada subsequent to those patterns of practice. The objective of this study is to estimate the incremental cost-effectiveness of alternative HER2 testing strategies to direct adjuvant trastuzumab treatment in early-stage BC patients in Canada based on current patterns of testing and treatment practice. We hypothesize that the scenario analyses will reveal that current deviations from testing and treatment result in higher incremental cost-effectiveness ratios due to higher rates of over and under treatment vs. guideline-adherence behaviour. Thus, this work will make a unique contribution to the literature by quantifying the impact that deviations from guidelines can have on outcomes and costs.

The literature on trastuzumab therapy cost-effectiveness was reviewed systematically (Chapter 2) to identify key methodological and knowledge gaps to guide the current study. Knowledge gaps specific to local patterns of practice were identified and subsequently shaped the design of a retrospective cohort study to understand how HER2 testing and trastuzumab treatment are conducted in current practice. Patterns of HER2 test access, test utilisation and sequencing, and trastuzumab use based on test results were explored in a population-level cohort

of Ontario early-stage BC patients diagnosed after trastuzumab approval in the adjuvant setting (Chapter 3) and informed key transition probabilities in a cost-effectiveness analysis (Chapter 4).

The systematic literature review (Chapter 2) also identified gaps in economic evaluation methods, particularly concerning the *joint evaluation* of HER2 test characteristics and treatment efficacy. This finding informed the design of a decision-analytic model to estimate the incremental cost-effectiveness of alternative HER2 testing strategies to direct trastuzumab in the adjuvant setting (Chapter 4). The model considered testing strategies advocated by current guidelines as well as newer chromogenic and silver *in situ* hybridisation (CISH; SISH) alternatives. These novel tests may offer advantages over IHC or FISH in terms of speed, accuracy, and ability to archive results. [29] The sequelae resulting from treatment assignment and true HER2 status were simulated within a Markov model of early-stage BC, calibrated to reflect the natural history of the disease in Canada. Calibration of the disease model was noted as a key methodological gap in the literature review, further limiting the relevance of model findings to local jurisdictions. The model also accommodated varying patterns of testing and treatment to estimate the impact of current practice on health outcomes and costs. Scenarios assuming practice consistent with guidelines were compared with current practice patterns to elucidate the consequences of deviation from testing or treatment guidelines. The analyses were conducted from the payer perspective to estimate

the lifetime incremental costs and outcomes associated with each test-treat strategy relative to current practice and guideline recommendations.

This sandwich thesis begins with a systematic review of cost-effectiveness analyses of targeted trastuzumab therapy in breast cancer (Chapter 2, Papers 1 and 2). Then, the methods and findings of a population-level retrospective cohort study of Ontario HER2 testing and trastuzumab treatment patterns are presented (Chapter 3, Paper 3). Finally, the findings of the systematic review and retrospective study were incorporated into the cost-effectiveness analysis of HER2 testing strategies to direct targeted trastuzumab therapy in early-stage breast cancer (Chapter 4, Paper 4).

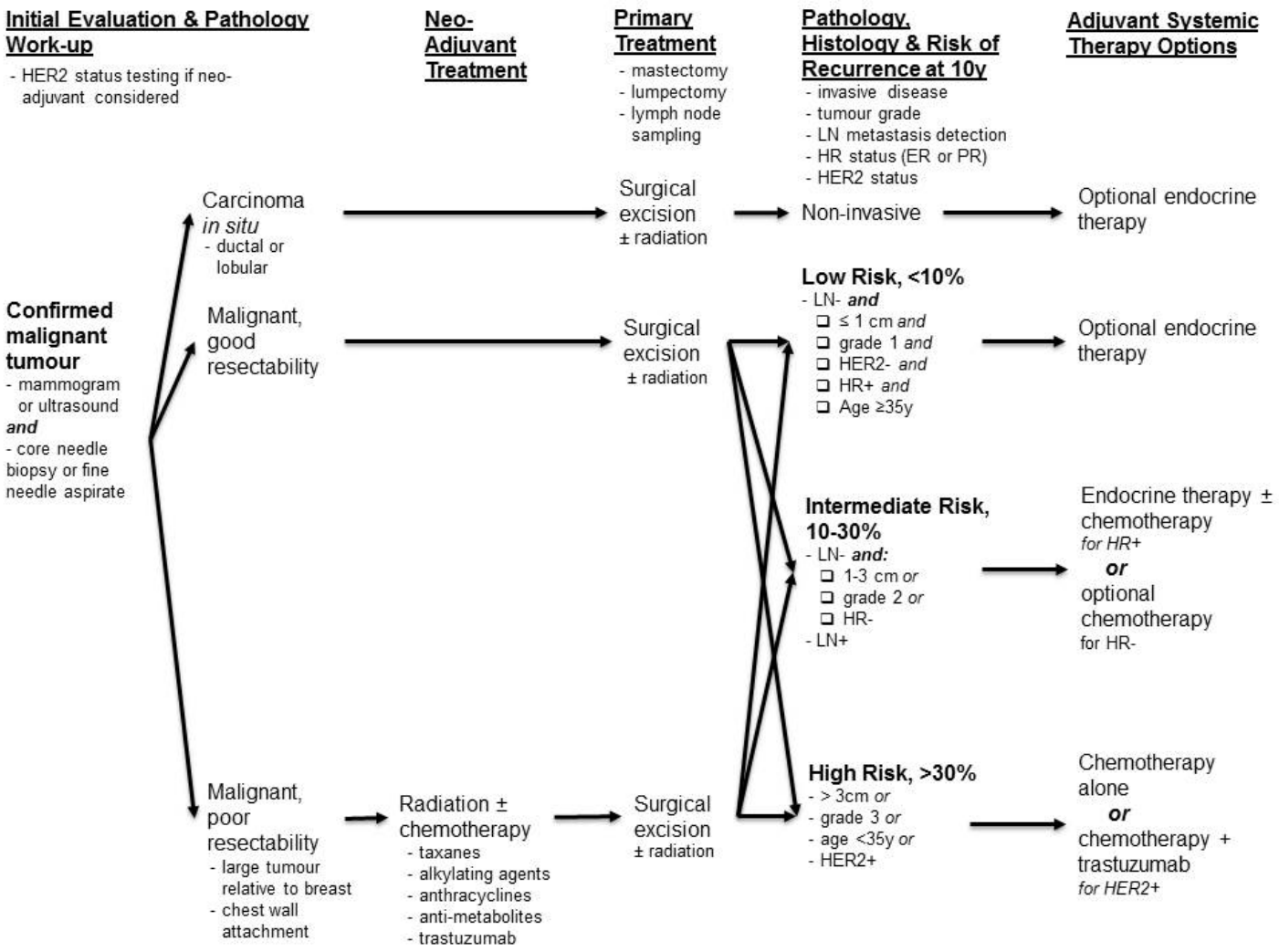

Initial Evaluation & Pathology Work-up
- HER2 status testing if neo-adjuvant considered
Neo-Adjuvant Treatment
Primary Treatment
- mastectomy
- lumpectomy
- lymph node sampling
Pathology, Histology & Risk of Recurrence at 10y
- invasive disease
- tumour grade
- LN metastasis detection
- HR status (ER or PR)
- HER2 status
Adjuvant Systemic Therapy Options
Confirmed malignant tumour
- mammogram or ultrasound
and
- core needle biopsy or fine needle aspirate
Carcinoma in situ
- ductal or lobular
Surgical excision ± radiation
Non-invasive
Optional endocrine therapy
Malignant, good resectability
Surgical excision ± radiation
Low Risk, <10%
- LN- and
❑ ≤ 1 cm and
❑ grade 1 and
❑ HER2- and
❑ HR+ and
❑ Age ≥35y
Optional endocrine therapy
Intermediate Risk, 10-30%
- LN- and:
❑ 1-3 cm or
❑ grade 2 or
❑ HR-
- LN+
Endocrine therapy ± chemotherapy
for HR+
or
optional chemotherapy
for HR-
Malignant, poor resectability
- large tumour relative to breast
- chest wall attachment
Radiation ± chemotherapy
- taxanes
- alkylating agents
- anthracyclines
- anti-metabolites
- trastuzumab
Surgical excision ± radiation
High Risk, >30%
- > 3cm or
- grade 3 or
- age <35y or
- HER2+
Chemotherapy alone
or
chemotherapy + trastuzumab
for HER2+

Figure 1. General overview of the diagnosis and treatment of post-menopausal, non-metastatic (early-stage) breast cancer in Canada. This figure is not an exhaustie representation of risk stratification and all possible treatment recommendations throughout Canada; it was adapted from The St. Gallen algorithm [12], guidelines issued by the Canadian Steering Committee on Clinical Practice Guidelines for the Care and Treatment of Breast Cancer [15–17] and Cancer Care Ontario guidelines [10,13,47] with clinician input (Dr. N. Leighl). Abbreviations: ER: estrogen receptor; HER2: human epidermal growth factor receptor-2; HR: hormone receptor (either estrogen or progesterone); LN: lymph node; PR: progesterone receptor; +: positive; -: negative.

1.6 REFERENCES

1. Clark GM, Sledge GW, Osborne CK, McGuire WL. Survival from first recurrence: relative importance of prognostic factors in 1,015 breast cancer patients. J. Clin. Oncol. 1987;5(1):55–61.
2. Clarke RB, Howell A, Andersen E. Control of proliferation in the normal and neoplastic breast. In: Pasqualini JR, editor. Breast Cancer: Prognosis, Treatment and Prevention. New York: Marcel Dekker; 2002. p xxi, 634 p.:73–91.
3. Formby B, Wiley TS. Inhibition of cell growth and induction of apoptosis. In: Pasqualini JR, editor. Breast Cancer: Prognosis, Treatment and Prevention. New York: Marcel Dekker; 2002. p xxi, 634 p.:125–47.
4. Hecht SS, Hawley Berger A, Pandolfi PP, Howley PM, Ganem D, Kieff E, et al. Etiology of Cancer. In: Rosenberg SA, editor. DeVita, Hellman, Rosenberg's Cancer Principals and Practice of Oncology, 9th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2011. p. xlvii, 2638 p. :150–232.
5. Canadian Cancer Society's Steering Committee on Cancer Statistics. Canadian Cancer Statistics 2012. Toronto: Canadian Cancer Society; 2012 p. 1–69.
6. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. Cochrane Database Syst. Rev. 2013 Jun 4;6:CD001877.
7. Cancer Care Ontario. Guidelines for Staging Patients with Cancer. Toronto: Cancer Care Ontario; 2005. p. 1–3.
8. Canadian Cancer Society, National Cancer Institute of Canada. Canadian Cancer Statistics 2007. Toronto: Canadian Cancer Society; 2007 p. 1–116.
9. Tonelli M, Connor Gorber S, Joffres M, Dickinson J, Singh H, Lewin G, et al. Recommendations on screening for breast cancer in average-risk women aged 40-74 years. Can. Med. Assoc. J. 2011;183(17):1991–2001.
10. Members of the Breast Cancer Disease Site Group. Surgical Management of Early-Stage Invasive Breast Cancer. Brackstone M, Tey R, editors. Toronto: Cancer Care Ontario; 2011. p. 1–34.
11. Clarke M, Collins R, Darby S, Davies C, Elphinstone P, Evans E, et al. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. Lancet. 2005;366(9503):2087–106.
12. Goldhirsch A, Wood WC, Gelber RD, Coates a S, Thürlimann B, Senn H-J. Progress and promise: highlights of the international expert consensus on

the primary therapy of early breast cancer 2007. Ann. Oncol. 2007;18(7):1133–44.

13. Members of the Breast Cancer Disease Site Group. Breast Irradiation in Women with Early Stage Invasive Breast Cancer Following Breast Conserving Surgery. Dayes I, Rumble R, editors. Toronto: Cancer Care Ontario; 2011.
14. Locker GY. Adjuvant systemic therapy of early breast cancer. In: Winchester DJ, Winchester DP, editors. Breast Cancer. Hamilton, Ont.: B.C. Decker; 1999. p. viii, 294 p. :201–18.
15. Levine M. Clinical practice guidelines for the care and treatment of breast cancer: adjuvant systemic therapy for node- positive breast cancer (summary of the 2001 update). Can. Med. Assoc. J. 2001;164(5):644–6.
16. Levine M. Clinical practice guidelines for the care and treatment of breast cancer: adjuvant systemic therapy for node-negative breast cancer (summary of the 2001 update). Can. Med. Assoc. J. 2001;164(2):213.
17. Shenkier T, Weir L. Clinical practice guidelines for the care and treatment of breast cancer: 15. Treatment for women with stage III or locally advanced breast cancer. Can. Med. Assoc. J. 2004;170(6):983–94.
18. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet. 2005;365(9472):1687–717.
19. Kayl AE, Meyers C a. Side-effects of chemotherapy and quality of life in ovarian and breast cancer patients. Curr. Opin. Obstet. Gynecol. 2006;18(1):24–8.
20. Montazeri A. Health-related quality of life in breast cancer patients: a bibliographic review of the literature from 1974 to 2007. J. Exp. Clin. Cancer Res. 2008;27:32.
21. National Library of Medicine. Personalized medicine (definition) [Internet]. Genet. Home Ref. US National Library of Medicine, National Institutes of Health, Department of Health & Human Services; 2013 [cited 2013 Jul 2]. Available from: http://ghr.nlm.nih.gov/glossary=personalizedmedicine
22. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science. 1987;235(4785):177–82.
23. Slamon DJ, Godolphin W, Jones L a, Holt J a, Wong SG, Keith DE, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science. 1989;244(4905):707–12.

24. De Luca A, Carotenuto A, Rachiglio A, Gallo M, Maiello MR, Aldinucci D, et al. The role of the EGFR signaling in tumor microenvironment. J. Cell. Physiol. 2008;214(3):559–67.
25. Pietras RJ, Arboleda J, Reese DM, Wongvipat N, Pegram MD, Ramos L, et al. HER-2 tyrosine kinase pathway targets estrogen receptor and promotes hormone-independent growth in human breast cancer cells. Oncogene. 1995;10(12):2435–46.
26. Hancock MC, Langton BC, Chan T, Toy P, Monahan JJ, Mischak RP, et al. A monoclonal antibody against the c-erbB-2 protein enhances the cytotoxicity of cis-diamminedichloroplatinum against human breast and ovarian tumor cell lines. Cancer Res. 1991;51(17):4575–80.
27. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N. Engl. J. Med. 2001;344(11):783–92.
28. Moja L, Tagliabue L, Balduzzi S, Parmelli E, Pistotti V, Guarneri V, et al. Trastuzumab containing regimens for early breast cancer. Cochrane Breast Cancer Group, editor. Cochrane Database Syst. Rev. 4th ed. Chichester, UK: Cochrane Database of Systematic Reviews; 2012;4:CD006243.
29. Penault-Llorca F, Bilous M, Dowsett M, Hanna W, Osamura RY, Rüschoff J, et al. Emerging technologies for assessing HER2 amplification. Am. J. Clin. Pathol. 2009;132(4):539–48.
30. Hanna W, O'Malley FP, Barnes P, Berendt R, Gaboury L, Magliocco A, et al. Updated recommendations from the Canadian National Consensus Meeting on HER2/neu testing in breast cancer. Curr. Oncol. 2007;14(4):149–53.
31. Wolff AC, Hammond MEH, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. J. Clin. Oncol. 2007;25(1):118–45.
32. Phillips KA, Liang S-Y, Van Bebber S. Challenges to the translation of genomic information into clinical practice and health policy: Utilization, preferences and economic value. Curr. Opin. Mol. Ther. 2008;10(3):260–6.
33. Elkin EB, Weinstein MC, Winer EP, Kuntz KM, Schnitt SJ, Weeks JC. HER-2 testing and trastuzumab therapy for metastatic breast cancer: a cost-effectiveness analysis. J. Clin. Oncol. 2004;22(5):854–63.
34. Duffy MJ, Crown J. Companion Biomarkers: Paving the Pathway to Personalized Treatment for Cancer. Clin. Chem. 2013;10(59):1447–56.

35. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. J. Am. Med. Assoc. 2003;290(12):1624–32.
36. Eggertson L. New network created to assess drug safety. Can. Med. Assoc. J. 2011;183(18):E1293–4.
37. Iglehart JK. Prioritizing comparative-effectiveness research--IOM recommendations. N. Engl. J. Med. 2009;361(4):325–8.
38. Tunis SR, Pearson SD. Coverage options for promising technologies: Medicare's "coverage with evidence development". Health Aff. (Millwood). 25(5):1218–30.
39. Goeree R, Levin L. Building bridges between academic research and policy formulation: the PRUFE framework - an integral part of Ontario's evidence-based HTPA process. Pharmacoeconomics. 2006 ;24(11):1143–56.
40. Chalkidou K, Tunis S, Lopert R, Rochaix L, Sawicki PT, Nasser M, et al. Comparative effectiveness research and evidence-based health policy: experience from four countries. Milbank Q. 2009;87(2):339–67.
41. Levin L, Goeree R, Sikich N, Jorgensen B, Brouwers MC, Easty T, et al. Establishing a comprehensive continuum from an evidentiary base to policy development for health technologies: the Ontario experience. Int. J. Technol. Assess. Health Care. 2007;23(3):299–309.
42. Garrison LP, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. Value Heal. 2007;10(5):326–35.
43. Calvert M, Wood J, Freemantle N. Designing "Real-World" trials to meet the needs of health policy makers at marketing authorization. J. Clin. Epidemiol. 2011;64(7):711–7.
44. McCabe CJ, Stafinski T, Edlin R, Menon D. Access with evidence development schemes: a framework for description and evaluation. Pharmacoeconomics. 2010;28(2):143–52.
45. Van Kammen J, de Savigny D, Sewankambo N. Using knowledge brokering to promote evidence-based policy-making: The need for support structures. Bull. World Health Organ. 2006;84(8):608–12.
46. Goeree R, Levin L, Chandra K, Bowen JM, Blackhouse G, Tarride J-E, et al. Health technology assessment and primary data collection for reducing uncertainty in decision making. J. Am. Coll. Radiol. 2009;6(5):332–42.
47. Trudeau M, Madarnas Y, McCready DR, Pritchard KI, Messersmith H. The Role of Trastuzumab in Adjuvant and Neoadjuvant Therapy in Women

with HER2 / neu-overexpressing Breast Cancer : A Clinical Practice Guideline. Toronto: Cancer Care Ontario; 2007. p. 4–8.

CHAPTER 2

SYSTEMATIC REVIEW OF ECONOMIC ANALYSES OF HER2 TESTING OR TRASTUZUMAB TREATMENT IN BREAST CANCER

PAPER 1[†]**:** Ferrusi, I. L., et al. Looking back at 10 years of trastuzumab therapy: What is the role of HER2 testing? A systematic review of health economic analyses. Personalized Medicine. 2009;6(2):193-215.

†Academic contributions by Ilia L. Ferrusi: concept and design, acquisition of data, analysis and interpretation of data, writing, critical revision and finalisation of the manuscript for important intellectual content.

PAPER 2[✧]**:** Ferrusi, I. L., et al. Do economic evaluations of targeted therapy provide support for decision makers? Am. J Manag. Care 2011;17(suppl 5):SP61-SP70.

✧ Academic contributions by Ilia L. Ferrusi: concept and design, acquisition of data, analysis and interpretation of data, writing, critical revision and finalisation of the manuscript for important intellectual content.

2.1 PREFACE

We conducted a systematic review of economic analyses of HER2 testing or trastuzumab treatment in BC to establish current methodological approaches and opportunities for novel evidence generation. The review identified 17 original, incremental economic evaluations of HER2 testing or trastuzumab treatment in early-stage BC, and revealed several lessons for the conduct of an economic analysis of HER2 testing to direct targeted trastuzumab therapy in a Canadian cohort (Chapter 4). More specifically, we sought to answer two questions to inform subsequent chapters of this thesis: (1) what are the modelling approaches and main conclusions of economic analyses that examined HER2-targeted trastuzumab therapy in BC?; and (2) how have economic analyses of HER2-targeted trastuzumab therapy for early-stage BC met the needs of decision-makers?

The first manuscript (Paper 1) summarises the main findings of economic analyses of targeted trastuzumab therapy in the peer-reviewed literature. Targeted trastuzumab was examined using hypothetical cohort models in several international settings for early-stage and metastatic treatment. Trastuzumab was unanimously considered “cost-effective” in early-stage BC largely due to the substantial gains in life-expectancy experienced by that population. In contrast, the lower gain in life years experienced by metastatic patients led one author to conclude

that trastuzumab was not cost-effective in the metastatic setting. Two of the 17 studies considered alternative test-treat strategies and reported the incremental cost-effectiveness of therapy in the context of HER2 test accuracy, sequencing and reflex testing. Models that considered HER2 testing strategy demonstrated that incremental cost-effectiveness estimates were more sensitive to changes in test characteristics (sensitivity, specificity) than any other model parameter in analyses of trastuzumab treatment in either early-stage or metastatic disease alone (e.g. treatment effect, drug cost, rates of side effects or underlying recurrence and mortality rates). Analyses of testing and treatment concluded that trastuzumab treatment was only "cost-effective" when given to HER2-positive patients identified with testing strategies that minimised false-positive and false-negative diagnoses: (1) re-testing patients with an equivocal or positive IHC result using FISH techniques, or (2) using FISH testing alone to determine HER2 status. This conclusion was driven by the lower sensitivity and specificity of IHC vs. FISH, and the high economic and clinical consequences of treating false-positive patients. Thus, the first paper of this chapter highlights the importance of modelling HER2 test properties and alternative test-treat strategies when evaluating targeted trastuzumab therapy in BC. It also reveals opportunities to add to the scientific literature by examining newer HER2 testing technologies, namely SISH and CISH techniques, which are being adopted in several Canadian provinces.

The second manuscript (Paper 2) focuses on methods used in economic analyses of targeted trastuzumab therapy with a view to meeting the needs of decision-makers. In the absence of empirically validated decision-support methods for economic analysis, decision-maker needs were defined through a review of economic analysis guidelines from Canada, the United Kingdom and the United States. Guidelines consistently emphasized the use of local data to inform model parameters, the use of probabilistic analysis to characterise joint uncertainty in model parameters, and representation of decision uncertainty using diagrams (cost-effectiveness plane plots, tornado diagrams, cost-effectiveness acceptability curves [CEACs]) as approaches to meeting the needs of decision-makers. These criteria were used to review analyses in the primary clinical focus of this dissertation, early-stage BC. Within that context, the second manuscript revealed deficiencies in the literature and opportunities to distinguish this work and improve our understanding cost-effective HER2-targeted therapy. Firstly, few studies modeled local practice patterns or used local inputs outside of cost parameters. This has particular importance for models considering HER2 testing strategies given regional variability in test practice and recent weight given to comparative effectiveness research methods. Therefore, this review directly informs the level of detail in the data collection strategy in Chapter 3, where HER2 documentation and testing practice in Ontario are reported. Secondly, probabilistic analysis of joint parameter uncertainty was reported in two-thirds of reviewed studies. This

reinforced that probabilistic methods were important for this dissertation, particularly to assess the joint uncertainty between test and treatment parameters. Thirdly, decision uncertainty was characterised infrequently and this represents another opportunity to contribute to the literature by including CEACs (Chapter 4).

2.2 PAPER 1

Ferrusi, I. L., et al. Looking back at 10 years of trastuzumab therapy: What is the role of HER2 testing? A systematic review of health economic analyses. Personalized Medicine. 2009;6(2):193-215.

SPECIAL REPORT

Looking back at 10 years of trastuzumab therapy: what is the role of HER2 testing? A systematic review of health economic analyses

Trastuzumab is a targeted therapy for human EGF receptor-2 (HER2)-positive breast cancer. The effectiveness and cost–effectiveness of trastuzumab hinges not only on its clinical efficacy in responding patients, but on the ability to accurately identify appropriate therapeutic candidates. We sought to systematically review the cost–effectiveness of trastuzumab with a focus on the impact of the test(s) used for HER2 diagnosis. Our review included 17 economic evaluations or health technology assessments of trastuzumab therapy or HER2 testing. Trastuzumab was considered cost-effective in all early-stage disease studies, while one author concluded that trastuzumab was not cost-effective for metastatic disease. Only two papers considered the joint effects of test accuracy and sequencing with trastuzumab therapy. These demonstrated that trastuzumab cost–effectiveness is sensitive to HER2-test properties.

KEYWORDS: cost–effectiveness · cost-utility · HER2 · herceptin · sensitivity · specificity · targeted therapy · trastuzumab

Ilia L Ferrusi[1], Deborah A Marshall[2†], Nathalie A Kulin[3], Natasha B Leighl[4] & Kathryn A Phillips[5]

[†]Author for correspondence:
[1]McMaster University, Hamilton, ON, Canada
[2]Centre for Evaluation of Medicines, Department of Clinical Epidemiology & Biostatistics, McMaster University, 105 Main St. E, Level P1, Hamilton, ON L8N 1G6, Canada
Tel.: +1 905 522 1155 ext. 34967;
Fax: +1 905 528 7386;
ferrusil@mcmaster.ca
[3]St Joseph's Hospital, Hamilton, ON, Canada
[4]Princess Margaret Hospital, Toronto, ON, Canada
[5]UCSF Institute for Health Policy Studies & UCSF Comprehensive Cancer Center, CA, USA

Breast cancer is the second most common malignancy and the second most common cause of cancer death in women in the USA. The overexpression of human EGF receptor-2 protein (HER2/*neu* or HER2) is prevalent in approximately 20–30% of breast cancer cases [1]. HER2 positivity is associated with increased rates of tumor growth and post-surgical disease recurrence, reduced survival and poorer response to standard chemotherapy [1,2].

Trastuzumab (Herceptin®) is a breast cancer therapy specifically targeted for women who overexpress HER2, and may only be provided to patients with evidence of HER2 overexpression [3]. It was first approved for the treatment of HER2-positive metastatic breast cancer in 1998 and has been commercially available for 10 years. Treatment with trastuzumab was linked to the concomitantly approved use of two commercially available tests for HER2 determination in that same year. FISH detects *HER2* gene amplification in paraffin-embedded tissue samples. It is considered the gold standard of HER2 testing but is not perfect; it is quite costly at US$300–400 per test and requires skilled laboratory personnel [4]. In contrast, immunohistochemistry (IHC) detects HER2 protein expression and has experienced widespread adoption owing to the ease of test conduct and its lower cost (<US$100). However, IHC is less accurate than FISH due to the susceptibility of proteins to damage when handling paraffin-embedded tissue samples [4]. The interpretation of IHC test results can also be subjective [5], and variations in inter-rater reliability, within and between analytic centers, are well documented [4]. Still, trastuzumab is considered a highly successful example of targeted therapy with worldwide sales in 2007 of US$1.48 billion over the 12 months ending in March 2008 [101]. Its efficacy was recently summarized in a systematic review of adjuvant therapy, and is associated with a combined 0.62 relative risk of disease-free survival and 0.56 relative risk of overall survival compared with those taking no trastuzumab [6].

Countries with single payer systems, such as the UK, have questioned the value of including trastuzumab in their formularies and decision-makers have demanded a better understanding of the clinical and cost–effectiveness of trastuzumab [7,8]. In Canada, patient demand and political pressure were reported to have pushed forward the approval of trastuzumab in the adjuvant setting, ahead of complete efficacy and safety data reports [9]. Other jurisdictions such as New Zealand [10–14] and the USA [15] have also debated the high cost of trastuzumab therapy. Given this decidedly publicized controversy, this paper examines evidence from cost–effectiveness analyses of trastuzumab treatment using systematic review methods. Cost–effectiveness analysis quantifies the consequences of two or more health interventions in terms of both health outcomes and associated costs in an incremental fashion [16]. However,

10.2217/17410541.6.2.193 © 2009 Future Medicine Ltd **Personalized Medicine** (2009) **6**(2), 193–215 ISSN 1741-0541

the effectiveness of trastuzumab as a targeted therapy hinges not only on its clinical efficacy in responding patients, but on the ability of clinicians to accurately identify the appropriate therapeutic candidates. In the case of trastuzumab, it is particularly important to investigate the impact of testing given that two commercially available tests are widely used, allowing different strategies for initial or confirmatory testing. If economic analysis and decision analytic modeling are used to inform decision making, it is important that all relevant variables to that decision are included. For example, the strategy used to identify HER2-positive patients could significantly impact the predicted cost–effectiveness of trastuzumab therapy if the rates of false-positive and false-negative diagnoses are high. However, repeated or confirmatory testing is associated with additional costs that may not justify the incremental gain in accurate diagnoses. Individual models of testing or treatment can isolate the most cost-effective treatment or diagnosis strategies, but will ignore the influence that inaccurate diagnosis has on the incremental cost–effectiveness of therapy. Moreover, decision makers could reach erroneous conclusions if the selection of patients in clinical practice differs from the clinical trial setting. We therefore sought to characterize the cost–effectiveness of trastuzumab in the context of HER2 testing.

Cardiotoxicity has emerged as a rare but serious side-effect in trastuzumab-treated patients [17]. This bears particular clinical importance in the early-stage breast cancer population given that a majority of patients can be expected to experience remission and live full lives. This issue is also particularly important in light of HER2 testing; an inaccurate test result places false-positive patients at a potentially higher risk of developing cardiotoxicity without actually experiencing any therapeutic benefit. Development of symptomatic or asymptomatic heart failure may also affect health-related quality of life. Therefore, we also examined how cardiotoxicity was addressed in the papers identified in the search.

Methods

Data sources & search strategy

A systematic search was undertaken to identify economic analyses and health technology assessments which included an economic analysis of trastuzumab in the treatment of breast cancer or HER2 testing. The search included publications from up to, and including, October 2008 indexed in BIOSIS, Cochrane, CRD, EconLit, the Excerpta Medica Database (EMBASE), the Health Economic Evaluations Database (HEED), MEDLINE and PubMed electronic databases. The BIOSIS database was included specifically due to its coverage of conference proceedings and abstracts. Our search strategy allowed for the inclusion of analyses conducted prior to the 1998 market approval of trastuzumab in the USA.

The search strategy used three filters to identify relevant publications: economic (cost and cost-analysis, health economics, economic evaluation, pharmacoeconomics, cost–effectiveness, cost-benefit, cost-utility); breast cancer (breast tumor, breast carcinoma, breast neoplasm, mammary AND [tumor or carcinoma or neoplasm or cancer]); and trastuzumab or HER2/*neu* (ERBB2 receptor, epidermal growth factor receptor, herceptin, trastuzumab, HER2). A detailed description of the search strategies used in each database is provided in Appendix 1. Due to the high potential for relevance to this review, abstracts from the San Antonio Breast Cancer Symposium (SABCS) published between 2004 and 2007 were hand-searched. Technology appraisals produced by NICE in the UK were also hand-searched as they were published in the English language from a country where the trastuzumab controversy is well-documented. Finally, the reference lists of key topical reviews and retrieved articles were also hand-searched.

Inclusion criteria

Among the English language publications meeting the search strategy criteria were (a citation was included if it met the following criteria):

- A paper detailing original research, defined as any study reporting findings based on a collection of new data or assembly of existing data in a new context with a new result;
- An analysis – an economic evaluation considering two or more alternatives and presenting results in an incremental ratio (this included cost-minimization, cost–effectiveness, cost–utility, cost–benefit and health technology assessments containing any of the aforementioned analyses);
- A study considering breast cancer as the primary focus;
- An analysis evaluating either HER2 testing strategies (consisting of IHC or FISH tests to identify HER2-overexpressing patients) or trastuzumab treatment of HER2-positive breast cancer.

Conference abstracts and presentations were included in the initial phase of the review and authors were contacted to determine whether fully reported results had been published in a peer-reviewed setting. Only peer-reviewed analyses were included in this review.

Study selection

The abstracts and titles of all search hits were reviewed independently by two reviewers (Ilia Ferrusi and Deborah Marshall), and each potentially relevant article was obtained in full for further review. Consensus was reached via discussion between the two reviewers to obtain agreement on any conflict during the title and abstract review process. Cohen's unweighted κ [18] was calculated to reflect the agreement between reviewers in the title and abstract review process. Selection of studies based on full article review was conducted by a single reviewer (Ilia Ferrusi) and verified by an independent reviewer (Deborah Marshall).

Data abstraction

Data abstraction characterized the following features of each economic evaluation:

- The aim of each economic analysis;
- The method of analysis (i.e., cost–effectiveness, cost–utility, cost–benefit or cost-minimization analysis; Markov, decision-tree model etc.);
- The population of interest;
- Testing and treatment alternatives examined;
- How the potential for cardiotoxicity and its management was modeled;
- The primary findings of the analysis (i.e., incremental cost–effectiveness ratio) converted to 2007 US dollars using annual average exchange rates [102–105] and the medical component of the consumer price index [106];
- Characterization of parameter uncertainty;
- The author's conclusions.

Results

Search results

Search results are shown in Figure 1. The search strategy identified 958 relevant citations from the following databases: Cochrane (13), CRD (27), EMBASE (677), HEED (7), MEDLINE (119) and PubMed (115). Searches conducted in BIOSIS and EconLit were not fruitful. Handsearching identified an additional ten citations. Among the 968 total hits, 199 were duplications leaving 769 titles and abstracts for review. A total of 75 citations were selected for full review based on the inclusion criteria (Cohen's unweighted $\kappa = 0.86$), and 58 were excluded after further application of the exclusion criteria. This left 17 studies for abstraction. Among those, 12 considered trastuzumab therapy only, two examined HER2 testing and trastuzumab treatment and three considered HER2 testing exclusively.

Several relevant conference abstracts and letters were identified during the literature search, but personal communication with the authors revealed that full analyses were not yet published in a peer-reviewed journal [19–22].

The majority of models were hypothetical cohort simulations (16/17). Models of trastuzumab efficacy were most sensitive to the drug cost (7/14), duration of survival benefit (6/14), discount rate (5/14) and the relative risk reduction (4/14) associated with trastuzumab. In models that considered HER2 testing, we noted that only models considering test–treat scenarios were sensitive to changes in IHC or FISH sensitivity or specificity (2/5). We also rated study quality per the Drummond criteria, which considers a well-defined study question and perspective, comprehensive description of treatment alternatives, evidence of effectiveness, inclusion of all relevant costs and consequences, appropriate units for costs and consequences, credible valuation of costs and consequences, temporal adjustment, incremental analysis and uncertainty analysis [16]. Each study specified a clear research question, and most (14/17) stated the perspective. While all studies did an excellent job of explaining the treatment alternatives, the reasons for selection and exclusion of other alternatives were generally poorly described. Evidence of effectiveness was only demonstrated in a single study, while the remainder (13/14) employed evidence from randomized, controlled trials to inform trastuzumab efficacy. Elements of costing were difficult to evaluate given the limited space permitted in publications. This element was quite variable in quality, although some authors gave a very thorough and detailed description [23,24]. Discounting was always reported and conducted appropriately per the context of the analysis, although discounting was not warranted in a cross-sectional analysis of trastuzumab and all testing-only analyses. Uncertainty analysis was the most poorly performed or reported element in this review. A total of five of the 17 studies were supported, at least in part, by the drug manufacturer (either Hoffmann-La Roche [NJ, USA] or Genentech

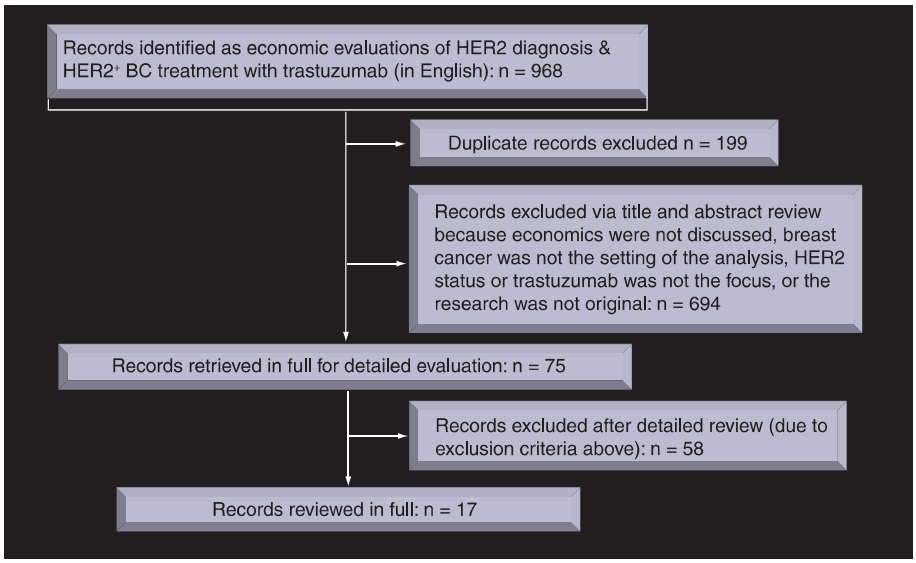


Figure 1. Quorum diagram depicting the flow of studies through the review process.
HER2: Human EGF-2 receptor.

[CA, USA]), while the remaining studies were supported by public funding agencies (6/17) or did not report a funding source (6/17).

Metastatic breast cancer

The four metastatic studies were conducted in European (3/4) and North American settings; these are shown in Table 1. Elkin *et al.* conducted a unique analysis in which the costs and consequences of HER2 testing and treatment were considered [25]. This study examined alternative test–treat strategies whereby patients were tested with IHC only, IHC with FISH confirmation under various circumstances, or FISH alone (see Table 1). It demonstrated that the choice of test–treat strategy has a significant impact on the cost–effectiveness of therapy. By quantifying the additional expense of treating false-positive patients and the opportunity lost by not treating false-negative patients, this study presented crucial evidence of test–treat interdependency on cost–effectiveness. The use of testing and the shorter time horizon of the metastatic analysis is likely to have produced the single largest incremental cost–effectiveness ratio (ICER) for trastuzumab therapy identified in this review at US$153,648–178,231 per quality adjusted life year (QALY) depending on the testing strategy. Norum *et al.* reported a similar ICER (US$92,584–217,264 per life year saved [LYS]) but concluded that trastuzumab was not cost-effective against standard chemotherapy [26]. The analysis by Poncet *et al.* was informed by a pragmatic evaluation of trastuzumab in French hospitals against matched geographic controls [27]. This study predicted a much lower ICER for trastuzumab of US$17,861 per LYS over nontrastuzumab in a 2-year time frame. The authors concluded that trastuzumab was an 'affordable' treatment option. The NICE technology assessment considered trastuzumab in combination and as monotherapy [28]. The assessment of monotherapy modeled an indirect comparison of trastuzumab with vinorelbine, but NICE reviewers were not confident with these findings given the indirect comparison and small size of the trials.

A single metastatic study considered the consequences of cardiotoxicity [25]. Poncet's pragmatic analysis [27] accounted for the exclusion of cardiac toxicities, as none were observed in the study population. Norum, Elkin and Poncet *et al.* included the costs of HER2 testing, and Norum and Elkin both estimated trastuzumab benefit from the same clinical trial [29], but only Elkin modeled test characteristics and repeated the testing. Studies of metastatic breast cancer were only sensitive to estimates of trastuzumab effectiveness when testing was not modeled. In fact, Elkin demonstrated that the cost–effectiveness of trastuzumab in metastatic disease was insensitive to treatment benefit when testing was considered; instead, the ICER was predominantly affected by changes in test properties. Moreover, Elkin revealed that a strategy combined of the two different tests resulted in more accurate patient identification and subsequently better therapeutic outcomes.

Table 1. Summary of economic analyses of trastuzumab in metastatic breast cancer.

Study (Country) (Perspective) (Time horizon) (Discount rate)	Study aim	Methods	Population	Alternatives examined	Results (2007 US$)	Sensitivity analysis	Author's conclusions	Ref.
Poncet *et al.* (2008) (France) (Hospital) (2 years) (None)	To estimate the ICER of trastuzumab + paditaxel as first-line CTX compared to standard CTX using data from a local prospective trial	– Pragmatic open-controlled clinical trial; four hospitals provided trastuzumab + paditaxel, while six control hospitals provided any CTX without trastuzumab – CEA – Threshold analysis also conducted	IHC3+ MBC patients 18 years or older	Trastuzumab + paditaxel versus range of standard CTX chosen at the investigator's discretion	$17,861/LYS	Not conducted	Trastuzumab-associated costs seem affordable within the French healthcare system	[27]
Norum *et al.* (2005) (Norway) (Payer) (Lifetime) (5%)	To compare expected health outcomes of trastuzumab in advanced BC with costs in a model-based CEA	– CEA – Modeling method not stated – Costs of cardiac monitoring and cardiologist visits included	IHC2+ and 3+ MBC patients	Trastuzumab + standard chemotherapy versus standard chemotherapy alone	$92,584–$217,264/LYS depending on survival benefit	– Multivariate examining: HER2 prevalence, IHC costs, outpatient costs, pharmacy administration costs, trastuzumab costs and survival benefit – Survival benefit and drug costs impacted CE	Trastuzumab not CE	[26]

BC: Breast cancer; CE: Cost-effective; CEA: Cost–effectiveness analysis; CUA: Cost–utility analysis; CTX: Chemotherapy; HER2: Human EGF receptor-2; HTA: Health technology assessment; ICER: Incremental cost–effectiveness ratio; IHC: Immunohistochemistry; LYG: Life years gained; LYS: Life years saved; MBC: Metastatic breast cancer; S1–7: Strategy 1–7; QALY: Quality adjusted life year.

Table 1. Summary of economic analyses of trastuzumab in metastatic breast cancer.

Study (Country) (Perspective) (Time horizon) (Discount rate)	Study aim	Methods	Population	Alternatives examined	Results (2007 US$)	Sensitivity analysis	Author's conclusions	Ref.
Elkin *et al.* (2004) (USA) (Not stated, but partial societal costs included) (Lifetime) (3%)	To estimate ICER of alternative testing and trastuzumab treatment strategies in MBC patients	– CUA – Markov model + Monte Carlo simulation – Costs of cardiac monitoring and treatment modeled	Hypothetical cohort of 65-year-old women with newly diagnosed with MBC	– S1: no testing; CTX for all – S2: IHC only; trastuzumab for IHC3$^+$ only – S3: IHC only; trastuzumab for IHC2$^+$ and IHC3$^+$ – S4: IHC, FISH confirmation of IHC2$^+$ and IHC3$^+$; trastuzumab for FISH$^+$ – S5: IHC, FISH confirmation of IHC2$^+$; trastuzumab for FISH$^+$ or IHC3$^+$ – S6: FISH only; trastuzumab for FISH$^+$ – S7: no testing; trastuzumab for all (chemotherapy = paclitaxel)	– Only two ICERS were not dominated: S4: $153,648/QALY for IHC With FISH confirmation of all positives versus S1; S6: $178,231/QALY for initial FISH versus S1 – All other strategies not CE	– Univariate and multivariate – Ranges generally from published literature – Only changes in test characteristics influenced results	It was more CE to use FISH alone or as confirmation of all positive IHC results, rather than using FISH to confirm only weakly positive results or using IHC alone	[25]
NICE (2002) (UK) (NHS) (Lifetime) (3.5%)	To provide guidance on the use of trastuzumab in advanced BC	– CUA within HTA – Systematic review that included data from an unpublished Roche (2000) submission to NICE	MBC patients with IHC3$^+$	– S1: trastuzumab + paclitaxel versus paclitaxel alone – S2: trastuzumab monotherapy versus vinorelbine monotherapy	– S1: $90,118/QALY for trastuzumab in combination with paditaxel – S2: $18,024/LYG for trastuzumab as monotherapy	– Methods not described – Sensitive to survival estimates in combination therapy	– Study S1 was well conducted while study S2 was less so – Trastuzumab recommended as a cost-effective treatment option for MBC	[28]

BC: Breast cancer; CE: Cost-effective; CEA: Cost–effectiveness analysis; CUA: Cost–utility analysis; CTX: Chemotherapy; HER2: Human EGF receptor-2; HTA: Health technology assessment; ICER: Incremental cost–effectiveness ratio; IHC: Immunohistochemistry; LYG: Life years gained; LYS: Life years saved; MBC: Metastatic breast cancer; S1–7: Strategy 1–7; QALY: Quality adjusted life year.

Early-stage breast cancer

A total of eight out of the 10 analyses of trastuzumab in early-stage disease concluded that it was a cost-effective treatment option against comparators, as outlined in TABLE 2. Authors were only uncertain of the cost–effectiveness when 52- and 9-week regimens were compared [23,24,30]. These studies suggested that the ICER for trastuzumab was smaller under the 9-week course, but were reluctant to state this conclusively due to the short-term results and the small size of the Finland Herceptin (FinHer) 9-week trial [1]. Trastuzumab benefit was usually assumed to last for 5 years (6/10) following therapy, while others assumed that benefit occured only for the duration of trial data (2/10) or did not state this assumption (2/10). We found that cost–effectiveness estimates in early-stage disease were more favorable than those in the metastatic breast cancer setting. Some variation was noted across geographical regions; European estimates ranged from $6,783 per life years gained (LYG) to US$65,250/QALY, North American estimates ranged from US$20,065 to US$43,330/QALY, while all other regions ranged from US$14,083/QALY to US$23,309/LYG. When comparing within outcome type, the ranges between estimates were more similar: US$13,361–65,250/QALY and US$6,783–51,976/LYG. Some of this variation can be explained by the choice of trastuzumab regimen. Estimates for 52-week therapy ranged from US$13,361–65,250/QALY, while estimates for the 9-week regimen were US$6,783/LYG to US$14,083/QALY.

Very few early-stage studies assessed the impact of HER2 testing strategies on the cost–effectiveness of trastuzumab, despite the greater potential of the patient population and Elkin's previous findings in the metastatic setting. Lidgren and colleagues conducted the only analysis to consider test accuracy and alternate testing strategies, including confirmatory testing (TABLE 2) [31]. Garrison and colleagues [32] attempted to account for test accuracy by assuming that five tests were performed for every HER2-positive patient, and in estimating test costs, that 30% of tests were FISH. While Garrison's model did not actually model test accuracy, Lidgren's examination of various test–treat strategies demonstrated that the choice of test strategy affected the ICER. Lidgren concluded that FISH testing for all patients was the preferred strategy because it resulted in the greatest gain in QALYs while falling below a willingness-to-pay threshold of $56,116 (€41,500 in 2005). However, if cost–effectiveness is the sole criterion, then confirmatory FISH for IHC2$^+$ and 3$^+$ would be the strategy of choice. In univariate, sensitivity analysis, Lidgren's model was sensitive to the trastuzumab-related risk reduction of an event, the duration of trastuzumab benefit and future costs. However, in a multivariate, sensitivity analysis, the model was sensitive to changes in the sensitivity and specificity of IHC, and the strategy of initial IHC with FISH confirmation of IHC2$^+$ and 3$^+$ was the only nondominated strategy.

Modeled rates of cardiac toxicity varied considerably. Some of this variation can be attributed to the different rates observed in the FinHer and Herceptin Adjuvant (HERA) trials; indeed, models considering FinHer and HERA modeled both rates (2/2). The HERA cardiotoxicity rate of 1.67% was most frequently modeled (5/10); others modeled cardiotoxicity at a rate of 2% (2/10) and rates of 2.9 and 3%. Cardiac side effects were modeled using various methods in all early-stage studies. Most analyses modeled the costs of monitoring (9/10) and treatment (10/10) of cardiotoxicity-related events, such as congestive heart failure, based on the rates observed in clinical trials. Three studies modeled actual health states [33–35] associated with cardiotoxicity, and two overtly assumed that death rates due to cardiac toxicity were negligible. Lidgren's analysis of test–treat strategies was the only one to model a utility decrement for patients developing cardiotoxicity.

Testing

The search strategy included economic analyses of HER2 testing outside the context of trastuzumab treatment (TABLE 3). Testing studies tended to vary widely in methodology and alternatives examined, which we believe is related more to regional practice patterns than was observed in analyses of treatment. For example, both Morelle *et al.* [36] and Dranitsaris *et al.* [37] considered the timing of HER2 testing, although each framed the analysis differently. Morelle's motivation behind the choice of alternative strategies and test timing was driven by the inconsistent practice of paraffin-embedding tissue samples under local recommendations at the time [36]. However, Dranitsaris' cost-minimization analysis sought to determine whether HER2 testing at initial diagnosis in stages I, II or III would save the cost of tissue retrieval when testing at the time of metastatic relapse [37]. This indicates that local practice and test availability are key considerations.

Table 2. Summary of economic analyses of trastuzumab in early-stage breast cancer.

Study (Country) (Perspective) (Time horizon) (Discount rate)	Study aim	Methods	Population	Alternatives examined	Results (2007 US$)	Sensitivity analysis	Author's conclusions	Ref.
Lidgren *et al.* (2008) (Sweden) (Societal) (Lifetime) (3%)	– To carry out a CEA of the addition of 1-year of trastuzumab after adjuvant CTX for patients with early BC – Five testing strategies examined	– CUA – Markov model – 1 year cycles – Monte Carlo simulation – Costs of cardiac monitoring and treatment modeled – Modeled utility decrement for congestive heart failure	Hypothetical cohort of 55-year-old women with completely excised BC, treated with at least four cycles of adjuvant CTX	– S1: no testing, no trastuzumab for all – S2: IHC for all, trastuzumab for IHC3+ – S3: IHC for all, trastuzumab for IHC2+ and 3+ – S4: IHC for all, FISH confirmation of IHC2+ and 3+, trastuzumab for FISH+ – S5: FISH for all, trastuzumab for FISH+	– S2 and 3 ruled out by classic or extended dominance – ICERs per QALY: S4 over S1: $48,645; S5 over S4: $56,077	– Univariate analysis of several model parameters ± 30%; trastuzumab risk reduction, duration of treatment effect and future costs had the greatest impact – Multivariate analysis of IHC sensitivity and specificity; ICER of S5 affected – PSA of costs, utilities and trastuzumab efficacy; CEACs provided	– 1 year adjuvant trastuzumab is a cost-effective treatment option compared to observation – FISH testing for all is a preferable strategy due to higher QALY gains and an ICER below many common WTP thresholds	[31]
Neyt *et al.* (2008) (Belgium) (Payer) (Lifetime) (3 and 1.5%)	To estimate the cost–effectiveness of 9-week (FinHer)* and 1-year (HERA)‡ trastuzumab therapy compared to standard adjuvant treatment	– CEA – Budget impact analysis – Model type not indicated – Incremental risk of heart failure with HERA and FinHer modeled – Costs of cardiac monitoring modeled – Mean cost of heart failure modeled	Hypothetical cohort of HER2+ patients	– S1: 9-week trastuzumab therapy (FinHer) – S2: 1-year trastuzumab therapy (HERA) – S3: standard adjuvant therapy	– S1: $11,719/LYS over standard therapy for all stages of disease; dominant strategy for stage II and III disease – S2: $54,976/LYS over standard therapy for all stages of disease	– Scenario analysis, subgroup analysis and PSA – Model most sensitive to the cost discount rate, progression to metastatic disease and trastuzumab efficacy	After accounting for medical, patient, budgetary and societal 'arguments', FinHer regimen is preferred; large-scale multinational trial results are needed to confirm the benefits of the FinHer trial	[24]

*See ref [1].
‡See ref [43].
BC: Breast cancer; BCIRG: Breast Cancer International Research Group; CE: Cost–effectiveness; CEA: Cost–effectiveness analysis; CEAC: Cost–effectiveness acceptability curve; CUA: Cost–utility analysis; CTX: Chemotherapy; ER: Estrogen receptor; FinHer: Finland Herceptin; FEC: 5-fluorouracil, epirubicin and cyclophosphamide; HERA: Herceptin adjuvant; HTA: Health technology assessment; ICER: Incremental cost–effectiveness ratio; IHC: Immunohistochemistry; LVEF: Left ventricular ejection fraction; LYS: Life years saved; LYG: Life years gained; MBC: Metastatic breast cancer; N9831: North Central Cancer Treatment Intergroup Trial N9831; NSABP: National Surgical Adjuvant Breast and Bowel Project; S1–5: Strategy 1–5; SA: Sensitivity analysis; PSA: Probabilistic sensitivity analysis; QALY: Quality adjusted life year; WTP: Willingness to pay.

Table 2. Summary of economic analyses of trastuzumab in early-stage breast cancer.

Study (Country) (Perspective) (Time horizon) (Discount rate)	Study aim	Methods	Population	Alternatives examined	Results (2007 US$)	Sensitivity analysis	Author's conclusions	Ref.
Shiroiwa *et al.* (2008) (Japan) (Payer) (50 years) (3%)	To estimate the cost–effectiveness of adjuvant trastuzumab compared to standard adjuvant CTX based on the results of the HERA‡ trial in a Japanese population	– CEA – Markov model – 1-month cycle – Budget impact analysis – Rates of severe, a/symptomatic heart failure modeled – Cardiotoxicity treatment costs modeled	Hypothetical cohort of patients, median 49 years old	Standard adjuvant CTX versus standard adjuvant CTX with trastuzumab	$23,309/LYS over standard assuming trastuzumab benefit lasts 5 years	– Univariate analysis of all model assumptions – PSA of all transition probabilities and costs – Model was most sensitive to the duration of trastuzumab efficacy	1-year trastuzumab therapy is superior to standard adjuvant CTX in terms of cost–effectiveness	[44]
Dedes *et al.* (2007) (Finland) (Provider) (5, 10 and 15 years) (Costs: 3%, outcomes: 0%)	CEA taking into account the long-term clinical benefits and adverse effect profile of trastuzumab based on interim results of HERA‡ and FinHer* trials	– CEA – Markov model – 1-year cycle – Costs of cardiac monitoring and treatment modeled	– Hypothetical cohort of 10,000 post-surgery women aged 50 years – ER+ patients included (50%)	Standard adjuvant CTX versus standard adjuvant CTX with trastuzumab	– HERA: $25,807/LYS over 15 years – FinHer: $6,783/LYS over 15 years	– Univariate analysis of: trastuzumab price, yearly cost of MBC, clinical efficacy of trastuzumab, cost of treating local–regional recurrence and discount rate – Results sensitive to the price of trastuzumab, trastuzumab efficacy and regimen	– Adjuvant trastuzumab may be CE in a long-term perspective based on HERA data – FinHer 9-week regimen may save costs compared to no adjuvant trastuzumab	[30]

*See ref [1].
‡See ref [43].
BC: Breast cancer; BCIRG: Breast Cancer International Research Group; CE: Cost–effectiveness; CEA: Cost–effectiveness analysis; CEAC: Cost–effectiveness acceptability curve; CUA: Cost–utility analysis; CTX: Chemotherapy; ER: Estrogen receptor; FinHer: Finland Herceptin; FEC: 5-fluorouracil, epirubicin and cyclophosphamide; HERA: Herceptin adjuvant; HTA: Health technology assessment; ICER: Incremental cost–effectiveness ratio; IHC: Immunohistochemistry; LVEF: Left ventricular ejection fraction; LYS: Life years saved; LYG: Life years gained; MBC: Metastatic breast cancer; N9831: North Central Cancer Treatment Intergroup Trial N9831; NSABP: National Surgical Adjuvant Breast and Bowel Project; S1–5: Strategy 1–5; SA: Sensitivity analysis; PSA: Probabilistic sensitivity analysis; QALY: Quality adjusted life year; WTP: Willingness to pay.

Table 2. Summary of economic analyses of trastuzumab in early-stage breast cancer.

Study (Country) (Perspective) (Time horizon) (Discount rate)	Study aim	Methods	Population	Alternatives examined	Results (2007 US$)	Sensitivity analysis	Author's conclusions	Ref.
Garrison *et al.* (2007) (USA) (Payer, partial societal) (Lifetime, 20 years) (3%)	To develop a model to estimate the incremental CE of adjuvant trastuzumab for early-stage HER2$^+$ BC in the USA	– CUA – Markov model – Costs of cardiac monitoring and treatment of cardiac events modeled – Assumed that five tests were performed for every HER2$^+$ patient and that 30% of tests were FISH	– Hypothetical cohort of 50-year old women – HER2$^+$ status defined as IHC3$^+$ or FISH$^+$	Adjuvant trastuzumab + standard of care versus standard of care alone	– Payer perspective: $27,584/QALY – Partial societal perspective: $28,858/QALY	– Univariate and multivariate for all parameters except control arm survival – Most sensitive to discount rate, drug price, probability of metastasis and time horizon, but findings are generally robust – Best case: $9,506; worst case: $72,404	– Cost-effective over projected life-time horizon – Lower CE ratio than many accepted oncology treatments	[32]
Kurian *et al.* (2007) (US) (Partial societal) (40 years) (3%)	To analyze costs and health benefits of anthracycline-based trastuzumab regimen with anthracycline and nonanthracycline regimens	– Markov model – 1-month cycle Monte Carlo simulation – Cardiotoxicity states modeled – Costs of cardiac monitoring and treatment modeled – Assumed death due to cardiotoxicity is negligible	Hypothetical cohort of post-surgery women aged 49 years, similar to those enrolled in NSABP B-31, N9831 and BCIRG 006 trials	– S1: no trastuzumab – S2: trastuzumab with nonanthracycline chemotherapy – S3: trastuzumab with anthracycline chemotherapy	– S2 regimen dominated – S3 regimen ICER of $43,330/QALY versus S1	– Univariate and PSA – Varied ± 20%; results most sensitive to: discount rate, cost of S3, cost of treatment and median survival after recurrence	Trastuzumab in combination with anthracycline CTX is cost-effective and trastuzumab with nonanthracycline CTX is dominated in base case and PSA	[33]

[*] *See ref* [1].
[‡] *See ref* [43].
BC: Breast cancer; BCIRG: Breast Cancer International Research Group; CE: Cost–effectiveness; CEA: Cost–effectiveness analysis; CEAC: Cost–effectiveness acceptability curve; CUA: Cost–utility analysis; CTX: Chemotherapy; ER: Estrogen receptor; FinHer: Finland Herceptin; FEC: 5-fluorouracil, epirubicin and cyclophosphamide; HERA: Herceptin adjuvant; HTA: Health technology assessment; ICER: Incremental cost–effectiveness ratio; IHC: Immunohistochemistry; LVEF: Left ventricular ejection fraction; LYS: Life years saved; LYG: Life years gained; MBC: Metastatic breast cancer; N9831: North Central Cancer Treatment Intergroup Trial N9831; NSABP: National Surgical Adjuvant Breast and Bowel Project; S1–5: Stragety 1–5; SA: Sensitivity analysis; PSA: Probabilistic sensitivity analysis; QALY: Quality adjusted life year; WTP: Willingness to pay.

Table 2. Summary of economic analyses of trastuzumab in early-stage breast cancer.

Study (Country) (Perspective) (Time horizon) (Discount rate)	Study aim	Methods	Population	Alternatives examined	Results (2007 US$)	Sensitivity analysis	Author's conclusions	Ref.
Liberato *et al.* (2007) (Italy) (Health care system) (15 years) (3%)	To evaluate the CE of 12-month adjuvant trastuzumab in women with HER2+ early BC	– CUA – Decision tree + Markov model – Monte Carlo simulation – Joint analysis of NSABP B-31 and N9831 trials – Cardiotoxicity states modeled – Assumed death due to cardiotoxicity is negligible – Costs of cardiac monitoring and treatment modeled	Hypothetical cohort aged 50 years and clinically similar to those enrolled in NSABP B-31 and N9831 trials	CTX alone versus weekly trastuzumab for 1 year	– $20,095/QALY or ≤$27,044 (€20,000) /QALY with 91% likelihood in Italian setting – $20,605/QALY ≤$25,000/QALY with 91% likelihood in USA setting	– PSA for all parameters – Multivariate SA with results of HERA‡ – Less CE in elderly, low-risk patients – CEAC provided	– In the subgroup of patients analyzed, trastuzumab amplified clinical advantage of adjuvant therapy at a cost generally considered appropriate for the added value produced – Highly CE in medium to high-risk patients of up to 70 years	[34]
Millar *et al.* (2007) (Australia) (Payer) (51 years) (3%)	To estimate the CE of trastuzumab for early BC from an Australian health system perspective	– CUA – Markov model – 1-year cycle length – Calibrated disease model – Joint analysis of NSABP B-31 and HERA‡ trials to estimate 52-week trastuzumab efficacy – Costs of cardiac monitoring and treatment modeled – Assumed cardiotoxicity cured 3 years after CTX	– Hypothetical cohort aged 50 years at diagnosis – 35- and 65-year-old cohorts examined	9- (FinHer*) and 52- (HERA‡) week trastuzumab treatment versus no trastuzumab	– $18,882/ QALY for 52-week regimen – $14,083/QALY for 9-week regimen (FinHer)	– Univariate and multivariate analysis of: efficacy, recurrence, trastuzumab price, cost of MBC, noncancer illness treatment and duration of trastuzumab benefit – Model was most sensitive to drug cost, duration of treatment benefit and discount rate	– Adjuvant trastuzumab may be CE – Trastuzumab group incurs greater costs due to increased survival later in life – Substantially improved CE with 9-week trastuzumab regimen	[23]

*See ref [1].
‡See ref [43].
BC: Breast cancer; BCIRG: Breast Cancer International Research Group; CE: Cost–effectiveness; CEA: Cost–effectiveness analysis; CEAC: Cost–effectiveness acceptability curve; CUA: Cost–utility analysis; CTX: Chemotherapy; ER: Estrogen receptor; FinHer: Finland Herceptin; FEC: 5-fluorouracil, epirubicin and cyclophosphamide; HERA: Herceptin adjuvant; HTA: Health technology assessment; ICER: Incremental cost–effectiveness ratio; IHC: Immunohistochemistry; LVEF: Left ventricular ejection fraction; LYS: Life years saved; LYG: Life years gained; MBC: Metastatic breast cancer; N9831: North Central Cancer Treatment Intergroup Trial N9831; NSABP: National Surgical Adjuvant Breast and Bowel Project; S1–5: Stragety 1–5; SA: Sensitivity analysis; PSA: Probabilistic sensitivity analysis; QALY: Quality adjusted life year; WTP: Willingness to pay.

Table 2. Summary of economic analyses of trastuzumab in early-stage breast cancer.

Study (Country) (Perspective) (Time horizon) (Discount rate)	Study aim	Methods	Population	Alternatives examined	Results (2007 US$)	Sensitivity analysis	Author's conclusions	Ref.
Norum *et al.* (2007) (Norway) (Societal) (Lifetime) (3%)	To explore CE of including trastuzumab in adjuvant BC treatment in Norway	– Decision–analytic model – Assumed 10% and 20% improvements in survival – Costs of cardiac monitoring and treatment modeled	Hypothetical cohorts of women aged 50–60 years and 60–70 years	Trastuzumab + current treatment (FEC_{100}) versus current treatment alone per November 2005 Norwegian Breast Cancer Group guideline	– 10% survival gain: $49,668–$63,480/QALY – 20%survival gain: $13,361–$27,171/QALY	– Univariate analysis of all parameters – Changes in LYG, trastuzumab price, production gained and disease recurrence all affected ICER	– Under Norwegian €50,000/LYG threshold for CE, adjuvant trastuzumab appears CE – A minimum of 8% improvement in survival over 10 years is needed for CE	[45]
NICE (2006) (UK) (NHS) (Lifetime) (3.5%)	To provide guidance on the use of trastuzumab in adjuvant therapy	– Manufacturer CUA within HTA – Markov cohort model based on results of HERA[‡] trial – Assumed all patients tested with IHC and FISH for confirmation of IHC2[+] – Cardiotoxicity states modeled – Cost of cardiac monitoring and management of minor and severe cardiac events modeled	– Women who had surgery and completed standard CTX – LVEF >55%	Standard adjuvant CTX (HERA) versus weekly trastuzumab for 1 year	NICE: with cardiotoxicity events in a clinical setting, new ICER ranged from $31,637–65.250/QALY – $11,254/QALY proposed by manufacturer is an underestimate	– Univariate, multivariate and PSA – Discount rate, trastuzumab cost and RRR have greatest impact – CEAC provided	– Adjuvant trastuzumab is cost-effective – NICE estimate assumes 100% of relapsed MBC patients are retreated with trastuzumab	[35]

[+]*See ref* [1].
[‡]*See ref* [43].
BC: Breast cancer; BCIRG: Breast Cancer International Research Group; CE: Cost–effectiveness; CEA: Cost–effectiveness analysis; CEAC: Cost–effectiveness acceptability curve; CUA: Cost–utility analysis; CTX: Chemotherapy; ER: Estrogen receptor; FinHer: Finland Herceptin; FEC: 5-fluorouracil, epirubicin and cyclophosphamide; HERA: Herceptin adjuvant; HTA: Health technology assessment; ICER: Incremental cost–effectiveness ratio; IHC: Immunohistochemistry; LVEF: Left ventricular ejection fraction; LYS: Life years saved; LYG: Life years gained; MBC: Metastatic breast cancer; N9831: North Central Cancer Treatment Intergroup Trial N9831; NSABP: National Surgical Adjuvant Breast and Bowel Project; S1–5: Strategy 1–5; SA: Sensitivity analysis; PSA: Probabilistic sensitivity analysis; QALY: Quality adjusted life year; WTP: Willingness to pay.

The findings of Dendukuri *et al.* echoed Elkin *et al.* [25] and Lidgren *et al.* [31], demonstrating that it was more cost-effective to either use FISH to confirm equivocal and positive IHC tests (see TABLE 4), or to test all patients with FISH up front. This lends support to the use of confirmatory testing for equivocal IHC test results, but does not fully capture the impact of false-negative or -positive results in the consideration of subsequent patient outcomes and treatment.

TABLE 5 provides a summary of test properties modeled in analyses considering testing and treatment. We found that models of therapy did not consider test timing (i.e., to test upon initial diagnosis at early-stage, at the time of metastasis or both). Test timing was probably not a concern in Lidgren's analysis, given the evidence demonstrating concordance between HER2 overexpression in early and later disease stages [31]. There are emerging data that suggest the rate of discordance in HER2 expression between initial diagnosis and relapse may be 6–8% [38,39]. However, the question of test timing may have been relevant in Elkin's analysis of metastatic therapy [25]. Interestingly, repeat IHC testing was not considered in either analysis. Indeed, IHC tests tend to be used as a first-line test owing to the ease of conduct despite well-documented inaccuracies. All studies considered FISH as the only test for confirmatory purposes in all stages of disease. The literature clearly demonstrates that the testing strategy is an influential factor when considering the cost–effectiveness of trastuzumab therapy.

Discussion

We have demonstrated that trastuzumab is widely considered cost-effective across a range of international settings, economic perspectives and clinical settings. Therapy was more cost-effective in early-stage disease due to the substantial gain in life expectancy observed with adjuvant therapy, allowing the additional cost of trastuzumab therapy to be distributed over a larger period. However, studies taking a societal or partial societal perspective, noted that the incremental life years obtained with trastuzumab resulted in additional costs accrued over the lifetime of patients. Studies that concluded that trastuzumab was not cost-effective against comparators were conducted in the metastatic setting. A few studies concluded that trastuzumab might be cost-effective given the difference in costs accrued over 52- and 9-week regimens and the uncertainty around optimal duration of adjuvant trastuzumab therapy. This discrepancy may well be clarified when long-term results of the FinHer and HERA trials become available. Our findings are consistent with a recent systematic review of 52-week trastuzumab in the early-stage setting [40]. McKeage *et al.* reviewed adjuvant trastuzumab and do note the importance of tesitng, but do not compare findings from studies that consider testing alone. Younis *et al.* also conducted a literature review of trastuzumab economic analyses and emphasized the clinical issues, particularly around concurrent or consecutive administration and duration of therapy [41]. This review also acknowledged the significance of HER2 testing within the context of trastuzumab treatment.

Testing emerged as an important influence on the cost–effectiveness of therapy. When test–treat strategies were considered in the metastatic or early-stage settings, the message was consistent – the choice of testing strategy can 'make or break' the cost–effectiveness of trastuzumab therapy in the adjuvant or metastatic setting. This was linked to the minimization of false-positive and false-negative diagnoses through the use of either alternative test–treat strategies, which favored initial IHC with FISH confirmation of equivocal and positive cases, or the alternative, FISH testing for all patients. The impact of testing was most apparent in Elkin's metastatic analysis, where trastuzumab was modeled in a population with a much shorter life expectancy. Lidgren's model was not as sensitive to test properties in univariate analysis, but two-way analysis of IHC sensitivity and specificity did result in the exclusion of the FISH testing alone strategy by dominance. This suggests that the accuracy of diagnostic tests and the resultant mistreatment of false-positive and false-negative patients can have a profound impact on the cost–effectiveness of subsequent therapy. However, it is important to note that all testing studies assumed FISH to be the gold standard despite the fact that its not 100% specific or 100% sensitive. A recent study of central laboratory testing found that FISH and IHC showed 100% and 84.2% sensitivity in predicting patient response to trastuzumab monotherapy [42]. Dendukuri's systematic review estimated the probability of a positive FISH result within each IHC result category (TABLE 4), and used those probabilities to inform the cost–effectiveness analysis. Similarly, Elkin estimated a weighted average probability of each IHC result category conditional upon FISH results from several studies; this weighted average informed Lidgren's analysis as well.

Table 3. Summary of economic analyses considering HER2 testing only.

Study (Country) (Perspective) (Time horizon) (Discount rate)	Study aim	Methods	Population	Alternatives examined	Results (2007 US$)	Sensitivity analysis	Conclusions	Ref.
Dendukuri *et al.* (2007) (Canada) (3rd party payer) (Cross-sectional) (N/A)	Meta-analysis of studies published between 2000 and 2005, to estimate the percentage of $HER2^+$ patients whose status was accurately determined by IHC and CEA of seven alternative strategies to test HER2 status	– Cross-sectional Bayesian meta-analysis to estimate the distribution of IHC scores and the probability of a positive FISH result in each IHC score category from a systematic review of IHC and FISH studies – CEA of screening strategies	Hypothetical cohort of BC patients tested with assays licensed by Health Canada	– S1: IHC + FISH confirmation of $IHC2^+$; $HER2^+$ defined as $IHC3^+$ or $FISH^+$ (base case) – S2: IHC only; $HER2^+$ defined as $IHC2^+$ or 3^+ – S3: IHC only; $HER2^+$ defined as $IHC3^+$ – S4: IHC + FISH confirmation of $IHC1^+$ or 2^+; $HER2^+$ defined as $IHC3^+$ or $FISH^+$ – S5: IHC + FISH confirmation of $IHC2^+$ and 3^+; $HER2^+$ defined as $FISH^+$ – S6: IHC + FISH confirmation of $IHC1^+$, 2^+ and 3^+; $HER2^+$ defined as $FISH^+$ – S7: FISH only	Strategies 2, 3 and 4 were ruled out by classic or extended dominance. The incremental costs per accurately determined HER2 status (ICERs) for the remaining nondominated strategies were: S5 over S1: $5,686; S6 over S5: $7,423; S7 over S6: $7,736.	– PSA performed by varying all parameters of the economic evaluation simultaneously over plausible ranges – The cost of FISH did effect CEA results	The strategy with the lowest ICER compared to current practice is to screen all BC patients with IHC plus confirmation of $IHC2^+$ and 3^+ by FISH ($5)	[4]

BC: Breast cancer; CE: Cost–effectiveness; CEA: Cost–effectiveness analysis; CMA: Cost-minimization analysis; CTX: Chemotherapy; HER2: Human EGF receptor-2; ICER: Incremental cost–effectiveness ratio; IHC: Immunohistochemistry; MBC: Metastatic breast cancer; PSA: Probabilistic sensitivity analysis; S1–7: Strategy 1–7.

Table 3. Summary of economic analyses considering HER2 testing only.

Study (Country) (Perspective) (Time horizon) (Discount rate)	Study aim	Methods	Population	Alternatives examined	Results (2007 US$)	Sensitivity analysis	Conclusions	Ref.
Morelle *et al.* (2006) (France) (Not stated) (20 years after first diagnosis) (0%)	To examine the CE of five different testing strategies to direct trastuzumab therapy at metastatic relapse in secondary care	– CEA – A decision tree modeled the five diagnostic strategies in a cohort of 10,000 patients. Test characteristics were adapted to French context using local laboratory testing guidelines, local fixatives and the accessibility of paraffin-embedded samples at time of relapse.	Hypothetical cohort of 10,000 infiltrating BC patients were used in the model. A prospective, multicentric study followed 2045 invasive BC patients from initial diagnosis to ascertain local diagnostic test characteristics.	– S1: IHC at first MBC relapse (base case) – S2: systematic IHC at diagnosis – S3: FISH at first MBC relapse when IHC is not possible due to fixation – S4: FISH at first MBC relapse + IHC to clarify status of IHC2+ and 3+ – IHC was only used in S4 and S5 when the tumor was fixed in formalin	– All strategies were more expensive but at least as effective, in comparison with the base case – ICER per criterion 1: S4 $839/false-negative avoided – ICER per criterion 2: S5 $5,491/false-positive avoided	– Methods unclear; authors varied the discount rate (3, 5%), FISH test cost and proportion of samples fixed in formalin – Results not sensitive to any of above	– Given that false-negatives were the only cause of patients with HER2 amplified tumor not being given trastuzumab at the first metastatic relapse, 'efficient strategies' were S1 and S4 – Determining HER2 status at diagnosis, as an indication for trastuzumab at metastatic relapse, incurs substantial incremental costs which do not appear to be justified	[36]
Dranitsaris *et al.* (2003) (Canada) (payer) (5 years) (3%)	To assess the economic benefit of undertaking HER2 testing at diagnosis or disease recurrence	– CMA – Decision tree – Trastuzumab treatment not modeled – IHC3+ for HER2+ – IHC2+ retested with FISH; FISH+ for HER2+ – Likelihood of disease recurrence at 5 years modeled	– Pre- and post-menopausal women diagnosed with stages I–III BC – Stage IV not included as HER2 diagnosis is standard of care at that stage	– S1: HER2 test at diagnosis (stages I–III) – S2: HER2 test at MBC recurrence (base case)	ICERS of testing at stages I–III versus MBC recurrence: – I: $30,369 – II: $66,296 savings – III: $53,616 savings Net savings of $89,543 if tested at first diagnosis versus recurrence	One-way analysis of: incidence of inconclusive results, effect of various adjuvant CTX protocols on survival and 5 year recurrence rates	HER2 testing of all newly diagnosed stage I–III BC patients would result in cost-savings for the Canadian healthcare system	[37]

BC: Breast cancer; CE: Cost–effectiveness; CEA: Cost–effectiveness analysis; CMA: Cost-minimization analysis; CTX: Chemotherapy; HER2: Human EGF receptor-2; ICER: Incremental cost–effectiveness ratio; IHC: Immunohistochemistry; MBC: Metastatic breast cancer; PSA: Probabilistic sensitivity analysis; S1–7: Strategy 1–7.

Table 4. Description of HER2 status determination by immunohistochemistry testing.

IHC grade	Description
0	Negative test; no staining
1+	Negative test; weak, incomplete membrane staining in any proportion of tumor cells
2+	Equivocal test; complete membrane staining that is either nonuniform or weak in intensity but with obvious circumferential distribution in at least 10% of cells
3+	Positive test; uniform intense membrane staining of >30% of invasive tumor cells

IHC: Immunohistochemistry; HER2: Human EGF receptor-2.

The Canadian meta-analysis and cost–effectiveness analysis of HER2 testing strategies [4] echoed the results demonstrated in the treatment setting: initial testing with IHC followed by confirmatory FISH for IHC2+ and 3+ results was the most cost-effective strategy. Garrison's analysis provided indirect support for the importance of modeling HER2 test characteristics. His analysis attempted to capture the impact of testing by assuming that five tests must be performed per accurate diagnosis. This approach only captures the cost of additional testing, and does not estimate the impact of improper diagnosis. Moreover, the assumption that one in five tests leads to an accurate diagnosis actually represents population-wide testing, in which optimistic estimates of HER2 prevalence are about 20%. This assumes that all 20% (1/5) of prevalent cases will be identified with a single test, while the remaining 80% of HER2-negative patients (4/5) will also be identified with a single test. Local practice patterns significantly influenced the selection of alternatives and analytic approach in testing-focused analyses, probably accounting for the greater variability in the findings of these studies. Unfortunately, local testing practices were not modeled in conjunction with treatment. Therefore, future analyses would benefit from a more 'real-world' reflection of testing practice in a model considering the joint effects of testing and treatment. We also found that IHC was only considered as the first test in all studies evaluating test sequencing; it is likely that this reflects the well-documented inaccuracy of this test.

Cardiac toxicity was considered in all analyses of early-stage disease, and compared to only one analysis of metastatic breast cancer. This probably reflects the better characterization of cardiac toxicity rates and the nature of this side effect after years of careful study. While several analyses captured the economic consequences of

Table 5. Summary of economic analyses of trastuzumab considering the choice or sequence of HER2 test used to identify potential responders.

Author, date (Country) (Perspective) (Time horizon) (Discount rate)	How were test properties modeled?	Was test timing considered?	Were alternative test sequences considered?	What impact did test properties have on the results?	Ref.
Lidgren *et al.* (2008) (Sweden) (Societal) (lifetime) (3%)	Test properties were modeled according to Elkin *et al.**	No; only testing at diagnosis of early-stage BC was considered	Yes; authors considered primary testing with IHC followed by FISH confirmation under various circumstances, or FISH testing alone (see Table 2 for test–treat strategies)	The CE of trastuzumab therapy was highly dependent on testing strategy; trastuzumab therapy was only CE in early BC if FISH testing was used to confirm all IHC-positive patients ($36,112/QALY), or if FISH was the only test used ($41,629/QALY)	[31]
Elkin *et al.* (2004) (USA) (Not stated) (Lifetime) (3%)	A total of ten studies of IHC and FISH test properties were reviewed and weighted averages were calculated to inform model transition probabilities	No; only testing at MBC was considered	Yes; authors considered primary testing with IHC followed by FISH confirmation under various circumstances, or FISH testing alone (see Table 1 for test–treat strategies)	– Testing had an important impact on the cost–effectiveness of trastuzumab therapy; Strategy 4 (IHC with FISH confirmation of IHC2+ and 3+; trastuzumab for FISH+) has an ICER of $153,648/QALY, while Strategy 6 (FISH only; trastuzumab for FISH+) had an ICER of $178,231/QALY – The authors concluded that society must be willing to pay for CE testing strategies if it is willing to pay for trastuzumab at all.	[25]

**See ref.* [25].
BC: Breast cancer; CE: Cost–effectiveness; HER2: Human EGF receptor-2; ICER: Incremental cost–effectiveness ratio; IHC: Immunohistochemistry; MBC: Metastatic breast cancer; QALY: Quality adjusted life year.

cardiac toxicity by modeling the costs associated with monitoring and treatment, only Elkin's and Lidgren's test–treat designs were able to capture the consequences of cardiotoxicity in light of the potential for inappropriate treatment. Lidgren's design further attempted to estimate changes in quality of life due to development of congestive heart failure. Although none of the models included in this review were sensitive to cardiac toxicity rates or associated costs, it is important to note that the ability to capture all effects of improper treatment can only be achieved through consideration of test–treat strategies. This may become a very important consideration for future targeted therapies.

Any decision regarding the cost–effectiveness of trastuzumab should consider the context of its application in conjunction with the evidence. This systematic summary of the evidence suggests that trastuzumab is a cost-effective treatment option in early-stage breast cancer and may be cost-effective in metastatic disease as well. This conclusion was consistent in early-stage disease across a variety of international viewpoints and economic perspectives. Moreover, all ICER estimates detailed herein, fall within the range of other accepted cancer treatments. Most models were sensitive to the efficacy and cost of trastuzumab therapy. The ability of clinicians to accurately identify appropriate therapeutic candidates using HER2 testing emerged as an important factor in determining the effectiveness and cost–effectiveness of trastuzumab therapy. This review has also demonstrated that the choice of test and test sequencing can influence how cost-effective trastuzumab treatment is by

Executive summary

The clinical scenario

- Trastuzumab is a targeted therapy for human EGF receptor-2 (HER2)-positive breast cancer.
- HER2-positive patients may be identified using immunohistochemistry or FISH testing; these tests differ in their associated costs, sensitivity and specificity.
- The cost–effectiveness of trastuzumab may therefore be linked to the accuracy of the strategy used to identify patients in addition to the clinical efficacy of the drug.
- We therefore aimed to identify and review the current evidence of cost–effectiveness of trastuzumab and HER2 testing.

Analytic approach

- Systematic review of economic evaluations or health technology assessments (containing economic evaluations) of trastuzumab therapy or HER2 testing.
- Electronic searches conducted in BIOSIS, Cochrane, CRD, EconLit, Excerpta Medica Database (EMBASE), Health Economic Evaluations Database (HEED), MEDLINE and PubMed.
- Data extraction included study aim, analytic methods, patient population, alternatives examined, primary findings, uncertainty analysis and conclusions.

Results

- A total of 17 economic evaluations were included in the review; 12 considered trastuzumab therapy alone, two considered HER2 testing and subsequent trastuzumab treatment and three considered HER2 testing alone.
- All studies examining trastuzumab in the early-stage setting (n = 10) considered trastuzumab to be a cost-effective treatment option.
- One study of trastuzumab in the metastatic setting concluded that trastuzumab was not cost-effective.
- HER2 testing studies tended to focus on test sequencing or timing relative to jurisdictional constraints, and therefore demonstrated greater discordance in their conclusions.
- Studies that examined testing and treatment (n = 2) reported that initial testing with immunohistochemistry followed by confirmatory FISH testing with trastuzumab provided to FISH positive patients, was the most cost-effective test–treat strategy.
- The same conclusion regarding test sequencing was reached in another study that considered testing exclusively.
- Use of FISH testing alone was also cost-effective.
- Studies that considered testing and treatment noted that test characteristics were influential model parameters.
- Costs associated with trastuzumab-induced cardiotoxicity were modeled in all analyses of early-stage disease compared to a single metastatic study.
- The methods used to model cardiotoxicity varied and rarely included health states (3/10) in early-stage disease.

Conclusion

- Trastuzumab was considered cost-effective in the early-stage setting because the significant up-front costs of therapy were offset by the significant gain in life for this patient population.
- Testing is an important factor in determining the cost–effectiveness of trastuzumab therapy and the choice of test strategy is key.
- Initial testing with IHC and confirmatory FISH testing has the lowest incremental cost–effectiveness ratio, as it reduces the excess cost associated with treating false-positive patients while minimizing the lost opportunity of not treating false-negative patients. FISH testing alone is also cost-effective.
- Studies that model test–treat strategies were also able to capture the consequences of improper diagnosis in terms of cardiotoxicity.
- Economic evaluations must consider testing in conjunction with trastuzumab treatment for breast cancer.

reducing the number of false-positive and false-negative HER2 diagnoses. In this regard, initial testing with IHC followed by FISH confirmation of IHC2$^+$ and 3$^+$ patients was associated with the lowest ICERs, while FISH testing was also cost-effective. Again, the choice of test–treat strategy in practice must take the local context into account. Nevertheless, these findings are generally consistent with the guidelines for HER2 testing issued by the College of American Pathologists (IL, USA)/American Society for Clinical Oncology (VA, USA) [5]. Future analyses would benefit from further exploration of testing factors, particularly the accuracy of testing in the community versus testing in reference laboratories, and whether tratuzumab is provided within the context of testing guidelines. The impact of these factors has not been examined using decision–analytic modeling. Additional exploration of the recently approved chromogenic *in situ* hybridization (CISH) test [107] for HER2 status determination is also warranted.

Future perspective

As the application of personalized medicine through targeted therapies continues to expand in medicine, it will be imperative to consider test accuracy and sequencing (where more than one test is available) when analyzing the effectiveness and cost–effectiveness of future targeted therapies. Only through careful consideration of test–treat strategies will analysts be able to capture the impact of improper diagnosis as it relates to missed therapeutic opportunities and needless exposure to treatment-related side effects.

Financial & competing interests disclosure

This research was funded in part by an award from the Ontario Council on Graduate Studies; it was also funded by two grants from the National Cancer Institute (R01CA101849 and P01CA130818). Deborah Marshall provides ad hoc consulting for i3 Innovus, a global health economics and outcomes research company. Natasha Leighl has previously received honoraria from Hoffmann-La Roche within the last 2 years for continuing medical education lectures that were not related to this project or herceptin. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Bibliography

Papers of special note have been highlighted as:
- ▪ of interest
- ▪▪ of considerable interest

1 Joensuu H, Kellokumpu-Lehtinen PL, Bono P *et al.*: Adjuvant docetaxel or vinorelbine with or without trastuzumab for breast cancer. *N. Engl. J. Med.* 354, 809–820 (2006).

2 McKeage K, Perry CM: Trastuzumab: a review of its use in the treatment of metastatic breast cancer overexpressing HER2. *Drugs* 62, 209–243 (2002).

3 De Laurentiis M, Cancello G, Zinno L *et al.*: Targeting HER2 as a therapeutic strategy for breast cancer: a paradigmatic shift of drug development in oncology. *Ann. Oncol.* 16(Suppl. 4), IV7–IV13 (2005).

4 Dendukuri N, Khetani K, McIsaac M, Brophy J: Testing for HER2-positive breast cancer: a systematic review and cost–effectiveness analysis. *Can. Med. Assoc. J.* 176, 1429–1434 (2007).

▪ **Presents one methodology for the synthesis of human EGF receptor-2 (HER2) test accuracy data and the integration of this information into cost–effectiveness analysis.**

5 Wolff AC, Hammond ME, Schwartz JN *et al.*: American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J. Clin. Oncol.* 25, 118–145 (2007).

6 Dahabreh IJ, Linardou H, Siannis F, Fountzilas G, Murray S: Trastuzumab in the adjuvant treatment of early-stage breast cancer: a systematic review and meta-analysis of randomized controlled trials. *Oncologist* 13(6), 620–630 (2008).

7 Minhas R: How much will herceptin really cost? The money is already there. *BMJ* 333(7580), 1219–1220 (2006).

8 Dillon A, Littlejohns P: How much will herceptin really cost? Balancing local and national priorities is not new. *BMJ* 333(7580), 1219–1220 (2006).

9 Kondro W, Sibbald B: Patient demand and politics push Herceptin forward. *Can. Med. Assoc. J.* 173(4), 347–348 (2005).

10 Metcalfe S, Evans J: PHARMAC responds on Herceptin assumptions and decisions. *N. Z. Med. J.* 120(1260), U2692 (2007).

11 Metcalfe S, Evans J, Priest G: PHARMAC funding of 9-week concurrent trastuzumab (Herceptin) for HER2-positive early breast cancer. *N. Z. Med. J.* 120(1256), U2593 (2007).

12 Rosevear M: PHARMAC and Herceptin for early-stage breast cancer in New Zealand: Herceptin or deception? *N. Z. Med. J.* 119(1235), U2014 (2006).

13 Petersen S: Roche responds to the 'Herceptin or deception' article. *N. Z. Med. J.* 119(1237), U2069 (2006).

14 Breast Cancer Advocacy Coalition; Burgess EP: Breast Cancer Advocacy Coalition responds to the 'Herceptin or deception' article. *N. Z. Med. J.* 119(1237), U2065 (2006).

15 Hillner BE, Smith TJ: Do the large benefits justify the large costs of adjuvant breast cancer trastuzumab? *J. Clin. Oncol.* 25, 611–613 (2007).

16 Drummond MF, Sculpher MJ, Torrance GW, O'Brian BJ, Stoddart GL: *Methods for the Economic Evaluation of Health Care Programmes.* Oxford University Press, Oxford, UK (2005).

17 Perik PJ, de Korte MA, van Veldhuisen DJ, Gietema JA, Sleijfer DT, de Vries EGE: Cardiotoxicity associated with the use of trastuzumab in breast cancer patients. *Expert Rev. Anticancer Ther.* 7(12), 1763–1771 (2007).

18 Fleiss JL: The statistical basis of meta-analysis. *Stat. Methods Med. Res.* 2, 121–145 (1993).

19 Vernaglia PR, Cunha FM, Correa M *et al.*: Cost–effectiveness analysis of trastuzumab therapy in patients with early HER-2 positive breast cancer in Brazil. *Breast Cancer Res. Treat.* 106(Suppl. 1) S106 (2007) (Abstract).

20 Hillner BE: [5040] Clinical and cost–effectiveness implications of the adjuvant transtuzumab in HER2+ breast cancer trials. *San Antonio Breast Cancer Symposium 2005.* San Antonio, TX, USA, 8–11 December (2005).

21 Hornberger J, Kerrigan M, Foutel V: Cost–effectiveness of trastuzmab (Herceptin®) for treatment of metastatic breast cancer. *Ann. Oncol.*13, 52 (2002)

22 Ragaz J, Spinelli JJ: [2029] Cost–benefit estimates of adjuvant (Adj) trastuzumab (herceptin, H) for early breast cancer (BrCa). *San Antonio Breast Cancer Symposium 2005.* San Antonio, TX, USA, 8–11 December (2005).

23 Millar JA, Millward MJ: Cost–effectiveness of trastuzumab in the adjuvant treatment of early breast cancer: a lifetime model. *Pharmacoeconomics* 25, 429–442 (2007).

24 Neyt M, Huybrechts M, Hulstaert F, Vrijens F, Ramaekers D: Trastuzumab in early stage breast cancer: a cost–effectiveness analysis for Belgium. *Health Policy* 87(2), 146–159 (2008).

25 Elkin EB, Weinstein MC, Winer EP, Kuntz KM, Schnitt SJ, Weeks JC: HER-2 testing and trastuzumab therapy for metastatic breast cancer: a cost–effectiveness analysis. *J. Clin. Oncol.* 22, 854–863 (2004).

▪▪ **Demonstrates how test accuracy and sequencing can be modeled in conjunction with treatment; it focuses on the metastatic setting.**

26 Norum J, Risberg T, Olsen JA: A monoclonal antibody against HER-2 (trastuzumab) for metastatic breast cancer: a model-based cost–effectiveness analysis. *Ann. Oncol.* 16, 909–914 (2005).

27 Poncet B, Bachelot T, Colin C *et al.*: Use of the monoclonal antibody anti-HER2 trastuzumab in the treatment of metastatic breast cancer: a cost–effectiveness analysis. *Am. J. Clin. Oncol.* 31(4), 363–368 (2008).

28 National Institute for Clinical Excellence: Guidance on the use of trastuzumab for the treatment of advanced breast cancer. 34, 1–14 (2002).

29 Slamon DJ, Leyland-Jones BSS, Fuchs H *et al.*: Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that over expresses HER2. *N. Engl. J. Med.* 344, 783–792 (2001).

30 Dedes KJ, Szucs TD, Imesch P, Fedier A, Fehr MK, Fink D: Cost–effectiveness of trastuzumab in the adjuvant treatment of early breast cancer: a model-based analysis of the HERA and FinHer trial. *Ann. Oncol.* 18, 1493–1499 (2007).

31 Lidgren M, Jonsson B, Rehnberg C, Willking N, Bergh J: Cost–effectiveness of HER2 testing and 1-year adjuvant trastuzumab therapy for early breast cancer. *Ann. Oncol.* 19, 487–495 (2008).

▪ **Illustrates the consequences of inappropriate diagnosis in the early-stage setting by modeling HER2 test accuracy and sequencing to direct trastuzumab treatment.**

32 Garrison LP, Lubeck D, Lalla D, Paton V, Dueck A, Perez EA: Cost–effectiveness analysis of trastuzumab in the adjuvant setting for treatment of HER2-positive breast cancer. *Cancer* 110(3), 489–498 (2007).

33 Kurian AW, Thompson RN, Gaw AF, Arai S, Ortiz R, Garber AM: A cost–effectiveness analysis of adjuvant trastuzumab regimens in early HER2/*neu*-positive breast cancer. *J. Clin. Oncol.* 25, 634–641 (2007).

34 Liberato NL, Marchetti M, Barosi G: Cost effectiveness of adjuvant trastuzumab in human epidermal growth factor receptor 2-positive breast cancer. *J. Clin. Oncol.* 25, 625–633 (2007).

35 National Institute for Health and Clinical Excellence: Trastuzumab for the adjuvant treatment of early-stage HER2-positive breast cancer. 107 (2006).

36 Morelle M, Haslé E, Treilleux I, Michot JP, Bachelot T: Cost–effectiveness analysis of strategies for HER2 testing of breast cancer patients in France. *Int. J. Technol. Assess. Health Care* 22, 396–401 (2006).

37 Dranitsaris G, Norris B, Hanna W, O'Malley F, Gelmon K: Identifying the optimal testing of HER2 testing in patients with breast cancer: a Canadian economic evaluation. *Curr. Oncol.* 10, 36–44 (2003).

38 Alvarez R, Albarracin C, Valero V, Sahin A, Johnson M, Hortobagyi G: Discordance rates of HER-2 expression between breast primary tumors and paired metastatic tumor specimens. *2008 American Society of Clinical Oncology Breast Cancer Symposium* (2008) (Abstract 121).

39 Simmons C, Miller N, Geddie W, Gianfelice D, Dranitsaris G, Oldfield M, Clemons M: Changes in breast tumor receptor status with time: a prospective study assessing the impact of obtaining confirmatory biopsy at metastatic recurrence on patient management. *2008 American Society of Clinical Oncology Breast Cancer Symposium* (2008) (Abstract 124).

40 McKeage K, Lyseng-Williamson KA: Trastuzumab: a pharmacoeconomic review of its use in early breast cancer. *Pharmacoeconomics* 6(8), 699–719 (2008).

▪▪ **Presents an alternative systematic review of cost–effectiveness analyses of trastuzumab in breast cancer.**

41 Younis T, Skedgel C: Is trastuzumab a cost-effective treatment for breast cancer? *Expert Rev. Pharmacoeconomics Outcomes Res.* 8(5), 433–442 (2008).

42 Hofmann M, Stoss O, Gaiser T *et al.*: Central HER2 IHC and FISH analysis in a trastuzumab (Herceptin) Phase II monotherapy study: assessment of test sensitivity and impact of chromosome 17 polysomy. *J. Clin. Pathol.* 61(1), 89–94 (2008).

43 Piccart-Gebhart MJ, Procter M, Leyland-Jones B *et al.*: Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N. Engl. J. Med.* 353, 1659–1672 (2005).

44 Shiroiwa T, Fukuda T, Shimozuma K, Ohashi Y, Kiichiro T: The model-based cost–effectiveness analysis of 1-year adjuvant trastuzumab treatment: based on 2-year follow-up HERA trial data. *Breast Cancer Res. Treat.* 109, 559–566 (2008).

45 Norum J, Olsen JA, Wist EA, Lonnin PE: Trastuzumab in adjuvant breast cancer therapy: a model based cost–effectiveness analysis. *Acta Oncol.* 46, 153–164 (2007).

Websites

101 Doherty D: Roche sales fall 4% on lower demand for Tamiflu (2008) www.bloomberg.com/apps/news?pid=20670001&refer=home&sid=aIW.jE9ReeJU

102 US Federal Reserve: Foreign exchange rates (annual). G5A(405) (1999) www.federalreserve.gov/releases/g5a/19990104/

103 US Federal Reserve: Foreign exchange rates (annual). G5A(405) (2007) www.federalreserve.gov/releases/g5a/20070103/

104 US Federal Reserve: Foreign exchange rates (annual). G5A(405) (2008) www.federalreserve.gov/releases/g5a/20030106/

105 US Federal Reserve: Foreign exchange rates (annual). G5A(405) (2008) www.federalreserve.gov/releases/g5a/current/

106 US Bureau of Labor Statistics: Consumer price index – all urban consumers, medical care. *Series ID: CUUR0000SAM* (2008) www.bls.gov/cpi/#data 52

107 US Food and Drug Administration: New device approval: invitrogen SPOT-Light® HER2 CISH™ Kit – P050040 (2009) www.fda.gov/cdrh/mda/docs/p050040.html

Appendix

Appendix 1. Search strategies.

No.	Database: BIOSIS previews
	Search history
1	TS=((Economic Impact OR Economic Value OR Pharmacoeconomics OR Health Care Cost OR Economic Factors OR Economics OR Cost Analysis OR Cost OR Economic Analysis OR Cost-Effectiveness OR Costs OR "Quality of Life" OR Health Care Costs OR Medical Costs OR Quality-of-life).mp.)
2	TS=((Trastuzumab or Herceptin or Epidermal Growth Factor Receptor 2 or her2 or Oncogene Neu or erbB-2 Receptor).mp.)
3	TS=(Trastuzumab or Herceptin or Epidermal Growth Factor Receptor 2 or her2 or Oncogene Neu or erbB-2 Receptor)
4	TS=((((Breast or Mammary) and (Neoplasm* or Cancer* or Carcinoma* or Tumour* or Tumor* or Metastas*)) or (Breast Neoplasm* or Breast Cancer* or Breast Carcinoma*)).mp.)
5	TS=(((Breast or Mammary) and (Neoplasm* or Cancer* or Carcinoma* or Tumour* or Tumor* or Metastas*)) or (Breast Neoplasm* or Breast Cancer* or Breast Carcinoma*))
6	#4 AND #2 AND #1
7	#5 AND #3 AND #1
DocType: All document types; LitType: All literature types; Language: English; Taxa Notes: All Taxa Notes.	
No.	**Database: Cochrane via Wiley (Cochrane reviews, technology assessments, NHS Economic evaluation database [NHS EED])**
	Search history
1	MeSH descriptor Breast Neoplasms explode all trees
2	((breast* or mammar*) near/3 (neoplasm* or metastas* or carcinoma* or tumour* or tumor* or cancer*)):ti,ab,kw
3	(#1 OR #2)
4	(trastuzumab* or herceptin*):ti,ab,kw
5	(cost effective* or cost benefit* or cost utility*):ti,ab,kw
6	MeSH descriptor Costs and Cost Analysis explode all trees
7	MeSH descriptor Economics, Pharmaceutical explode all trees
8	(#5 OR #6 OR #7)
9	(#3 AND #4 AND #8)
No.	**Database: Centre for reviews & dissemination (Database of Abstracts of Reviews of Effects [DARE], NHS EED, Health Technology Assessment Database [HTA])**
	Search history
1	MeSH Breast Neoplasms
2	MeSH Breast
3	mammary
4	#2 OR # 3
5	MeSH Neoplasms
6	metastas*
7	MeSH Carcinoma
8	carcinoma*
9	tumour*
10	cancer*
11	tumor*
12	neoplasm*
13	#5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12
14	breast OR mammary NEAR neoplasm OR tumor OR tumour OR cancer OR carcinoma
15	#4 and #13
16	#1 or #15
17	#1 or #14
18	trastuzumab
19	herceptin
20	MeSH Receptor, erbB-2 EXPLODE 2

Appendix 1. Search strategies.	
21	her2
22	or/21-24
23	#18 or # 22 or # 23 or #21
24	#16 and #23
25	#17 and #23
26	#17 or #16
27	#26 and #23
28	MeSH Costs and Cost Analysis EXPLODE 1
29	(value AND of AND life OR economics, AND medical OR economics, AND pharmaceutical OR models, AND economic OR markov AND chains OR monte AND carlo AND method OR uncertainty) AND .sh.
30	("value of life" OR economics, AND medical OR economics, AND pharmaceutical OR models, AND economic OR "markov chains" OR "monte carlo method" OR uncertainty) AND .sh.
31	"value of life" OR economics
32	economics
33	"medical economics"
34	"markov chain"
35	"markov chain*"
36	"monte carlo"
37	uncertainty
38	(quality NEAR life) OR (willingness NEAR pay) OR (quality NEAR "adjusted life year*") OR (sensitivity NEAR analysis)
39	QOL OR QOLY* OR HRQOL OR QALY*
40	decision NEAR tree* OR analy* OR model*
41	MeSH Cost-Benefit Analysis EXPLODE 1
42	MeSH Economics, Pharmaceutical EXPLODE 1
43	#28 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42
44	#24 and #43
45	#25 and #43
46	english:la
47	#44 and #46
48	#45 and #46
No.	**Database: EconLit (earliest to 2008)**
	Search history
1	((Economic Impact) or (Economic Value) or Pharmacoeconomic*) or ((Health Care Cost) or (Economic Factor*) or Economic*) or ((Cost Analysis) or Cost* or (Economic Analysis)) or (Cost-Effectiveness or (Quality of Life) or (Health Care Cost*)) or ((Medical C ost*) or Quality-of-life or Cost-Benefit) or (discount* or HRQOL or QOL) or (QALY* or (sensitivity analysis) or (Health Economic*)) or ((Economic Evaluation) or (willingness to pay))
2	(breast cancer) or (breast carcinoma) or (breast tumour)
3	(herceptin or trastuzumab or (epidermal growth factor receptor 2)) or (erbB-2 or Her2 or Her2/neu) or ((erbB-2 receptor$) or (neu oncogene))
4	#1 and #2 and #3
Terms searched 'anywhere'.	
No.	**Database: The Excerpta Medica Database (EMBASE)**
	Search history
1	exp Breast Tumor/
2	((breast$ or mammar$) adj3 (neoplasm$ or metastas$ or carcinoma$ or tumour$ or tumor$)).ti,ab.
3	Trastuzumab/ or 180288-69-1.rn.
4	herceptin$.ti,ab,tn.
5	(cost effective$ or cost benefit$ or cost utility$).ti,ab.
6	exp Health economics/ or exp Economic Evaluation/ or exp Pharmacoeconomics/ or exp Economic Aspect/ or pe.fs.
7	(1 or 2) and (3 or 4) and (6 or 5)
8	limit 7 to english language

Appendix 1. Search strategies.

No.	Database: Health Economic Evaluations Database ([HEED] online) 2008
	Search history
1	(((breast* OR mammar*) AND (neoplasm* OR metastas* OR carcinoma* OR tumour* OR tumor*)) AND (Trastuzumab OR herceptin)): all fields
No.	**Database: Ovid MEDLINE® in-process & other nonindexed citations & Ovid MEDLINE® 1950 to present**
	Search history
1	exp "Costs and Cost Analysis"/
2	(value of life or economics, medical or economics, pharmaceutical or models, economic or markov chains or monte carlo method or uncertainty).sh.
3	economics.fs.
4	(quality of life or quality-adjusted life years).sh.
5	economics.sh.
6	((econom$ or cost or costly or costing or costed or prices or pricing or discount or discounts or discounted or discounting or budget$ or afford$ or pharmacoeconomic$ or pharmaco) adj1 economic$).ti,ab.
7	(decision adj1 (tree$ or analy$ or model$)).ti,ab.
8	((value or values or valuation) adj2 (money or monetary or life or lives)).ti,ab.
9	(QOL or QOLY or QOLYs or HRQOL or QALY or QALYs).ti,ab.
10	((quality adj1 life) or (wilingness adj1 pay) or (quality adj1 adjusted life year$) or (sensitivity adj analysis) or quality adjusted life expectanc$).ti,ab.
11	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
12	*Breast Neoplasms/
13	*Breast/
14	mammary.mp.
15	13 or 14
16	*Neoplasms/
17	metastas$.mp.
18	*Carcinoma/
19	tumour$.mp. [mp=ti, ot, ab, nm, hw]
20	cancer$.mp. [mp=ti, ot, ab, nm, hw]
21	tumor$.mp. [mp=ti, ot, ab, nm, hw]
22	neoplasm$.mp. [mp=ti, ot, ab, nm, hw]
23	neoplasm$.mp. [mp=ti, ot, ab, nm, hw]
24	16 or 17 or 18 or 19 or 20 or 21 or 22 or 23
25	15 and 24
26	12 or 25
27	trastuzumab.mp. [mp=ti, ot, ab, nm, hw]
28	herceptin.mp. [mp=ti, ot, ab, nm, hw]
29	Receptor, erbB-2/
30	Receptor, Epidermal Growth Factor/
31	her2.mp.
32	27 or 28 or 29 or 30 or 31
33	26 and 32
34	11 and 33
35	limit 34 to english language
No.	**Database: PubMed**
	Search history
23	Search ("1"[Publication Date] : "2008/10/31"[Publication Date]) AND (#22)
22	Search #21 Limits: English
21	Search #20 Limits: Humans
20	Search #19 and #16 and #7

Appendix 1. Search strategies.	
19	Search #17 or #18
18	Search "receptor, erbb 2"[MeSH Terms]
17	Search herceptin or trastuzumab or her2[Text Word]
16	Search "breast neoplasms"[MeSH Terms] or #15
15	Search #10 and #14
14	Search #11 or #12 or #13
13	Search metastas* OR tumour* OR tumor* OR cancer* OR neoplasm* OR carcinoma*[Text Word]
12	Search "carcinoma"[MeSH Terms]
11	Search "neoplasms"[MeSH Terms]
10	Search #8 or #9
9	Search mammary[Text Word]
8	Search "breast"[MeSH Terms]
7	Search #1 or #2 or #3 or #4 or #5 or #6
6	Search QOL or HRQOL or QALY*
5	Search cost effectiveness or cost benefit or cost utility
4	Search "cost*effectiveness" or "cost*benefit" or "cost*utility"
3	Search "value of life" or economics or "medical economic*" or "pharmaceutical model*" or economic or "markov chain*" or "monte carlo" or "uncertainty analys*"[Text Word]
2	Search "quality of life"[All Fields] AND "quality adjusted life"[All Fields] AND "health related quality of life"[All Fields]
1	Search "cost benefit analysis"[MeSH Terms] OR "cost control"[MeSH Terms] OR "cost of illness"[MeSH Terms] OR "cost savings"[MeSH Terms] OR "costs and cost analysis"[MeSH Terms].

2.3 PAPER 2

Ferrusi, I. L., et al. Do economic evaluations of targeted therapy provide support for decision makers? Am. J Manag. Care 2011; 17(suppl 5): SP61-SP70.

POLICY

Do Economic Evaluations of Targeted Therapy Provide Support for Decision Makers?

Ilia L. Ferrusi, PhD(c); Natasha B. Leighl, MD; Nathalie A. Kulin, MSc; and Deborah A. Marshall, PhD

Objective: Decision makers must make decisions without complete information. That uncertainty can be decreased when economic evaluations use local data and can be quantified by considering the variability of all model inputs concurrently per international evaluation guidelines. It is unclear how these recommendations have been implemented in evaluations of targeted cancer therapy. By using economic evaluations of adjuvant trastuzumab, we have assessed the extent to which decision support recommendations were adopted.

Study Design: Systematic review.

Methods: Published economic evaluations of adjuvant trastuzumab treatment in early-stage breast cancer were examined as an established example of targeted therapy. Canadian, United Kingdom, and US economic evaluation guidelines were reviewed to establish extraction criteria. Extraction characterized the use of effectiveness evidence and local data sources for model parameters, sensitivity analysis methods (scenario, univariate, multivariate, and probabilistic), and uncertainty representation (ie, cost-effectiveness plane, scatterplot, confidence ellipses, tornado diagrams, cost-effectiveness acceptability curve).

Results: Fifteen economic evaluations of adjuvant trastuzumab were identified in the literature. Local data were used to estimate costs (15 of 15) and utilities rarely (2 of 15) but not trastuzumab efficacy. Univariate sensitivity analysis was most common (12 of 15), whereas probabilistic analysis was less frequent (10 of 15). Two-thirds of all studies provided visual representation of results and decision uncertainty.

Conclusion: Authors of adjuvant trastuzumab economic evaluations rarely use local data beyond costs. Quantification of uncertainty and its representation also fell short of guideline recommendations. This review demonstrates that economic evaluations of adjuvant trastuzumab, as an example of targeted cancer therapy, can be improved for decision-making support.

(Am J Manag Care. 2011;17(5 Spec No.):SP61-SP70)

For author information and disclosures, see end of text.

Economic evaluation is a tool used by policy and decision makers to address the relationship between clinical effects and costs associated with diagnosis, treatment, adverse effects, supportive healthcare, and life gained or lost. Payers, providers, and physicians can use economic evaluations to inform drug formulary listing, procedure or device reimbursement, and patient care decisions.[1-8] Decision analytic models have provided valuable support for health policy decisions ever since the Centers for Disease Control first presented such evidence to support vaccine recommendations in the late 1960s.[2] More recently, these methods are being applied to targeted drug therapies.

Targeted therapies, or personalized medicines, allow physicians to tailor treatment to individual patients. These medicines exert their effect by specifically targeting biologic processes via gene or protein expression[9] and, though costly, can potentially offer substantial clinical and economic offsets by avoiding ineffectual treatment and minimizing adverse effects. Therefore, decision analytic modeling and economic evaluation of targeted therapies are powerful tools with which clinical efficacy and costs can be weighed against standard care. Nonetheless, care must be taken to ensure that analyses are conducted in a manner that supports informed healthcare decision making. Many countries have outlined explicit economic evaluation guidelines to encourage appropriate conduct for decision-making purposes. To date, it is unclear how closely researchers have followed guidelines. Understanding how economic evaluations of targeted therapies are designed to inform decision making could enhance the health policy and managed-care environments.

In this article, we examine how economic analyses of targeted therapy were conducted with a focus on informing healthcare decisions from the payer's perspective. Given its widespread uptake and considerable success in the treatment of breast cancer, trastuzumab (Herceptin; Genentech, South San Francisco, CA) was chosen for assessment. Two decades of clinical study and application have facilitated several economic evaluations of the drug and this systematic review examines those evaluations to understand whether analyses of targeted therapy were reported in a manner that supports informed healthcare decision making. We used economic evaluation guidelines from Canada, the United Kingdom, and the United States to establish decision support criteria. Our review focuses on recommendations specifically designed to aid the decision-making process by increasing the

In this issue
Take-Away Points / SP62
www.ajmc.com
Full text and PDF

Take-Away Points

Economic evaluations of adjuvant trastuzumab, as an example of targeted therapy, can better support informed decision making through increased use of local data to inform model parameters and quantification and graphic communication of decision uncertainty. Data reflecting local practice is rarely used to inform model parameters beyond costs.

- Joint analysis of parameter uncertainty using probabilistic sensitivity analysis was reported in two-thirds of reviewed studies.
- Graphics aimed at communicating the results of economic analysis and sensitivity analysis were provided in two-thirds of all studies.
- These methods are supported by several international guidelines.

relevance of the economic evaluation to the decision maker's setting and encouraging quantification and representation of decision uncertainty.

METHODS

Systematic Search Strategy and Study Selection

A search strategy was previously developed[10] to identify published, peer-reviewed economic analyses of trastuzumab in the adjuvant treatment of breast cancer. The search encompassed literature published through October 2008 that were indexed in Biosis, Cochrane, the Centre for Reviews and Dissemination, EconLit, EMBASE, the Health Economic Evaluations Database, MEDLINE, and PubMed electronic databases; we updated EMBASE and MEDLINE searches to February 2011. Only English language citations were considered. Economic evaluations were included if they represented original research; considered 2 or more alternatives in an incremental of cost-effectiveness, cost-utility, cost-benefit, or cost-minimization; and focused on the evaluation of trastuzumab therapy in the adjuvant setting. Abstracts were reviewed independently by 2 assessors, and relevant articles were obtained in full for additional review. Selection of studies on the basis of reviews of the full articles was conducted by a single reviewer, and a random sample was verified independently.

Data Extraction

We reviewed Canadian,[11] United Kingdom,[12] and US[13] national drug or drug and device economic evaluation guidelines to identify recommendations for increasing the relevance of the analysis to the decision maker's setting and quantification and representation of decision uncertainty. The items identified from each guideline were then extracted from included studies. The items selected for abstraction are listed in Table 1 along with the relevant guidance from each country. We excluded the recommendation to model local standard care and practice patterns because of the difficulty in identifying and verifying local patterns across international treatment settings and language barriers.

Data was extracted to a single form for data input and decision uncertainty. Here, we use decision uncertainty to represent our understanding of the likelihood that the result predicted by an economic evaluation will occur in practice. To understand how authors made each evaluation relevant to the decision maker's setting, we extracted the source for the following parameters and categorized the source as "local data" or "literature": human epidermal growth factor receptor 2 (HER2) test properties, trastuzumab efficacy, risk of recurrence or survival, cost, and utility estimates. For an item to be considered local data, the model parameter needed to be derived from actual practice in the jurisdiction of the evaluation or measured from the disease population of that jurisdiction. For example, health state utilities used in an economic evaluation in the United States were considered local if the utilities were measured from a US population of patients with the disease of interest.

Quantification and representation of decision uncertainty was also documented. We extracted parameter type (stochastic [point estimate selected at random from a distribution] or deterministic [single point estimate]) to gain an understanding of the approach used to represent the "best guess" estimate of any variable considered in the evaluation. The methods of assessing uncertainty in those parameters and assumptions (termed sensitivity analysis) were subsequently extracted. Use of univariate, multivariate, scenario, or probabilistic sensitivity analysis was noted, including which parameters were assessed by each method. It was crucial to distinguish the methods of sensitivity analysis, as each serves a different purpose. Univariate analysis involves changing a single parameter estimate at a time to understand how that parameter influences results.[14] Multivariate or scenario analysis involves changing multiple variables simultaneously, usually to represent some alternative set of circumstances, to understand the impact on results.[14] Univariate and multivariate analysis most frequently employ deterministic parameters. Finally, probabilistic analysis involves assigning distributions to model parameters (stochastic) and allowing each to vary randomly and concurrently to generate an empirical distribution for the cost-effectiveness ratio.[14] We also documented whether visual representation of results and uncertainty was provided and the type of graphic used to represent that uncertainty (collectively termed decision aids). In this context, provision of decision aids was defined as clear graphic presentation of the cost-effectiveness plane with a scatterplot or confidence ellipses or of univariate sensitivity analysis results or cost-effectiveness acceptability curves (CEACs) with tornado diagrams per the reviewed guidelines. Results presented

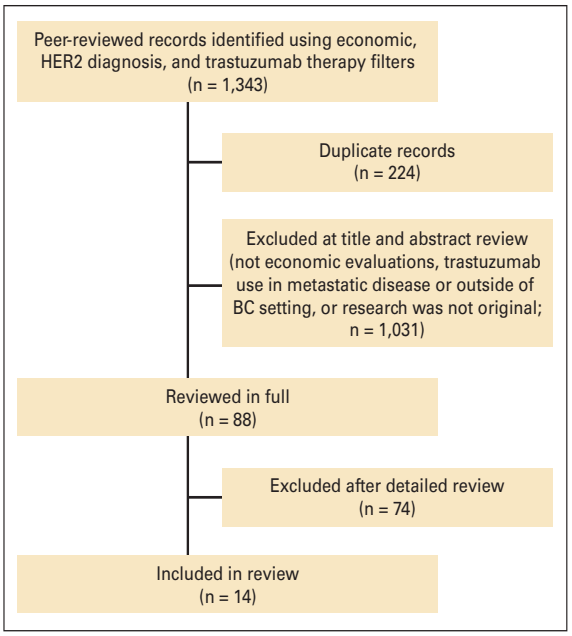
■ **Table 1.** Summary of Canadian, United Kingdom, and US Guidelines Specific to Efforts to Aid in the Decision-Making Process

	Data Input Recommendation					Methodology Recommendation	
Country	**Guideline Type**	**Purpose**	**Effectiveness**	**Cost**	**HRQOL**	**Sensitivity Analysis**	**Decision Uncertainty**
Canada[11]	Drug and nondrug technologies	To support the production of "credible and standardized" economic evaluations for use in the Canadian publicly funded decision-making environment	Systematic review; incorporate real-world factors that modify the effect if feasible (ie, patient adherence, diagnostic screening accuracy, physician compliance and skill, entire episode of care for devices)	Identify, measure, and value resources relevant to the decision makers' setting (including use of local unit costs) Base resource use estimates on Canadian routine practice	Preferred measurement from the general public that funds the healthcare system	Minimum of deterministic analysis for all inputs but should include additional methods Probabilistic analysis encouraged to assess parameter uncertainty	Quantify the contribution of each parameter to decision uncertainty Value of information methods suggested Tornado diagrams, cost-effectiveness planes, and CEACs recommended as visual representation
United Kingdom[12]	Drug and nondrug technologies	To provide an overview of appropriate methods for technology appraisal (including systematic review, economic evaluation) for submission to NICE	Evidence of effectiveness preferred over efficacy Systematic review required Typical patients, routine clinical practice, clinically relevant outcomes should all be included	Resources should be valued using prices relevant to the NHS	Should be elicited from a representative sample of the United Kingdom population	Distributions should be assigned to characterize parameter uncertainty Probabilistic sensitivity analysis is preferred Scenario analysis should be used to examine assumptions	Value of information analysis suggested Confidence ellipses, scatter plots, and CEACs appropriate for visually representing uncertainty
United States[13]	Drug technologies	To improve the "scope, quality, and relevance" of economic evaluations provided to the healthcare system evaluators and formulary decision makers	Transformation of efficacy estimates into effectiveness is encouraged by incorporating estimates of real-world factors (eg, adherence, comorbidities)	Opportunity costs relevant to decision maker	Not described	One-way deterministic analysis of all variables Scenario analysis, net benefit, probabilistic analysis, and CEACs recommended	One-way sensitivity analysis results should be presented in a figure (eg, tornado diagram) CEAC recommended if probabilistic analysis conducted

CEAC indicates cost-effectiveness acceptability curve; HRQOL, health-related quality of life; NHS, National Health Services; NICE, National Institute for Health and Clinical Excellence.

on the cost-effectiveness plane as a scatterplot or with confidence ellipses give the reader a sense of the distribution of incremental cost-effectiveness ratio (ICER) results. The CEAC shows the probability that a given intervention is more cost-effective than its comparator(s) over a range of willingness-to-pay values, providing the decision maker with an estimate of the likelihood that choosing to adopt the intervention would in fact be the right choice.[15] We also considered value of information (VOI) analysis, because this method was suggested by both Canadian and United Kingdom guidelines.[11,12]

Figure. The Flow of Studies Through the Review Process to Identify Health Economic Evaluations of Adjuvant Trastuzumab in Early-Stage Breast Cancer



BC indicates breast cancer.

Moreover, VOI relates the decision uncertainty of the model or specific parameters to the cost of conducting additional research to decrease that uncertainty[16] and therefore provides information to support decision making.

RESULTS

Search Results

The updated MEDLINE and EMBASE searches returned an additional 385 citations to the 958 citations previously identified. Duplicate citations accounted for 224 of the total, which left 1119 for review. Abstract review resulted in the exclusion of an additional 694 citations. A total of 15 studies remained after application of the inclusion criteria during full citation review. The Figure summarizes the study identification and selection process. The 2006 National Institute for Health and Clinical Excellence (NICE) report on the use of trastuzumab in early-stage breast cancer was included with the additional extraction of data from the related manufacturer's submission, which was available from the NICE Web site.[17] Several conference abstracts were identified but not included in the review because complete peer-reviewed articles were not available.[18-24]

A brief synopsis of economic evaluation methods, settings, and findings of the reviewed articles is presented in 2009 US dollars in Table 2. Overall, trastuzumab therapy was associated with an ICER deemed cost-effective in early-stage breast cancer by the majority study authors.[10] Additional studies identified in this updated review are generally consistent with that finding. However, Skedgel et al[27] noted that the cost-effectiveness of adjuvant trastuzumab in Canada could exceed the widely cited $50,000 per quality-adjusted life-year and $100,000 per quality-adjusted life-year thresholds and that this finding was largely dependent on the assumed duration of trastuzumab benefit. Indeed, 10 of 15 studies in the United States and international settings noted sensitivity to the assumed duration of trastuzumab benefit (typically 5 years) or the relative risk reduction associated with therapy. This suggests that follow-up on the long-term benefits of trastuzumab and the relative benefit of 52-week therapy compared with 9-week therapy will be crucial to understanding its cost-effectiveness in the adjuvant setting. Most authors did not consider local willingness-to-pay thresholds when concluding the cost-effectiveness of trastuzumab. The choice of testing strategy significantly impacted that ICER when test properties were modeled in conjunction with treatment. Some analyses suggested that a 9-week trastuzumab regimen[28] could result in potential cost savings compared with 52-week therapy[29] but that additional long-term data were needed. The results of several studies were sensitive to the cost of trastuzumab therapy.

Relevance to the Decision Maker's Setting

Table 3 lists the data sources reported among analyses of trastuzumab. Cost data were locally derived in all studies. Costing methods often reflected trial protocols or other published cost studies, although 2 authors used microcosting to reflect local practice patterns.[42,48] All remaining parameter categories were rarely informed by locally derived sources. Measures of treatment efficacy and utility estimates were sourced from the literature in all reports. Two authors used at least some utilities derived from local studies.[42,48]

Quantification of Decision Uncertainty and Decision Aids

Univariate analysis of deterministic parameters was conducted in the majority of the adjuvant trastuzumab studies. Despite the widespread guideline support for probabilistic sensitivity analysis, this technique was used in 10 of 15 studies, particularly in more recent publications. Additionally, multivariate or scenario analyses were conducted in 6 studies, and 2 studies used univariate analysis exclusively.[34,42] CEACs were presented in 8 of 10 studies that used probabilistic analysis. Beyond the CEAC, the tornado diagram[43] and the scatterplot[30,47] were the only other decision aids provided. No trends were noted with respect to international settings and sensitiv-

Table 2. Updated Summary of Findings From a Previous Review of Adjuvant Trastuzumab Therapy Economic Analyses

Analysis Feature	Updated Summary of Review Findings for Adjuvant Therapy (n = 15)
Perspectives[a]	Payer (n = 12) Societal (n = 2) Partial societal (n = 2)
International settings[a,b]	Australia, Belgium (n = 2), Canada, China, Finland, Italy, Japan, the Netherlands, Norway, Sweden, Switzerland, United Kingdom, United States (n = 3)
Methods	Analytic approach: cost-effectiveness (n = 3), cost-utility (n = 12) Model type: Markov model (n = 12), mixed model (n = 1), unknown (n = 2) Simulation type: hypothetical cohort (n = 15)
Time horizons[a]	2 to 5 years (n = 1) 6 to 15 years (n = 3) 16 to 51 years (n = 8) Lifetime (n = 7)
Results (ICER against comparator)[c]	ICER estimates varied widely for 52-week therapy ranging from $7902/QALY in Beijing (lifetime) to $126,580/QALY in Canada (lifetime) and from $24,822/LYS in Japan (lifetime) to $57,544/LYS in Belgium (lifetime)
Most influential parameters	Drug cost (n = 10) Discount rate (n = 8) Duration of survival benefit (n = 8) Trastuzumab relative risk reduction (n = 6) Test sensitivity/specificity (n = 2)
Concluding remarks	Trastuzumab was largely considered cost-effective in the adjuvant setting with 9-week therapy demonstrating lower ICERs than 52-week treatment; hesitation was expressed over the quality of the evidence for 9-week therapy vs 52-week The choice of testing strategy can significantly impact trastuzumab cost-effectiveness Drug price and the duration of trastuzumab benefit were seen as key sources of variation in ICER estimates

Data adapted.[10]
ICER indicates incremental cost-effectiveness ratio; LYS, life-year saved; QALY, quality-adjusted life-year.
[a]Does not add up to 15 because some authors considered more than 1 international setting, perspective, or time horizon.
[b]n = 1 unless otherwise noted.
[c]All currencies converted to 2009 US$ by using annual average exchange rates[25] and the medical component of the US consumer price index.[26]

ity analysis conduct. A summary of uncertainty analysis methods is provided in Table 4.

DISCUSSION

Our findings point to several avenues along which economic analyses of trastuzumab-targeted therapy can be improved for decision-making purposes. We noted that local data were rarely incorporated to inform parameters beyond the cost category, thus limiting outcome relevance to the decision maker. Although the inclusion of local costs was unanimous in this review, most authors derived resource use from clinical trial protocols or other published cost analyses. We are unable to comment on the treatment modalities with local practice patterns, given that many authors failed to comment on the similarity of the comparator treatment option to local standard care. Other parameters, such as treatment efficacy, present important challenges in terms of availability of local data and relevance, which often necessitate the use of preexisting trial data for economic evaluations.

However, as Phillips et al[50,51] point out, there are important and unanswered questions about the use of targeted therapies in the real world. Which patients get tested and treated? How accurate is HER2 testing in the clinical setting? What testing and treatment approaches are used to direct targeted HER2 therapy in actual clinical practice? These questions cannot be answered with hypothetical cohort simulations informed primarily by data from trials or early-stage applications. We recognize that not all economic analyses can be informed by local or pragmatic trial data, but the establishment of the Coverage With Evidence Development framework by the Centers for Medicare and Medicaid Services and the recent National Institutes of Health push for comparative effectiveness research[52] attest to the growing need for this type of evidence in decision making.[53]

Table 3. Relevance of the HER2 Testing or Trastuzumab Treatment Economic Evaluation to the Decision Maker's Setting

	Source of Data			
Author	Test Accuracy	Trastuzumab Efficacy, Risk of Recurrence, or Survival	Cost	Utility Estimates
Blank[30]	Literature[31,32]	Literature[29]	Local unit costs and literature	Literature; Swedish estimates
Chen[33]	Not modeled	Literature[29]	Local unit costs	Literature; United Kingdom estimates
Dedes[34]	Not modeled	Literature[18,28,29]	Local unit costs	Not modeled
Essers[35]	Not modeled	Literature[29]	Local unit costs but on the basis of United Kingdom resource use	Literature; Swedish estimates
Garrison[36]	Not modeled	Literature[37]	Local unit costs and literature	Literature
Kurian[38]	Not modeled	Literature[37]	Local unit costs and literature	Literature
Liberato[39]	Not modeled	Literature[21,37]	Local unit costs for 2 countries	Literature
Lidgren[48]	Literature[40]	Literature[19]	Local unit costs; future costs from literature	Literature; studies conducted locally
Millar[41]	Not modeled	Literature[42,37]	Local unit costs	Literature
Neyt[43]	Not modeled	Literature[28,29]	Local unit costs and literature	Not modeled
NICE[44]	Not modeled	Literature[29]	Manufacturer study and data obtained from Abacus	Manufacturer study and literature
Norum[42]	Not modeled	Literature[37]	Local unit costs	Literature; some utilities derived locally
Shiroiwa[45]	Not modeled	Literature[46]	Local unit costs	Not modeled
Skedgel[27]	Not modeled	Literature[29,46]	Local cost studies at cancer unit	Literature
Van Vlaenderen[47]	Not modeled	Literature[29]; underlying progression rates from Belgian registry	Local unit costs and literature	Literature; largely US estimates

Data sources were extracted and reviewed for relevance by evaluating whether the source was locally derived from the decision maker's setting.
HER2 indicates human epidermal growth factor receptor 2; NICE, National Institute for Health and Clinical Excellence.

Our work demonstrates that analyses of targeted therapy generally fall short of ensuring relevance through the use of local data. Locally derived, pragmatic evidence requires both the infrastructure and the time to collect long-term follow-up in the early-stage breast cancer setting given the significant improvement in mortality experienced as a result of early detection and adjuvant therapy.[54] However, Poncet et al[55] provide an example of how to use local, pragmatic evidence to inform health economic analyses in the metastatic setting. The difficulty in obtaining local clinical evidence is not exclusive to targeted therapy, but the importance of local relevance is heightened because of variability in real-world test performance and the high cost of many (biologic) targeted therapies. For example, a Canadian analysis of testing strategies[56] suggests that variation in national testing practice significantly impacted cost-effectiveness estimates.

Quantification of decision uncertainty and presentation of decision aids also fell short of supporting informed decision making. Exclusive use of univariate analysis is no longer sufficient for assessing parameter uncertainty[57] given that the ICER is affected by the shared uncertainty in multiple model inputs. Probabilistic sensitivity analysis is accepted as a much more powerful tool for addressing this. Such information is crucial to decision makers, who must make decisions for highly variable populations (eg, according to age or comorbidity) or for treatment settings that may differ from the studies used to inform the evaluation. Decision aids were provided in two-thirds of adjuvant trastuzumab evaluations; Blank et al[30] pro-

■ **Table 4.** Quantification of Representation of Decision Uncertainty in Published Economic Evaluations of HER2 Testing and Trastuzumab Targeted Therapy

Author	Model Parameter (deterministic vs stochastic)	Scenario Analyses	Univariate Sensitivity Analysis (variables)	Multivariate Sensitivity Analysis (variables)	Probabilistic Sensitivity Analysis (variables)	Decision Aids Provided?
Blank[30]	Deterministic and stochastic	None	Trastuzumab price, HER2 test costs, local/regional and metastatic costs (all ± 30%); discount rate (0, 6%); HER2 prevalence (15, 25%)	None	All	Tornado diagram, scatterplot on cost-effectiveness plane, CEAC
Chen[33]	Deterministic	None	MBC, local/regional recurrence Y1 costs, and trastuzumab costs (± 20%) Probability of progression to MBC, local/regional recurrence (+50%) Discount rate for life-years (3%), costs (0%, 5%)	None	None	CEAC (unclear how CEAC was produced without stochastic parameters)
Dedes[34]	Deterministic	All conducted with HERA data only Clinical benefit of trastuzumab limited to 3 years 20% rate of trastuzumab retreatment in MBC 80% rate of trastuzumab retreatment in MBC Trastuzumab administered in centers saving redundant drug discount rate, 3%	HERA[29] data: trastuzumab price, cost of MBC and local/regional recurrence, efficacy of trastuzumab (all ± 10%, ± 20%, ± 30%); discount rate FinHer[28] data: risk of recurrence and MBC (5% and 95% CI)	None	None	No
Essers[35]	Deterministic and stochastic	None	Unclear, but univariate analyses conducted; model sensitive only to use of trastuzumab at MBC recurrence	None	Yes, although stochastic variables were unclear	CEAC
Garrison[49]	Deterministic	Payer perspective Societal perspective	Discount rate, trastuzumab, death and HER2 and cardiac diagnostic costs, recurrence rates, utilities, cost of cardiac toxicity, recurrence-associated mortality	All varied to upper or lower bound	None	No
Kurian[38]	Stochastic	None	Discount rate, trastuzumab, MBC, and cardiac toxicity costs; survival; cardiac toxicity recovery; utility of trastuzumab and cardiac toxicity; cardiac toxicity–associated mortality	None	All	No
Liberato[39]	Deterministic and stochastic	Assumed characteristics of HERA trial[29]	None	Conducted as scenario	All	CEAC
Lidgren[48]	Deterministic and stochastic	None	Costs of trastuzumab, recurrence and AEs, recurrence risks, mortality, and trastuzumab efficacy all varied ± 30%; future costs, duration of trastuzumab effect, HER2 prevalence also varied	IHC test sensitivity and specificity	All	CEAC

(Continued)

Table 4. Quantification of Representation of Decision Uncertainty in Published Economic Evaluations of HER2 Testing and Trastuzumab Targeted Therapy *(Continued)*

Author	Model Parameter (deterministic vs stochastic)	Scenario Analyses	Univariate Sensitivity Analysis (variables)	Multivariate Sensitivity Analysis (variables)	Probabilistic Sensitivity Analysis (variables)	Decision Aids Provided?
Millar[41]	Deterministic	Age 50 years given 52-week trastuzumab across 95% CI Age 70 years given 52-week trastuzumab across 95% CI	Discount rate, age at diagnosis, cost of treating MBC, relapse, other diseases and cardiac toxicity, duration of trastuzumab effect, risk of distant recurrence after 52-week trastuzumab across 95% CI bounds, risk of distant recurrence after 9-week trastuzumab across 95% CI bounds	Conducted within scenarios	None	No
Neyt[43]	Stochastic	Subgroup analyses according to age group, stage, and discount rate	None	Conducted as scenario	All but HER2 test, administration, and follow-up costs	Tornado diagram
NICE[44]	Deterministic and stochastic	None	Utilities, discount rate, trastuzumab efficacy, costs, recurrence, and progression rates	None	All	CEAC
Norum[42]	Deterministic	None	All parameters ± 25%; discount rate	None	None	No
Shiroiwa[45]	Stochastic	None	Discount rate, recurrence rate, PE costs, palliative care costs	None	All	CEAC
Skedgel[27]	Stochastic	Assumed 3-year trastuzumab benefit Assumed 5-year trastuzumab benefit	Discount rate, cost of palliative treatment; threshold analyses to determine what parameters could produce ICER estimate below CA$50,000/QALY and CA$100,000/QALY thresholds	None	All	CEAC
Van Vlaenderen[47]	Stochastic	None	None	None	All	Scatterplot with willingness-to-pay threshold

Methods of sensitivity analysis, parameters analyzed, and decision aids, if any, were abstracted.
AEs indicates adverse events; CEAC, cost-effectiveness acceptability curve; CI, confidence interval; FinHer, Finland Herceptin; HER2, human epidermal growth factor receptor 2; HERA, Herceptin Adjuvant; ICER, incremental cost-effectiveness ratio; IHC, immunohistochemistry; MBC, metastatic breast cancer; NICE, National Institute for Health and Clinical Excellence; QALY, quality-adjusted life-year.

vide an excellent example of the use of scatterplot, tornado, CEAC, and diagrams to represent decision uncertainty graphically. These graphic presentations of base case or sensitivity analysis results are promoted as a tool to communicate aspects of structural, parameter, or assumption-based sensitivity to the reader in a nontechnical manner. This makes the interpretation of the ICER and its variability more accessible to decision makers such as plan managers, who may not have technical expertise in economic evaluation. The use of graphic decision aids was unanimously supported by Canadian, United Kingdom, and US economic evaluation guidelines and is, in fact, promoted by several other jurisdictions.[58-60] CEACs are also a natural extension of probabilistic sensitivity analysis; it was therefore surprising to observe that CEACs were not provided in all studies conducting probabilistic analysis. Conversely, VOI is much more complex to conduct, and the absence of this decision aid was expected.

We recognize that there are some limitations to this review. Several methodologic aspects of cost-effectiveness analysis are challenging to incorporate in publications given current word limits. For example, model calibration was rarely reported in review of cancer simulation models, despite the importance

of this method in ensuring that a model accurately predicts outcomes for the population of interest.[61] Calibration was also recommended by Canadian economic evaluation guidelines as a method of ensuring relevance to the decision maker.[11] We encourage authors and journals to provide this information via online support materials whenever possible.

The exclusive assessment of trastuzumab as a targeted therapy in this review may limit the generalizability of our findings. We believe that the trends observed herein are not restricted to trastuzumab and that many relate directly to economic evaluations of other targeted therapies, particularly those with prerequisite diagnostic tests. We also acknowledge that some of the citations included in this review may not be intended to inform specific local decisions. Indeed, hypothetical cohort evaluations are designed to predict outcomes in a theoretical population. Conversely, readers of the scientific literature, including physicians and formulary managers, require a more realistic understanding of the potential impacts of treatment or funding decisions. Many review agencies, such as the Canadian Association for Drugs and Technologies in Health and NICE, require a systematic review of economic evaluations within drug reimbursement submissions. This suggests that hypothetical cohort evaluations do factor into reimbursement decisions.

Additionally, the use of Academy of Managed Care Pharmacy (United States), the Canadian Association for Drugs and Technologies in Health (Canada), and NICE (United Kingdom) guidelines as benchmarks for economic evaluation methods is not validated. However, the recommendations included in this review for data sources and decision uncertainty are consistent with Danish,[58] French,[59] and German[60] guidelines and the final recommendations of the International Society for Pharmacoeconomics and Outcomes Research Real World Data[49] and Good Research Practices-Modeling Studies[62] task forces. The Canadian, United Kingdom, and US guidelines were applied to strike a balance between current methods and highly regarded guidelines in the absence of empirically validated economic evaluation methods for decision support.

Our findings suggest that economic evaluations of targeted therapy can be improved to support high-quality, informed decision making. Although real-world effectiveness estimates are often unavailable or difficult to generate, several other steps can be taken to ensure relevance to the decision maker's setting, including the incorporation of local utilization patterns to improve costing and behavioral assumptions. Quantification and representation of decision uncertainty can also be improved through the regular conducting of probabilistic sensitivity analysis, the provision of decision aids, and the practical application of VOI methods.

Acknowledgment

We would like to thank Elena Elkin, PhD, for her insightful commentary and contributions to this article.

Supported in part by an award from the Ontario Council on Graduate Studies and by Grants No. R01CA101849 and P01CA130818 from the National Cancer Institute (Bethesda, MD).

Authors' Disclosures of Potential Conflicts of Interest: The authors indicated no potential conflicts of interest.

Author Contributions

Conception and design: Ilia L. Ferrusi, Natasha B. Leighl, Deborah A. Marshall. ***Financial support:*** Deborah A. Marshall. ***Administrative support:*** Deborah A. Marshall. ***Collection and assembly of data:*** Ilia L. Ferrusi. ***Data analysis and interpretation: Ilia L. Ferrusi, Natasha B. Leighl, Nathalie A. Kulin, Deborah A. Marshall Manuscript writing:*** Ilia L. Ferrusi, Natasha B. Leighl, Deborah A. Marshall. ***Final approval of manuscript:*** Ilia L. Ferrusi, Natasha B. Leighl, Nathalie A. Kulin, Deborah A. Marshall.

Address Correspondence to: Deborah A. Marshall, PhD, Faculty of Medicine, University of Calgary, Health Research Innovation Centre, 3280 Hospital Dr NW, Room 3C56, Calgary, AB T2N 4Z6; e-mail: damarsha@ucalgary.ca.

REFERENCES

1. Berger ML, Teutsch S: Cost-effectiveness analysis: From science to application. *Med Care* 43:49-53, 2005 (suppl 7)

2. Buxton MJ, Drummond MF, Van Hout BA, et al: Modelling in economic evaluation: an unavoidable fact of life. *Health Econ* 6:217-227, 1997

3. O'Brien B: Economic evaluation of pharmaceuticals: Frankenstein's monster or vampire of trials? *Med Care* 34:DS99-DS108, 1996 (suppl 12)

4. Briggs AH, Sculpher MJ, Logan RP, et al: Cost effectiveness of screening for and eradication of *Helicobacter pylori* in management of dyspeptic patients under 45 years of age. *BMJ* 312:1321-1325, 1996

5. Pauker SG: Coronary artery surgery: the use of decision analysis. *Ann Intern Med* 85:8-18, 1976

6. Schoenbaum SC, McNeil BJ, Kavet J: The swine-influenza decision. *N Engl J Med* 295:759-765, 1976

7. Snider DE, Holtgrave DR, Dunet DO: Decision analysis, in Haddix AC, Teutsch SM, Shaffer PA, et al (eds): *Prevention Effectiveness. A Guide to Decision Analysis and Economic Evaluation.* Oxford University Press, New York, NY, 1996, pp 27-45

8. Weinstein MC, Toy EL, Sandberg EA, et al: Modeling for health care and other policy decisions: uses, roles, and validity. *Value Health* 4:348-361, 2001

9. Mancinelli L, Cronin M, Sadée W: Pharmacogenomics: the promise of personalized medicine. *AAPS PharmSci* 2:E4, 2000

10. Ferrusi IL, Marshall DA, Kulin NA, et al: Looking back at 10 years of trastuzumab therapy: What is the role of HER2 testing?—a systematic review of health economic analyses. *Per Med* 6:193-215, 2009

11. Canadian Agency for Drugs and Technologies in Health (CADTH): Guidelines for the Economic Evaluation of Health Technologies: Canada (ed 3). Canadian Association for Drugs and Technologies in Health, Ottawa, Ontario, Canada, 2006

12. National Institute for Health and Clinical Excellence (NICE): Guide to the Methods of Technology Appraisal. National Institute for Health and Clinical Excellence, London, United Kingdom, 2008

13. Academy of Managed Care Pharmacy (AMCP): AMCP Format for Formulary Submissions Version 2.1. Academy of Managed Care Pharmacy, Alexandria, VA, 2005

14. Drummond MF, Sculpher MJ, Torrance GW, et al: Critical assessment of economic evaluation, in Drummond MF, Sculpher MJ, Torrance GW, et al: *Methods for the Economic Evaluation of Health Care Programmes (ed 3).* Oxford University Press, New York, NY, 2005, pp 27-53

15. Fenwick E, Claxton K, Sculpher M: Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econom* 10:779-787, 2001

16. Claxton K: Bayesian approaches to the value of information: implications for the regulation of new pharmaceuticals. *Health Econ* 8:269-274, 1999

17. Hoffmann-La Roche: Herceptin (trastuzumab) NICE submission: achieving clinical excellence in the adjuvant treatment of HER-2 positive breast cancer. http://www.nice.org.uk/nicemedia/live/11585/33437/33437.pdf

18. Vernaglia PR, Cunha FM, Correa M, et al: Cost-effectiveness analysis of trastuzumab therapy in patients with early HER-2 positive breast cancer in Brazil. *Breast Cancer Res Treat* 106:S106, 2007

19. Hillner BE: Clinical and cost-effectiveness implications of adjuvant trastuzumab in HER2+ breast cancer trials. Presented at 2005 San Antonio Breast Cancer Symposium, San Antonio, TX, December 8-11, 2005 (abstr 5040)

20. Hornberger J, Kerrigan M, Foutel V: Cost-effectiveness of trastuzumab (Herceptin) for treatment of metastatic breast cancer. *Ann Oncol* 13:52, 2002

21. Ragaz J, Spinelli JJ: Cost-benefit estimates of adjuvant trastuzumab for early breast cancer. Presented at 2005 San Antonio Breast Cancer Symposium, San Antonio, TX, December 8-11, 2005 (abstr 2029)

22. Lopes GD: An evaluation of the societal costs and benefits of adjuvant trastuzumab in the treatment of early HER2neu-positive breast cancer in Singapore. *J Clin Oncol* 28, 2010 (abstr e16514)

23. Clark O, Paladini L, Clark LG, et al: Trastuzumab on adjuvant breast cancer treatment: evidence synthesis and health technology evaluation. *Value Health* 12:A289, 2009 (abstr PCN166)

24. Guiliani G, Ray JA, Urspruch A: Cost-effectiveness analysis of adjuvant therapy with trastuzumab for HER2-positive breast cancer in Italy utilizing follow-up data. *Value Health* 12:A276, 2009 (abstr PCN102)

25. US Federal Reserve: Foreign exchange rates (weekly). http://www.federalreserve.gov/releases/H10/default.htm

26. US Bureau of Labor Statistics: Consumer price index - all urban consumers, medical care; Series ID: SUUR0000SAM (2011)

27. Skedgel C, Rayson D, Younis T: The cost-utility of sequential adjuvant trastuzumab in women with Her2/Neu-positive breast cancer: an analysis based on updated results from the HERA trial. *Value Health* [epub ahead of print on March 10, 2009]

28. Joensuu H, Kellokumpu-Lehtinen P-L, Bono P, et al: Adjuvant docetaxel or vinorelbine with or without trastuzumab for breast cancer. *N Engl J Med* 354: 809-820, 2006

29. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, et al: Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* 353:1659-1672, 2005

30. Blank PR, Schwenkglenks M, Moch H, et al: Human epidermal growth factor receptor 2 expression in early breast cancer patients: a Swiss cost-effectiveness analysis of different predictive assay strategies. *Breast Cancer Res Treat* 124:497-507, 2010

31. Press MF, Sauter G, Bernstein L, et al: Diagnostic evaluation of HER-2 as a molecular target: an assessment of accuracy and reproducibility of laboratory testing in large, prospective, randomized clinical trials. *Clin Cancer Res* 11:6598-6607, 2005

32. Dendukuri N, Khetani K, McIsaac M, et al: Testing for HER2-positive breast cancer: a systematic review and cost-effectiveness analysis. *CMAJ* 176:1429-1434, 2007

33. Chen W, Jiang Z, Shao Z, et al: An economic evaluation of adjuvant trastuzumab therapy in HER2-positive early breast cancer. *Value Health* 12:S82-S84, 2009

34. Dedes KJ, Szucs TD, Imesch P, et al: Cost-effectiveness of trastuzumab in the adjuvant treatment of early breast cancer: a model-based analysis of the HERA and FinHer trial. *Ann Oncol* 18:1493-1499, 2007

35. Essers BA, Seferina SC, Tjan-Heijnen VC, et al: Transferability of model-based economic evaluations: the case of trastuzumab for the adjuvant treatment of HER2-positive early breast cancer in the Netherlands. *Value Health* 13:375-380, 2010

36. Garrison LP Jr, Lubeck D, Lalla D, et al: Cost-effectiveness analysis of trastuzumab in the adjuvant setting for treatment of HER2-positive breast cancer. *Cancer* 110:489-498, 2007

37. Romond EH, Perez EA, Bryant J, et al: Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med* 353:1673-1684, 2005

38. Kurian AW, Thompson RN, Gaw AF, et al: A cost-effectiveness analysis of adjuvant trastuzumab regimens in early HER2/neu-positive breast cancer. *J Clin Oncol* 25:634-641, 2007

39. Liberato NL, Marchetti M, Barosi G: Cost effectiveness of adjuvant trastuzumab in human epidermal growth factor receptor 2-positive breast cancer. *J Clin Oncol* 25:625-633, 2007

40. Elkin EB, Weinstein MC, Winer EP, et al: HER-2 testing and trastuzumab therapy for metastatic breast cancer: a cost-effectiveness analysis. *J Clin Oncol* 22:854-863, 2004

41. Millar JA, Millward MJ: Cost effectiveness of trastuzumab in the adjuvant treatment of early breast cancer: a lifetime model. *Pharmacoeconomics* 25:429-442, 2007

42. Norum J, Olsen JA, Wist EA, et al: Trastuzumab in adjuvant breast cancer therapy: a model based cost-effectiveness analysis. *Acta Oncol* 46:153-164, 2007

43. Neyt M, Huybrechts M, Hulstaert F, et al: Trastuzumab in early stage breast cancer: a cost-effectiveness analysis for Belgium. *Health Policy* 87:146-159, 2008

44. National Institute for Health and Clinical Excellence: Trastuzumab for the adjuvant treatment of early-stage *HER2*-positive breast cancer. National Institute for Health and Clinical Excellence, London, United Kingdom, 2008

45. Shiroiwa T, Fukuda T, Shimozuma K, et al: The model-based cost-effectiveness analysis of 1-year adjuvant trastuzumab treatment: based on 2-year follow-up HERA trial data. *Breast Cancer Res Treat* 109:559-566, 2008

46. Smith I, Procter M, Gelber RD, et al: 2-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer: a randomised controlled trial. *Lancet* 369:29-36, 2007

47. Van Vlaenderen I, Canon JL, Cocquyt V, et al: Trastuzumab treatment of early stage breast cancer is cost-effective from the perspective of the Belgian health care authorities. *Acta Clinica Belgica* 64:100-112, 2009

48. Lidgren M, Jönsson B, Rehnberg C, et al: Cost-effectiveness of HER2 testing and 1-year adjuvant trastuzumab therapy for early breast cancer. *Ann Oncol* 19:487-495, 2008

49. Garrison LP Jr, Neumann PJ, Erickson P, et al: Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. Value Health 10:326-335, 2007

50. Phillips KA, Liang SY, Van Bebber S, et al: Challenges to the translation of genomic information into clinical practice and health policy: utilization, preferences and economic value. *Curr Opin Mol Ther* 10:260-266, 2008

51. Phillips KA: Closing the evidence gap on the use of emerging testing technologies in clinical practice. *JAMA* 300:2542-2544, 2008

52. Steinbrook R: Health care and the American Recovery and Reinvestment Act. *N Engl J Med* 360:1057-1060, 2009

53. Tunis SR, Pearson SD: Coverage options for promising technologies: Medicare's 'coverage with evidence development.' *Health Aff (Millwood)* 25:1218-1230, 2006

54. Berry DA, Cronin KA, Plevritis SK, et al: Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med* 353:1784-1792, 2005

55. Poncet B, Bachelot T, Colin C, et al: Use of the monoclonal antibody anti-HER2 trastuzumab in the treatment of metastatic breast cancer: a cost-effectiveness analysis. *Am J Clin Oncol* 31:363-368, 2008

56. Dranitsaris G, Norris B, Hanna W, et al: Identifying the optimal testing of HER2 testing in patients with breast cancer: a Canadian economic evaluation. *Curr Oncol* 10:36-44, 2003

57. Drummond MF, Sculpher MJ, Torrance GW, et al: Critical assessment of economic evaluation, in Drummond MF, Sculpher MJ, Torrance GW, et al (eds): *Methods for the economic evaluation of health care programmes* (ed 3). Oxford University Press, New York, NY, 2005, pp 27-53

58. Alban A, Gyldmark M, Pedersen AV, et al: The Danish approach to standards for economic evaluation methodologies. *Pharmacoeconomics* 12:627-636, 1997

59. Collège des Économistes de la Santé: French Guidelines for the Economic Evaluation of Health Care Technologies. Collège des Économistes de la Santé, Paris, France, 2004, pp 1-86

60. Graf von der Schulenburg JM, Greiner W, Jost F, et al: German recommendations on health economic evaluation: third and updated version of the Hanover Consensus. Value Health 11:539-544, 2008

61. Stout NK, Knudsen AB, Kong CY, et al: Calibration methods used in cancer simulation models and suggested reporting guidelines. *Pharmacoeconomics* 27:533-545, 2009

62. Weinstein MC, O'Brien B, Hornberger J, et al: Principles of good practice of decision analytic modeling in health care evaluation: report of the ISPOR Task Force on Good Research Practices-Modeling Studies. *Value Health* 6:9-17, 2003 ■

CHAPTER 3

RETROSPECTIVE COHORT STUDY OF HER2 TESTING AND TRASTUZUMAB USE IN ONTARIO PRACTICE

PAPER 3†: Ferrusi, I. L., et al. Closing the Personalized Medicine Information Gap in Breast Cancer: HER2 Test Documentation Practice. Am. J Manag. Care 2013;19(1):838-44.

†Academic contributions by Ilia L. Ferrusi: concept and design, acquisition of data, analysis and interpretation of data, writing, critical revision and finalisation of the manuscript for important intellectual content.

SUPPLEMENTAL ANALYSES: Ferrusi, I. L. Earle C. C., Trudeau, M., Leighl, N. B., Pullenayegum, E., Khong, H., Hoch, J.S., Marshall, D. A.

3.1 PREFACE

Chapter 2 presented some of the challenges to assessing a personalised medicine having a companion diagnostic, such as trastuzumab, within a decision-analytic framework. A key finding was the importance of modelling the relationship between the companion diagnostic and its accompanying targeted therapy; in the case of trastuzumab as a personalised medicine, this includes HER2 test utilisation, testing alternatives, individual test sensitivity and specificity, test sequencing or confirmatory testing and assumptions about the concordance of treatment with test results. Phillips *et al.* noted the scarcity of information about routine personalised medicine clinical practice in the literature, and identified this as a key knowledge gap for cost-effectiveness evaluations of trastuzumab. [1] At the time that this study was undertaken, HER2 testing practice in Ontario, and the wider Canadian landscape, was largely unknown. Although guidelines were issued by a Canadian consensus panel of pathologists [2] to encourage optimal HER2 testing practice, uptake by provincial pathology processes was voluntary. Those guidelines recommend HER2 testing at diagnosis for all patients diagnosed with invasive BC using standardised and validated tests. Initial testing with IHC is recommended, with confirmatory retesting of IHC equivocal cases using FISH or another validated brightfield *in situ* hybridisation method. Studies of HER2 testing in Canadian practice were limited to a conference abstract reporting a Nova Scotian chart review.[3] This review at a

single Halifax hospital revealed HER2 testing was provided to 81% of early-stage BC patients [3], but did not detail the types of test(s) used or confirmatory testing patterns. Therefore, a study of HER2 testing practice in Ontario was undertaken to inform influential test parameters in a subsequent decision analysis of testing strategies to direct targeted adjuvant trastuzumab treatment in Ontario (Chapter 4). Specifically, information was gathered to answer the following questions for the Ontario setting:

1. Do early-stage BC patients receive a HER2 test at diagnosis?
2. What type of HER2 test (e.g. IHC, FISH) is used most commonly for HER2 testing?
3. When is confirmatory testing for HER2 used?
    a. What type of test is used for confirmatory testing?
4. What proportion of HER2 positive patients receive trastuzumab treatment?

The study design, described in detail in Paper 3, was shaped by the need to capture practice variation throughout the province or practice setting and therefore aimed to document testing practice as widely as possible. We considered several potential sources of provincial HER2 test information for this study. At the time of study initiation, provincial pathology registries (e.g. the Pathology Information Management System [PIMS] electronic data collection system managed by Cancer Care Ontario [CCO]) did not capture HER2-specific tumour pathology. In

fact, province-wide collection of HER2 status was not added to the Ontario Cancer Registry (OCR) until 2009 under new Collaborative Staging efforts.[4] We explored an alternative provincial database, the Ontario Laboratory Information System (OLIS). However, this eHealth Ontario initiative also lacked the HER2 test documentation required for our analysis.

Therefore, we deliberately pursued opportunities to generate a unique dataset through a combination of original data collection and linkage to administrative health data. Detailed information about HER2 testing could be obtained via chart or pathology report review. We elected to work with the repository of pathology reports held by CCO for the purposes of maintaining the OCR. Through consultations with CCO, we believed that this data source would provide a comprehensive snapshot of HER2 testing practice across the entire province for all incident BC cases, capturing reports from all Local Health Integration Networks (LHINs), practice settings (e.g. academic vs. community setting or cancer centre vs. non-cancer centre), physician types and urban settings. Moreover, these pathology reports had a pre-existing identifier relating to patient records in the OCR, and therefore other administrative health data, facilitating straightforward linkage of disparate data elements into individual records in the final dataset. Weighed against the cost and time required to conduct a chart review to derive similar HER2 data, our team of co-investigators agreed that centrally-held, registry-linked pathology reports provided the optimal balance of

detailed information and comprehensiveness. Indeed, our methodical review of OCR-linked pathology reports produced a rich dataset detailing the type, number and sequence of HER2 tests provided to early-stage BC patients. However, we also identified several reporting biases in the OCR-linked pathology source suggestive of more complete reporting for patients with advanced disease at diagnosis (Paper 3). The unexpected variability observed for HER2 test reporting ultimately limited our ability to conclusively comment on the first research question, and necessitated a separate analysis of reporting patterns from a healthcare system factor and clinical perspective (Paper 3).

Appendices following the paper provide all univariates (Appendix 1) a comparison of all models examining HER2 documentation, including sensitivity analyses and grouped variable results (Appendix 2, 3) not shown in the published version of Paper 3. The subsequent Supplemental Analyses section details the methods of analysis and results to address research questions about HER2 test types, confirmatory FISH testing and trastuzumab treatment in routine clinical practice (research questions two through four). The Supplemental Analyses inform the decision analytic model of alternative HER2 testing strategies to direct targeted trastuzumab treatment of early-stage BC patients (Chapter 4).

3.2 PAPER 3

Ferrusi, I. L., et al. Closing the Personalized Medicine Information Gap in Breast Cancer: HER2 Test Documentation Practice. Am. J Manag. Care 2013;19(1):838-44.

CLINICAL

Closing the Personalized Medicine Information Gap: HER2 Test Documentation Practice

Ilia L. Ferrusi, PhD; Craig C. Earle, MD; Maureen Trudeau, MD; Natasha B. Leighl, MD; Eleanor Pullenayegum, PhD; Hoa Khong, MD; Jeffrey S. Hoch, PhD; and Deborah A. Marshall, PhD

Background: Uncertainty about human epidermal growth factor receptor-2 (HER2) testing practice in Canada continues to hinder efforts to improve personalized medicine. Pathologists routinely perform HER2 assessment for all tumors > 1 cm, and pathology is reported centrally to the provincial cancer registry.

Objectives: To understand patterns of HER2 test documentation for early-stage breast cancer (BC) patients in Ontario's centralized pathology reporting system.

Study Design: Retrospective cohort study of central HER2 test documentation in early-stage BC patients diagnosed in 2006-2007.

Methods: Cohort and staging information was derived from cancer registry and admissions data. Linkage across administrative databases provided data on surgical and radiologic treatment, sociodemographic factors, diagnosis setting, and comorbidities. Pathology reports from the provincial cancer registry were reviewed for HER2 testing, hormone receptor, and grade. Unadjusted and adjusted odds ratios were calculated to determine factors related to HER2 documentation.

Results: A HER2 test was documented for 66% of 13,396 patients. HER2 documentation was associated with stage, hormone receptor, and tumor grade documentation. Higher stage and grade at diagnosis were also associated with HER2 documentation. All models suggested variable regional documentation patterns. Documentation did not differ by sociodemographic factors, presence of comorbidities, or surgical procedure.

Conclusions: Despite a universal testing policy, the rate of centralized HER2 test documentation was lower than expected and related to disease severity. Differences in regional reporting likely reflect ascertainment bias inherent to centralized pathology reporting rather than testing access. Improved HER2 reporting is encouraged for cancer registration, quality-of-care measurement, and program evaluation.

(Am J Manag Care. 2013;19(1):17-26)

For author information and disclosures, see end of text.

Breast cancer (BC) is the most frequently diagnosed cancer among Canadian women, with a projected incidence of 23,400 in 2011.[1] Early diagnosis and adjuvant treatment provide significant gains in life expectancy for women diagnosed with early-stage disease.[2] Recent advances in treatment focus on using genetic information to target treatments to patients who are likely to respond. One example is the human epidermal growth factor receptor-2 (HER2) oncogene and protein, first noted as predictors of overall survival and time to relapse in BC.[3] Amplification of the HER2 oncogene in 20% of cancers is associated with poor prognosis, aggressive tumor proliferation, and poorer response to chemotherapy.[3,4] Trastuzumab therapy has demonstrated significant improvements in disease-free survival and mortality in patients whose tumors overexpress HER2.[5,6] Testing for HER2 to identify treatment candidates is typically conducted by immunohistochemistry (IHC) or in situ hybridization techniques, most commonly fluorescence (FISH).[7,8] Differences in test accuracy and cost prompted the development of testing guidelines[7] to ensure efficient and accurate diagnosis of HER2-positive patients. Although less expensive, IHC is also less accurate.[7] Quantification of HER2 gene overexpression by FISH is more accurate, but also more expensive and difficult to conduct. Guidelines recommend the use of either IHC or FISH to detect HER2 overexpression and promote reflex FISH testing to clarify the HER2 status of IHC-equivocal tumors.[7,8]

Accurate identification of HER2-positive patients is crucial given the high cost of adjuvant trastuzumab therapy and the potential exposure of false-positive patients to cardiotoxic side effects. Adjuvant treatment guidelines recommend that all incident patients with invasive disease receive a HER2 workup.[9] Economic evaluations of HER2 testing and treatment demonstrate the clinical and economic costs of failure to accurately classify IHC-equivocal patients.[10-12] Studies of HER2 testing in the early days of metastatic therapy in the United States suggested an information gap[13] in HER2 documentation for 48% of eligible patients. More information is needed to gauge the quality of current practice and to establish a foundation for assessing the optimal use of personalized medicine in the real world. Without this information, it is difficult for administrators or researchers to understand issues related to access to testing, appropriateness of treatment, and cost-effective care. HER2 testing practice in Canada remains largely unreported in the

In this article
Take-Away Points / p18
www.ajmc.com
Full text and PDF

Take-Away Points

Population-based analysis indicated that tumor pathology for human epidermal growth factor receptor-2 (HER2) testing was not consistently reported to the central registry, despite universal access to HER2 testing in the Ontario public healthcare system.

- Locally performed HER2 tests were documented for 66% of patients at the centralized registry.
- Although HER2 test documentation was unrelated to income or urban residence, it was related to documentation of other pathology factors and disease severity measured by stage and tumor grade.
- Without improved or mandatory HER2 reporting to the central registry, program evaluation and health quality improvement studies are limited.

literature, particularly with respect to how testing is documented, what tests were performed, test results, and whether reflex testing is conducted. A sample of early-stage patients in Nova Scotia suggests that 81% of patients received a HER2 test, but provides no insight into the type of test(s) used.[14]

We aimed to describe centralized HER2 test documentation and testing patterns in Ontario and to gain insight into how to use and interpret these data. Our specific objectives were to (1) assess the availability of data to evaluate HER2 testing practices from a centralized source in a real-world setting; (2) describe reporting system, clinical, or sociodemographic factors associated with HER2 documentation in Ontario; and (3) describe HER2 test utilization with respect to test type and test sequencing.

METHODS

Study Design and Setting

A retrospective cohort design was used to study patients diagnosed with early-stage BC between 2006 and 2007 in the Canadian province of Ontario. This time frame allowed 6 months of lag time subsequent to the approval of adjuvant trastuzumab therapy in mid-2005.[15] The associated treatment guideline and relevant policies were implemented province-wide under the auspices of Cancer Care Ontario (CCO). As the provincial cancer agency, CCO is involved in screening, diagnostic, treatment, recovery, and palliative services to all patients diagnosed with cancer in the publicly funded Ontario healthcare system. The New Drug Funding Program of CCO administers the reimbursement of new, expensive systemic therapies, including trastuzumab. At the time of this study, adjuvant trastuzumab treatment was available to patients with HER2-positive tumors larger than 1 cm that were previously treated with chemotherapy.[9,15] Ontario's policy is to follow Canadian testing guidelines "…to test all patients with invasive breast cancer for HER2/*neu* at the time of diagnosis."[7,8] Testing was funded by the Ontario Ministry of Health and Long-term Care and routinely performed by pathologists irrespective of other clinical or pathologic factors. Tumor pathology should be reported centrally to the provincial cancer agency for maintenance of the cancer registry. Reimbursement for trastuzumab requires evidence of a positive HER2 test, but this is provided separately from registry reporting.

Research ethics board approval was obtained from St. Joseph's Healthcare Hamilton. The protocol was also approved by the privacy committees of CCO and the Institute for Clinical Evaluative Sciences. These agencies provided access to provincial and national administrative health data, and facilitated record linkage across data sources using anonymous patient identifiers. This manuscript was reviewed by CCO, the Institute for Clinical Evaluative Sciences, and the Ministry of Health and Long-term Care.

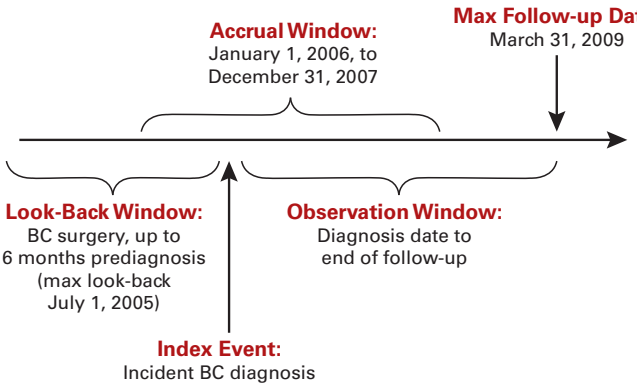
Participants

All female patients who were diagnosed with early-stage, invasive BC between January 1, 2006, and December 31, 2007, and were treated with surgery (modified radical or partial mastectomy, lumpectomy) within 6 months of diagnosis were eligible for the study. Patients with BC (*International Classification of Diseases, Ninth Revision* code 174) were identified from the Ontario Cancer Registry (OCR). The early-stage cohort was identified by eliminating metastatic and miscoded noninvasive carcinoma (stage 0) patients per clinical staging data.[16] Patients with metastatic disease were additionally eliminated if either of the following were identified: (1) metastatic treatment protocols in New Drug Funding Program records or (2) advanced cancer diagnosis on inpatient admission records within 4 months of incident diagnosis. Finally, the cohort was limited to patients who received surgical treatment within 6 months of diagnosis per inpatient, ambulatory care, and health insurance billing records. Surgery was defined as modified radical mastectomy, partial mastectomy, or lumpectomy with the exclusion of needle biopsies or lymph node excision alone. Surgical treatment was limited to 6 months following diagnosis to allow for sufficient access to HER2 testing in patients who were diagnosed in late 2007. These exclusions isolated a patient population indicated for HER2 testing and potentially eligible for adjuvant trastuzumab treatment in Ontario (**Figure 1**). All follow-up was captured through administrative health data or medical records.

Variables, Sources, and Measurement

We collected variables measuring clinical, demographic, and healthcare system factors likely to influence HER2 test documentation or test usage. Administrative data were de-

■ **Figure 1.** Time Frame Definitions for Selection of the Early-Stage Breast Cancer Cohort and Follow-Up Data Collection



BC indicates breast cancer.

rived from several sources: (1) inpatient and outpatient procedures and diagnoses from Canadian Institute for Health Information Same Day Surgery,[17] National Ambulatory Care Reporting System,[17] and Discharge Abstract Databases[18]; (2) professional and procedure billing codes from the Ontario Health Insurance Program database; and (3) demographic information from the Institute for Clinical Evaluative Sciences Physician and Registered Persons databases.[19] Incident diagnosis, institution of diagnosis, staging, laterality, and vital status were obtained from the OCR. Pathology reports submitted to CCO for OCR purposes were reviewed for HER2 testing, estrogen or progesterone receptor status, tumor grade, and testing laboratory. Whenever possible, we derived exclusion variables from multiple data sources to reduce the impact of nonreporting biases associated with a single data source. For example, surgical procedures were derived from Same Day Surgery, National Ambulatory Care Reporting System, and Ontario Health Insurance Program databases.

Data on the primary outcome, HER2 test documentation, were collected from pathology reports. We considered a patient to have HER2 test documentation when there was evidence in the pathology report that a HER2 test was requested or conducted. Detailed information was collected to document the type of test provided and sequencing. Testing was documented as IHC or FISH when any prespecified keyword (Table 1) was found in the report. A HER2 test was recorded as unknown when evidence of a HER2 test was present but the type was not distinguishable. The date of each HER2 test was recorded.

Potential predictors of documentation included age at diagnosis, income, laboratory type, diagnosing physician specialty, and tissue source. Income was categorized into quintiles according to Statistics Canada methodology, which uses postal code–derived census data to estimate household size–adjusted family income.[20,21] The specialty of the treating physician was derived from chemotherapy billing records. Finally, laboratory type and tissue source predictors were drawn from pathology reports. We considered urbanicity, treatment setting, and comorbidity as potential confounders of HER2 test documentation. Urbanicity was defined using postal code–derived census data.[22] The institution of diagnosis assigned in the OCR was used to determine the local health integration network (LHIN) diagnosis setting. LHINs are health authorities responsible for providing, planning, integrating, and funding all public healthcare within a defined geographic region. Charlson Comorbidity Index scores were computed from inpatient diagnosis codes[23-25] to categorize patients as having no prior comorbidities or 1 or more prior comorbidities in the 3 years prior to incident diagnosis. Hormone receptor status and histologic tumor grade were determined from pathology reports. This 3-year look-back window was chosen to maximize capture of patients with prior comorbidities, particularly cardiac conditions, that might not be identified in a single year prior to diagnosis via inpatient codes. Our a priori clinical reasoning was that trastuzumab therapy would be more likely to be contraindicated in patients with comorbidities, which might introduce a systematic bias against HER2 testing in the cohort. Stage at diagnosis was captured from the OCR. Finally, we collected several other clinical indicators to describe the cohort. Breast-conserving (partial excision, partial mastectomy, tumor excision, with or without reconstruction) or nonconserving (modified radical mastectomy, modified radical excision, with or without reconstruction) surgical procedures were determined from billing or inpatient/outpatient procedure codes. Radiation treatment was captured from billing records. Tumor laterality and vital status were determined from the OCR.

Table 1. Keywords Used in Determination of IHC or FISH Test Documentation From Pathology Reports

IHC Keywords	FISH Keywords
% Positive cells	FISH
A0485 (antibody)	Fluorescence in situ hybridization
ABC IHC technique	HER2/CEP 17
CB11 (antibody)	HER2/neu gene amplification
HER2/neu protein overexpression	HER2/neu oncogene amplification
HER2/neu oncoprotein overexpression	PathVysion
HercepTest	
Immunohistochemical testing	
IHC	
LSAB	
Polymeric IHC technique	
SP3 (antibody)	
TAB250 (antibody)	

FISH indicates fluorescence in situ hybridization; HER2, human epidermal growth factor receptor-2; IHC, immunohistochemistry.

Collection of Tumor Pathology From Medical Records

All OCR-related pathology reports available for the cohort were reviewed for HER2 test, tumor grade, and hormone receptor data using a study-specific protocol adapted from the 2009 College of American Pathologists guidelines.[26] Pathology report extraction was completed as a patient-level analysis by prioritizing the most aggressive tumor in multifocal cases for consistency with clinical decision making. Synoptex software (Artificial Intelligence in Medicine[27]) was used to assist in data capture from pathology reports. The software was used in a semiautomated way, whereby a study-specific data entry form was initially populated using customized algorithms adapted for this study. The data capture was then evaluated and either accepted or corrected by trained reviewers. Pathology reports were reviewed on an individual basis, although patients with 3 or more documented HER2 tests were subject to a holistic review of all pathology reports to reduce duplicate reporting of HER2 tests. The pathology review incorporated reports with and without linkage to the patient record in the cancer registry in an effort to reduce any reporting or system-related factors that might bias HER2 documentation.

Statistical Analysis

A minimum of 160 patients were required to detect a 0.05% deviation in HER2 test documentation with 80% power. Descriptive statistics were used to review the distribution of all variables. The relationship between each variable and the primary outcome was first analyzed by univariate odds ratios (ORs) and χ^2 tests. Adjusted ORs were estimated by multiple logistic regression. Variables were selected for inclusion in logistic models based on the significance of univariate ORs at the $P = .10$ level. We prespecified the inclusion of urbanicity and income variables in each model to address sociodemographic questions of access.

Documentation of HER2 testing was modeled from 2 perspectives to account for variability due to reporting system or clinical factors. From the system perspective, we assessed whether HER2 test reporting to the OCR was associated with reporting of other pathology factors by regressing HER2 test documentation against the documentation of other pathology variables (eg, histologic tumor grade documented vs undocumented). From the clinical perspective, we examined whether HER2 documentation was related to clinical factors such as disease severity by regressing HER2 documentation against clinically defined categories for relevant variables (eg, histologic tumor grade 2 or 3 vs 1). An alpha level of .05 indicated statistical significance in multivariable analysis.

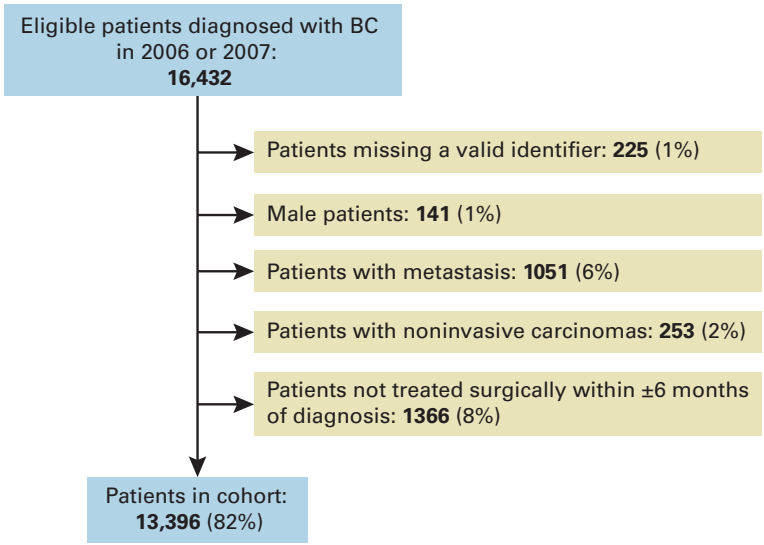
The significance of regional variation was estimated for LHINs as a group by using the global likelihood ratio test for each model. The robustness of model conclusions was tested in sensitivity analyses examining alternative modeling scenarios: (1) all clinically important and statistically significant predictors; (2) all clinically important and statistically significant predictors, excluding variables with >25% missingness; (3) only statistically significant predictors; (4) only clinically significant predictors. The preferred final model was the one that maximized the χ^2 likelihood ratio and pseudo R^2 values while also having clinical face validity. Final models were then used to examine prespecified interaction terms selected by systems or clinical reasoning. Finally, documentation of IHC and FISH tests was analyzed descriptively. All analyses were performed with STATA version 10 (StataCorp, College Station, Texas).

RESULTS

Cohort

A total of 16,432 patients diagnosed with incident BC in 2006 and 2007 were identified from the OCR. After application of exclusion criteria (**Figure 2**), the final analytic data set

Figure 2. Flow Diagram of Patient Exclusions From the Cohort and Pathology Reports Reviewed[a]



BC indicates breast cancer.
[a]Percentages shown were calculated on the basis of eligible patients diagnosed with breast cancer.

consisted of 13,396 patients. We reviewed 29,764 reports of breast tissue. At baseline, the cohort was similar to the Ontario BC population on the basis of age, stage at diagnosis, laterality, radiation usage, and vital status[28] as of March 2009 (Table 2). The vital status of all patients was known at the end of follow-up.

HER2 Documentation

HER2 test documentation was noted for 66% of the cohort (n = 8854). Age at diagnosis, LHIN, and documentation of other pathology variables were significant predictors of HER2 test documentation after adjusting for other predictors in the reporting system perspective (Table 3). Increasing age at diagnosis was associated with lower odds of having a documented HER2 test. Documentation of other tumor pathology factors was associated with higher odds of having a documented HER2 test; in the extreme case, patients with documented hormone receptor status had 12.5 times higher odds of having a documented HER2 test than those without hormone receptor status documentation. Practice varied significantly by regional LHIN (global likelihood ratio test P <.001; data not shown). The presence of comorbidities was a significant predictor in univariate analysis, but was not associated with HER2 test documentation after adjusting for other predictors. Sociodemographic factors were not associated with HER2 test documentation in the cohort. We excluded treating physician specialty and laboratory variables from models because of high proportions of missing values (47% and 84%, respectively). The direction and significance of all variables were robust to all sensitivity analysis (data not shown).

Clinical-perspective results (Table 4) were similar to reporting-system results. After accounting for regional LHIN variation (global likelihood ratio test P <.001; data not shown), sociodemographic factors and comorbidity were not associated with HER2 documentation. Older patients had lower odds of HER2 test documentation. Clinical indicators suggest that disease severity was associated with higher odds of HER2 test documentation. Patients diagnosed with clinical stage II and III disease had 1.45 and 1.83 times higher odds of HER2 documentation, respectively, than patients diagnosed with stage I disease. Similarly, patients diagnosed with histologic tumor grades 2 and 3 had 1.25 and 1.51 times higher odds of HER2 documentation, respectively, than patients diagnosed with grade 1. The direction and significance of all predictors remained consistent in sensitivity analyses (data not shown). From the clinical perspective, we also investigated interactions between urbanicity and income, age and stage, age and comorbidity, and stage and grade (capturing aggressive disease); none was significant (data not shown).

Documented HER2 Testing Patterns

The quality and consistency of HER2 test documentation varied widely, but test type was distinguishable for 96% and 95% of first and second tests, respectively. Among tested pa-

Table 2. Characteristics of Cohort Patients Diagnosed With Early-Stage Breast Cancer in 2006 or 2007[a]

Characteristic	Cohort, % (n)	HER2 Test Documentation, % (n)	
		Undocumented	Documented
All patients	100 (13,396)	34 (4542)	66 (8854)
Age at diagnosis, y[b]			
<50	23 (3034)	21 (961)	23 (2073)
50-59	25 (3376)	25 (1117)	26 (2259)
60-69	24 (3149)	23 (1063)	24 (2086)
70-79	18 (2449)	19 (859)	18 (1590)
≥80	10 (1388)	12 (542)	10 (846)
Diagnosis year[b]			
2006	49 (6622)	48 (2162)	50 (4460)
2007	51 (6774)	52 (2380)	50 (4394)
Vital status			
Alive at end of follow-up	95 (12,704)	95 (4319)	95 (8385)
Urbanicity[b]			
Urban	87 (11,706)	85 (3876)	88 (7830)
Rural	13 (1684)	15 (665)	12 (1019)
Stage at diagnosis[b]			
I	35 (4712)	38 (1710)	34 (3002)
II	32 (4319)	28 (1258)	35 (3061)
III	10 (1358)	8 (347)	11 (1011)
Unknown	22 (3007)	27 (1227)	20 (1780)
Hormone receptor status[b]			
Positive	50 (6660)	19 (866)	65 (5794)
Negative	12 (1599)	3 (157)	16 (1442)
Unknown	38 (5137)	77 (3519)	18 (1618)
Histologic tumor grade[b]			
1	13 (1668)	11 (486)	13 (1182)
2	27 (3594)	18 (804)	31 (2790)
3	19 (2601)	10 (468)	24 (2133)
Unknown	41 (5533)	61 (2784)	31 (2749)
Laterality[b]			
Left	44 (6137)	40 (1818)	49 (4319)
Right	47 (5948)	38 (1721)	48 (4227)
Bilateral	2 (302)	1 (60)	3 (242)
Unknown	8 (1009)	21 (943)	1 (66)
Breast-conserving surgery	60 (8048)	60 (2717)	60 (5331)
Radiation treatment[b]	67 (8966)	62 (2832)	69 (6134)
Income			
Income quintile 1 (lowest)	17 (2334)	17 (771)	18 (1563)
Income quintile 2	19 (2577)	20 (910)	19 (1667)
Income quintile 3	20 (2649)	19 (879)	20 (1770)
Income quintile 4	21 (2812)	21 (956)	21 (1856)
Income quintile 5 (highest)	22 (2991)	22 (1014)	22 (1977)
Comorbidity[b]			
Charlson Comorbidity Index score ≥1	7 (988)	8 (367)	7 (621)

HER indicates human epidermal growth factor receptor-2.
[a]Numbers may not add to 100% due to rounding. Follow-up to March 31, 2009.
[b]Statistically significant differences between documented and undocumented patients at the $P = .05$ level per Pearson's χ^2 test.

■ **Table 3.** Unadjusted and Adjusted Odds Ratios for HER2 Documentation in Centrally Held Pathology Reports of Patients Diagnosed With Early-Stage Breast Cancer in 2006 or 2007: Reporting-System Perspective[a]

Reporting-System Variable	HER2 Test Documentation	
	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Age at diagnosis, y		
50-69 vs <50	0.92 (0.84-1.01)	0.88 (0.78-0.99)[b]
≥70 vs <50	0.81 (0.73-0.89)[b]	0.74 (0.65-0.85)[b]
Income quintile		
Quintile 1, 2, 3, or 4 vs quintile 5	1.00 (0.92-1.09)	1.00 (0.90-1.12)
Urbanicity		
Rural vs urban	0.76 (0.68-0.84)[b]	0.90 (0.77-1.04)
Stage documentation		
Documented vs undocumented	1.47 (1.35-1.60)[b]	1.58 (1.42-1.77)[b]
Hormone receptor documentation		
Documented vs undocumented	15.38 (14.09-16.80)[b]	12.54 (11.36-13.85)[b]
Histologic tumor grade documentation		
Documented vs undocumented	3.52 (3.26-3.79)[b]	2.18 (1.98-2.41)[b]
Comorbidity		
Charlson Comorbidity Index score ≥1 vs 0	0.86 (0.75-0.98)[b]	1.02 (0.85-1.22)

CI indicates confidence interval; HER, human epidermal growth factor receptor-2; OR: odds ratio.
[a]All pathology variables were modeled as either documented or undocumented to assess reporting-system relationships. Adjusted odds ratios account for variability by local health integration network in addition to the variables shown. This model considered 13,363 observations (99.7% of cohort).
[b]Significant at the $P = .05$ level.

tients, 95% had 1 or more IHC tests and IHC tests accounted for 94% of all first tests. Conversely, 15% of tested patients had 1 or more FISH tests, and 2% received FISH as the first test. The majority of patients (73%) had a single test documented. Secondary tests were noted for 24% of tested patients, while the remainder received more than 2 tests (3%). The second test type was split almost evenly among IHC and FISH at 49% and 46%, respectively. A maximum of 6 tests were noted for a patient across multiple pathology reports, which could not be ruled as duplicate reporting.

DISCUSSION

This study addresses a knowledge gap about HER2 test documentation and testing patterns in the largest cohort of female early-stage BC patients reported in the literature to date. In this population-based cohort, we documented a HER2 test for 66% of patients from pathology reports held by the Ontario provincial cancer agency within the context of a universal access environment where HER2 testing was standard practice for all early-stage patients.

Test documentation was related to documentation of other pathology factors such as hormone receptor testing, tumor stage, and histologic grade. Similarly, clinical measures of tumor pathology indicated that patients with more aggressive disease (advanced stage and grade) were more likely to have HER2 test documentation. Higher odds for testing in stage II or III versus stage I disease were expected given that some stage I tumors were likely smaller than 1 cm and therefore not indicated for trastuzumab treatment or HER2 testing. We also found that hormone receptor–negative patients were less likely to have HER2 documentation. We would expect the opposite trend, with hormone receptor negative–patients having a higher rate of documentation, if the potential clinical need for trastuzumab were a determinant of HER2 documentation. These findings do not suggest that patients with less aggressive disease lacked access to HER2 testing, but rather suggest that completeness of centralized reporting or documentation might be related to disease severity. That is more likely to be a function of the registry reporting system than of the clinical need for trastuzumab, as HER2 reporting for trastuzumab reimbursement is provided independent of the registry. However, the potential for residual confounding in the cohort limits our ability to draw conclusions about the role of disease severity in HER2 documentation. In contrast, HER2 test documentation was not related to sociodemographic factors such as urbanicity

Table 4. Unadjusted and Adjusted Odds Ratios for HER2 Documentation in Centrally Held Pathology Reports of Patients Diagnosed With Early-Stage Breast Cancer in 2006 or 2007: Clinical Perspective[a]

	HER2 Test Documentation	
Clinical Variable	**Unadjusted OR (95% CI)**	**Adjusted OR (95% CI)**
Age at diagnosis, y		
50-69 vs <50	0.92 (0.84-1.01)	0.78 (0.66-0.93)[b]
≥70 vs <50	0.81 (0.73-0.89)[b]	0.60 (0.49-0.74)[b]
Income quintile		
Quintile 1, 2, 3, or 4 vs quintile 5	1.00 (0.92-1.09)	1.00 (0.85-1.17)
Urbanicity		
Rural vs urban	0.76 (0.68-0.84)[b]	0.92 (0.75-1.14)
Stage at diagnosis		
II vs I	1.39 (1.27-1.51)[b]	1.45 (1.25-1.69)[b]
III vs I	1.66 (1.45-1.90)[b]	1.83 (1.45-2.30)[b]
Hormone receptor status		
Negative vs positive	1.37 (1.15-1.64)[b]	1.44 (1.11-1.87)[b]
Histologic tumor grade		
2 vs 1	1.43 (1.25-1.63)[b]	1.25 (1.05-1.48)[b]
3 vs 1	1.87 (1.62-2.17)[b]	1.51 (1.24-1.86)[b]
Comorbidity		
Charlson Comorbidity Index score ≥1 vs 0	0.86 (0.75-0.98)[b]	1.19 (0.88-1.60)

CI indicates confidence interval; HER2, human epidermal growth factor receptor-2; OR, odds ratio.
[a]All pathology variables were modeled using clinical categorizations to address relationships between disease pathology and HER2 documentation. Adjusted odds ratios account for variability by local health integration network in addition to the variables shown. This model considered 6142 observations (45.8% of cohort).
[b]Significant at the $P = .05$ level.

or income, or to comorbidity, consistent with the policies of the Ministry of Health and Long-term Care. These observations were robust across different modeling perspectives and sensitivity analyses.

Our multivariable analysis results both support and challenge the findings of other studies. A report from Nova Scotia, Canada, found that untested patients tended to be older and have smaller tumors.[14] Similarly, a study in Swansea, Southwest Wales, also suggested that elderly patients were less likely to have HER2 testing.[29] These age-related differences in HER2 testing and documentation are consistent with the results of the current study. A study of Kaiser Permanente Northwest patients found that HER2 testing was more common among those also tested for estrogen receptor status.[30] That is consistent with documented practice in Canada, where HER2 and hormone receptor testing by IHC are typically conducted simultaneously in centralized laboratories.[31] However, HER2 testing in the Kaiser Permanente population was lower in patients with Medicare or Medicaid insurance.[30] This situation differs from that in Ontario, where HER2 testing is universally available, regardless of enrollment in either provincial or private insurance plans. Indeed, we demonstrated that sociodemographic factors did not play a role in test documentation. A recent analysis of early-stage BC in Aetna patients demonstrated a 97% rate of HER2 testing, and showed that documentation was not related to age, tumor pathology, or sociodemographic factors.[32] The contrasting rates of HER2 documentation between the Aetna study and this analysis underscore the regional reporting variation across Ontario LHINs. This variation provides an important caveat for researchers studying personalized medicine practice using a centralized, cancer registry–driven data source and provides support for the ongoing effort to adopt collaborative staging standards in Ontario.[33]

Incomplete HER2 test documentation in centrally held reports makes interpretation of these findings challenging, which in turn has implications for efforts to improve the quality of personalized medicine through research and monitoring. Recent calls for better evidence to evaluate personalized medicine practice have highlighted 2 challenges: (1) pricing-based reimbursement for diagnostic tests does not facilitate individual test identification in administrative data and (2) registries documenting testing and subsequent treatment decisions are lacking.[34]

We assessed the ability to document HER2 testing in Ontario across treatment settings and geographical locations, but found that HER2 status was not detailed in the OCR, and HER2 tests were not discernible from other diagnostics in billing data (despite being billed individually vs as a bundle). CCO's repository of OCR pathology submissions provided a central source for pathology reports, but we identified several deficiencies in this source. Although submission of pathology reports to CCO to inform cancer diagnoses and staging in the OCR is mandatory, HER2 testing was not required for reporting at the time of this study. Moreover, HER2 tests tend to be reported on addendum pathology reports, which may not be submitted centrally unless the addendum alters diagnosis or staging for registry purposes. Differential use of electronic systems to manage and submit pathology reports at the laboratory level may also account for variability. These system factors may explain the high missingness of HER2 and other pathology factors in the cohort. Pathology reports as a source of HER2 information are also highly variable in format, detail, and structure.

This study provides a baseline measure of HER2 test documentation using multivariable methods in the largest early-stage BC cohort reported in the literature. We have demonstrated the potential for biased reporting of more aggressive disease, which may be related to the capabilities of individual laboratory information systems or incentives to gain access to drug reimbursement. Information about testing practice can be used to independently assess testing guideline adherence, identify variations in testing practice across laboratories, and improve the consistency and quality of HER2 testing. However, improvements to the reporting system, or mandatory reporting of HER2 test results, are necessary to produce high-quality data for provincial assessment. Program evaluators and health services researchers need to be aware of potential ascertainment bias before using this information for quality improvement or policy development purposes.

Acknowledgments

Pathology report review and extraction: Jenita Chrysostuoum, Nicole Findlay, Janna Fung, Sumeha Khurana, Andrea Poletto, Kennedy Reimersma, Raphael Sabio, Sama Thomas.

Institute for Clinical Evaluative Sciences data analysis support: Paul Cascagnette, BSc; Refik Saskin, MSc.

Expert consultation: Artificial Intelligence in Medicine, Peter Brueckner, MD, MBA; George Cernile, PMP; Nelson Chong, BSc; Thi Ho, MSc; Sharon Nofech-Mozes, MD; Carol Sawka, MD, FRCPC.

Author Affiliations: From Centre for Evaluation of Medicines (ILF, DAM), St. Joseph's Healthcare Hamilton, Hamilton, Ontario, Canada; Department of Clinical Epidemiology and Biostatistics (ILF, EP, DAM), McMaster University, Hamilton, Ontario, Canada; Ontario Institute for Cancer Research (CCE), Toronto, Ontario, Canada; Sunnybrook Health Sciences Centre (MT), Toronto, Ontario, Canada; Department of Medicine, University of Toronto (MT, NBL), Toronto, Ontario, Canada; Princess Margaret Hospital, University Health Network (NBL), Toronto, Ontario, Canada; Biostatistics Unit, St Joseph's Healthcare Hamilton (EP), Hamilton, Ontario, Canada; Alberta Bone and Joint Health Institute (HK), Calgary, Alberta, Canada; Pharmacoeconomics Research Unit, Cancer Care Ontario (JSH), Toronto, Ontario, Canada; Institute of Health Policy, Management and Evaluation, University of Toronto (JSH), Toronto, Ontario, Canada; Department of Community Health Sciences, University of Calgary (DAM), Calgary, Alberta, Canada.

Funding Source: This study was supported through provision of data by the Institute for Clinical Evaluative Sciences (ICES) and Cancer Care Ontario (CCO) and through funding support to ICES from an annual grant by the Ministry of Health and Long-Term Care and the Ontario Institute for Cancer Research (OICR). The opinions, results and conclusions reported in this paper are those of the authors. No endorsement by ICES, CCO, OICR, or the government of Ontario is intended or should be inferred. Support was also provided by a grant from the National Cancer Institute (P01CA130818).

Author Disclosures: Dr Trudeau reports paid consultancies from Roche for advisory boards. The other authors (ILF, CCE, NBL, EP, HK, JSH, DAM) report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

Authorship Information: Concept and design (ILF, CCE, NBL, DAM); acquisition of data (ILF, CCE, NBL, HK, DAM); analysis and interpretation of data (ILF, CCE, MT, NBL, EP, HK, JSH, DAM); drafting of the manuscript (ILF, HK, JSH, DAM); critical revision of the manuscript for important intellectual content (ILF, CCE, MT, NBL, EP, JSH, DAM); statistical analysis (ILF, EP, HK, JSH, DAM); provision of study materials or patients (ILF); obtaining funding (ILF, CCE, NBL, DAM); administrative, technical, or logistic support (ILF, DAM); and supervision (DAM).

Address correspondence to: Deborah A. Marshall, PhD, Faculty of Medicine, University of Calgary, Rm 3C56, Health Research Innovation Centre, 3280 Hospital Drive NW, Calgary, AB T2N 4Z6. E-mail: damarsha@ucalgary.ca.

REFERENCES

1. Canadian Cancer Society's Steering Committee on Cancer Statistics. *Canadian Cancer Statistics 2011.* Toronto, Canada: Canadian Cancer Society; 2011.

2. Bray F, McCarron P, Parkin DM. The changing global patterns of female breast cancer incidence and mortality. *Breast Cancer Res.* 2004;6(6):229-239.

3. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science.* 1987;235(4785):177-182.

4. Slamon DJ, Godolphin W, Jones LA, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science.* 1989; 244(4905):707-712.

5. Dahabreh IJ, Linardou H, Siannis F, Fountzilas G, Murray S. Trastuzumab in the adjuvant treatment of early-stage breast cancer: a systematic review and meta-analysis of randomized controlled trials. *Oncologist.* 2008;13(6):620-630.

6. Perez EA, Romond EH, Suman VJ, et al. Four-year follow-up of trastuzumab plus adjuvant chemotherapy for operable human epidermal growth factor receptor 2-positive breast cancer: joint analysis of data from NCCTG N9831 and NSABP B-31. *J Clin Oncol.* 2011;29(25): 3366-3373.

7. Hanna W, O'Malley FP, Barnes P, et al. Updated recommendations from the Canadian National Consensus Meeting on HER2/neu testing in breast cancer. *Curr Oncol.* 2007;14(4):149-153.

8. Wolff AC, Hammond ME, Schwartz JN, et al; American Society of Clinical Oncology; College of American Pathologists. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol.* 2007;25(1):118-145.

9. Trudeau M, Madarnas Y, McCready D, Pritchard KI, Messersmith H; Breast Cancer Disease Site Group. *The Role of Trastuzumab in Adjuvant and Neoadjuvant Therapy in Women With HER2/Neu-Overexpressing Breast Cancer: A Clinical Practice Guideline.* Toronto: Cancer Care Ontario; May 12, 2006. Report 1-24.

10. Lidgren M, Jönsson B, Rehnberg C, Willking N, Bergh J. Cost-effectiveness of HER2 testing and 1-year adjuvant trastuzumab therapy for early breast cancer. *Ann Oncol.* 2008;19(3):487-495.

11. Elkin EB, Weinstein MC, Winer EP, Kuntz KM, Schnitt SJ, Weeks JC. HER-2 testing and trastuzumab therapy for metastatic breast cancer: a cost-effectiveness analysis. *J Clin Oncol.* 2004;22(5):854-863.

12. Dendukuri N, Khetani K, McIsaac M, Brophy J. Testing for HER2-positive breast cancer: a systematic review and cost-effectiveness analysis. *CMAJ.* 2007;176(10):1429-1434.

13. Phillips KA, Liang SY, Van Bebber S; Canpers Research Group. Challenges to the translation of genomic information into clinical practice and health policy: utilization, preferences and economic value. *Curr Opin Mol Ther.* 2008;10(3):260-266.

14. Snow S, Rayson D, Barnes P, Kara T, Younis T. Does HER2/neu status affect adjuvant chemotherapy utilization? a population-based analysis. Poster presented at: ASCO 2008 Breast Cancer Symposium; September 5-8, 2008; Washington, DC.

15. Spencer D, Strasbourg D. *McGuinty Government Improving Access to Cancer-Fighting Drugs: Herceptin, Navelbine, Taxotere to Be Funded by Province. Toronto, Canada: Ontario Ministry of Health and Long-Term Care.* http://news.ontario.ca/archive/en/2005/07/22/McGuinty-government-improving-access-to-cancerfighting-drugs.html. Published 2005. Accessed December 5, 2012.

16. American Joint Committee on Cancer. *AJCC Cancer Staging Manual.* 6th ed. http://www.cancerstaging.org/products/pasteditions.html. 2002. Accessed December 4, 2012.

17. Gibson D, Richards H, Chapman A. The national ambulatory care reporting system: factors that affect the quality of its emergency data. *Int J Inf Qual.* 2008;2(2):97-114.

18. Richards J, Brown A, Homan C. The data quality study of the Canadian Discharge Abstract database. Statistics Canada International Symposium Series: Proceedings. http://www5.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=11-522-X20010016282&lang=eng. Published 2001. Accessed December 5, 2012.

19. Iron K, Zagorski B, Sykora K, Manuel D. *Living and Dying in Ontario: An Opportunity for Improved Health Information.* Toronto, Canada: Institute for Clinical Evaluative Science. http://www.ices.on.ca/webpage.cfm?site_id=1&org_id=31&morg_id=0&gsec_id=0&item_id=4720. Published March 2008. Accessed December 5, 2012.

20. Murphy B, Zhang X, Dionne C. *Revising Statistics Canada's Low Income Measure (LIM).* Income Research Paper Series. Ottawa: Statistics Canada. Pages 1-31. http://www.statcan.gc.ca/pub/75f0002m/75f0002m2010004-eng.htm. Published 2010. Accessed December 5, 2012.

21. Wolfson MC, Evans JM; Organisation for Economic Co-operation and Development. *Statistics Canada's Low Income Cut-Offs Methodological Concerns and Possibilities.* Ottawa: Statistics Canada, Analytical Studies Branch; 1989.

22. Statistics Canada. *Postal Code Conversion File (PCCF), Reference Guide.* 92-153-G ed. Ottawa, Canada: Statistics Canada. http://www5.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=92-153-G&lang=eng. Published May 2011. Accessed December 5, 2012.

23. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care.* 2005;43(11):1130-1139.

24. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *J Clin Epidemiol.* 2004;57(12):1288-1294.

25. Quail JM, Lix LM, Osman BA, Teare GF. Comparing comorbidity measures for predicting mortality and hospitalization in three population-based cohorts. *BMC Health Serv Res.* 2011;11:146.

26. Lester SC, Bose S, Chen YY, et al; Members of the Cancer Committee, College of American Pathologists. Protocol for the examination of specimens from patients with invasive carcinoma of the breast. *Arch Pathol Lab Med.* 2009;133(10):1515-1538.

27. Synoptex NLP [software]. Artificial Intelligence in Medicine. http://www.aim.on.ca/system-integration/. 2011. Accessed October 16, 2011.

28. Quan ML, Hodgson N, Schultz SE, et al. Surgery for breast cancer. In: Urbach DR, Simunovic M, Schultz SE, eds. *Cancer Surgery in Ontario: ICES Atlas.* Toronto, Canada: Institute for Clinical Evaluative Sciences; 2008.

29. Webster R, Palaniappan N, Abraham J, Bertelli G, Jasani B, Barrett-Lee P. HER2 testing, adjuvant trastuzumab use and results: our experience in South Wales. *Clin Oncol (R Coll Radiol).* 2010;22(10):894.

30. Goddard KA, Weinmann S, Richert-Boe K, Chen C, Bulkley J, Wax C. HER2 evaluation and its impact on breast cancer treatment decisions. *Public Health Genomics.* 2012;15(1):1-10.

31. O'Malley FP, Thomson T, Julian J, et al. HER2 testing in a population-based study of patients with metastatic breast cancer treated with trastuzumab. *Arch Pathol Lab Med* 2008;132(1):61-65.

32. Haas JS, Phillips KA, Liang SY, et al. Genomic testing and therapies for breast cancer in clinical practice. *J Oncol Pract.* 2011;7(3)(suppl):e1s-e7s.

33. Cancer Care Ontario. Collaborative staging. *Surgical Oncology Newsletter.* 2009;4(1). https://www.cancercare.on.ca/SurgicalOncologyNews/766_776.htm. Accessed December 5, 2012.

34. Peacock S. Economics, personalized health care and cancer control. Presented at: 2011 CAPT Annual Conference; April 17, 2011; Ottawa, Ontario, Canada. ■

3.3 PAPER 3 APPENDICES

APPENDIX 3.3.1:UNIVARIATE ODDS RATIOS, ALL VARIABLES

HER2 Documentation Yes vs. No	Odds Ratio	Standard Error	z	P>z	95% Confidence Interval		Observations	% of Cohort
Age (years)								
50-69 vs. <50	0.92	0.04	-1.68	0.093	0.84	1.01	13,396	100%
≥ 70 vs. <50	0.81	0.04	-4.19	<0.001	0.73	0.89		
Breast conserving surgery								
Y vs. N	1.02	0.04	0.44	0.662	0.94	1.09	13,395	100%
Charlson comorbidity score								
≥1 vs. 0	0.86	0.06	-2.23	0.026	0.75	0.98	13,396	100%
Histologic tumour grade								
2 vs. 1	1.43	0.10	5.30	<0.001	1.25	1.63	7,836	58%
3 vs. 1	1.87	0.14	8.46	<0.001	1.62	2.17		
Histologic grade documentation								
Y vs. N	3.52	0.13	32.96	<0.001	3.26	3.79	13,396	100%
Hormone receptor documentation								
Y vs. N	15.38	0.69	60.85	<0.001	14.09	16.80	13,396	100%
Hormone receptor status								
Positive vs. negative	0.73	0.07	-3.46	<0.001	0.61	0.87	8,259	62%
Income quintile								
(1 or 2 or 3 or 4) vs. 5	1.00	0.04	0.00	0.998	0.92	1.09	13,363	100%
Physician specialty								
GP/FP vs. medical oncologist	0.91	0.08	-1.11	0.268	0.77	1.07	7,115	53%
Other vs. medical oncologist	0.71	0.07	-3.34	0.001	0.58	0.87		
Urbanicity								
Rural vs. urban	0.76	0.04	-5.16	<0.001	0.68	0.84	13,390	100%
Clinical stage documentation								
Y vs. N	1.47	0.06	9.05	<0.001	1.35	1.60	13,396	100%
Clinical stage	1.39	0.06	7.23	<0.001	1.27	1.51	10,389	78%

HER2 Documentation Yes vs. No	Odds Ratio	Standard Error	z	P>z	95% Confidence Interval		Observations	% of Cohort
Stage II vs. I								
Stage III vs. I	1.66	0.11	7.32	<0.001	1.45	1.90		
Tissue sample type								
Tumor block vs. biopsy	2.93	0.37	8.43	<0.001	2.28	3.76	7,172	54%

Abbreviations: GP: general practitioner, FP: family practitioner

GLOBAL LIKELIHOOD TEST FOR LHIN VARIABLE

The LHIN variable is comprised of a series of 14 nominal categories representing integrated health networks in Ontario. Healthcare is managed at the level of the network (health regions) within the province. We used this variable to account for variations that may occur across practice settings. We believed that practice would likely differ between regional cancer centres and community treatment settings, or between academic and community treatment settings, however we could not determine these variables from the administrative data or pathology reports available. We attempted to determine whether the patient was treated at a regional cancer centre by examining the institution codes associated with chemotherapy billing codes in claims data, but this procedure assigned a number of institutions that could not be categorised as regional cancer centres or otherwise. Thus, the LHIN variable was the only alternative measure of healthcare setting available from the administrative health data, complete for the cohort and validated by the Institute for Clinical Evaluative Sciences (ICES). Given that previous studies of Ontario clinical practice have noted differences in practice across LHINS, we felt it was necessary to adjust for this variability to better estimate the effects of other variables.

Usual coding of the LHIN nominal variable with a reference category was difficult to conceptualise as no single LHIN seemed to be an appropriate referent. Moreover, interpretation of the the effects of one LHIN relative to an arbitrarily

chosen referent LHIN would be challening without first knowing how the referent compared against the guidelines. It was not our goal to determine specifically which LHINs behaved more or less closely to guidelines, and we therefore did not need to precisely estimate the nature of the relationship between LHIN and HER2 documentation. Therefore, we chose to account for the variability in practice setting by evaluating LHIN with the global likelihood test. In univariate analyses, the likelihood ratio was significant, indicating that HER2 documentation did vary significantly by LHIN. We then chose to adjust for the grouped LHIN variable in our final models (Appendix 2, 3)

Block 1 = LHIN

Block	LL	LR	df	Pr > LR	AIC	BIC
1	-8465.414	220.67	1	0.0000	16934.83	16949.83

APPENDIX 3.3.2: SYSTEM PERSPECTIVE SENSITIVITY ANALYSIS MODELS

Documentation of HER2 testing was modeled from two perspectives to account for variability due to reporting system or clinical factors. In the system perspective, we assessed whether HER2 test reporting to the OCR was associated with reporting of other pathology factors. Clinical pathology variables were recoded as either documented or undocumented per pathology records (e.g. tumour stage documented vs. undocumented) and regressed against HER2 test documentation in the system perspective.

Sensitivity analyses (shown here) were conducted to test the robustness of model conclusions under alternative modeling scenarios:

(1) all clinically (or system) important and statistically significant predictors in the model,

(2) all clinically important and statistically significant predictors in the model, excluding variables with >25% missingness

(3) only statistically significant predictors in the model, and

(4) only clinically (or system) important predictors.

This approach sought to determine whether the exclusion of certain variables had an impact on the interpretation of other variables or the outcome. For example, results from modeling scenario (2) were compared to scenario (1) to understand whether the exclusion of variables with high missingness altered the nature of the

relationships between the dependent variable (HER2 documentation) and the independent variables included in both models.

The preferred model from scenarios (1) through (4) was selected as the one that maximised the Chi-square likelihood ratio and the pseudo R^2 value while maintaining face validity across the independent and dependent variable relationships. The preferred model was then used to examine pre-specified interaction terms selected by systems reasoning. In the systems perspective we did not identify potential interactions. The final model was adjusted for the effect of LHIN as a grouped variable using the global likelihood test.

Model A: all clinically important and statistically significant variables, no interactions

Selected for final model; global test for LHIN variable results shown (individual LHIN results confidential)

Logistic regression — Number of observations = 13363
LR chi^2 = 5670.18 — Prob > chi^2 = 0.0000
Log likelihood = -5722.1554 — Pseudo R^2 = 0.3313

HER2 Documentation Yes vs. No	Odds Ratio	Standard Error	z	P>z	95% Confidence Interval	
Urbanicity						
Rural vs. urban	0.90	0.07	-1.40	0.163	0.77	1.04
Age (years)						
50-69 vs. <50	0.88	0.05	-2.12	0.034	0.78	0.99
≥ 70 vs. <50	0.74	0.05	-4.36	<0.001	0.65	0.85
Clinical stage documentation						
Y vs. N	1.58	0.09	8.10	<0.001	1.42	1.77
Income quintile						
(1 or 2 or 3 or 4) vs. 5	1.00	0.06	0.06	0.951	0.90	1.12
Histologic grade documentation						
Y vs. N	2.18	0.11	15.85	<0.001	1.98	2.41
Hormone receptor documentation						
Y vs. N	12.54	0.63	50.14	<0.001	11.36	13.85
Charlson comorbidity score						
≥1 vs. 0	1.02	0.09	0.21	0.837	0.85	1.22

Global likelihood test for LHIN group variable

Block 1: Urbanicity
Block 2: Age
Block 3: Clinical stage documentation
Block 4: Income quintile
Block 5: Histologic grade documentation
Block 6: Hormone receptor documentation
Block 7: Charlson comorbidity score
Block 8: LHIN**

Block	LL	LR	df	Pr > LR	AIC	BIC
1	-8544.667	25.12	1	0.0000	17093.33	17108.34
2	-8535.465	18.40	1	0.0000	17076.93	17099.43
3	-8497.38	76.17	1	0.0000	17002.76	17032.76
4	-8497.294	0.17	1	0.6784	17004.59	17042.09
5	-7935.033	1124.52	1	0.0000	15882.07	15927.07
6	-6103.533	3663.00	1	0.0000	12221.07	12273.57
7	-6103.453	0.16	1	0.6884	12222.91	12282.91
8	-6069.733	67.44	1	0.0000	12157.47	12224.97

Model B: all clinically important and statistically significant variables, no interactions, excluding variables missing >25%

This model examined whether the exclusion of variables with high missingness impacted the interpretation of other variables, such as the significance and directionality of the relationship between other variables and the outcome.

Logistic regression — Number of obs = 13363
LR chi^2 = 119.89 — Prob > chi^2 = 0.0000
Log likelihood = -8497.2829 — Pseudo R^2 = 0.0070

HER2 Documentation Yes vs. No	Odds Ratio	Standard Error	z	P>z	95% Confidence Interval	
Urbanicity						
Rural vs. urban	0.74	0.04	-5.46	<0.001	0.67	0.83
Age (years)						
50-69 vs. <50	0.93	0.04	-1.59	0.112	0.85	1.02
≥ 70 vs. <50	0.85	0.04	-3.09	0.002	0.77	0.94
Clinical stage documentation						
Y vs. N	1.46	0.06	8.74	<0.001	1.34	1.59
Income quintile						
(1 or 2 or 3 or 4) vs. 5	1.02	0.04	0.42	0.676	0.93	1.11

Note: LHIN variable not shown due to confidentiality

Model C: only statistically significant univariate variables, no interactions

This model examined whether the exclusion of certain non-significant univariate variables, included originally to answer clinical questions, would impact the interpretation of other variables, such as the significance and directionality of the relationship between other variables and the outcome.

Logistic regression	Number of obs = 13390
LR chi^2 = 4924.79	Prob > chi^2 = 0.0000
Log likelihood = -6113.3552	Pseudo R^2 = 0.2871

HER2 Documentation Yes vs. No	Odds Ratio	Standard Error	z	P>z	95% Confidence Interval	
Urbanicity						
Rural vs. urban	0.74	0.05	-4.36	<0.001	0.65	0.85
Age (years)						
50-69 vs. <50	0.89	0.05	-1.92	0.054	0.80	1.00
≥ 70 vs. <50	0.77	0.05	-4.10	<0.001	0.68	0.87
Clinical stage documentation						
Y vs. N	1.50	0.08	7.42	<0.001	1.34	1.66
Histologic grade documentation						
Y vs. N	2.10	0.10	16.18	<0.001	1.92	2.30
Hormone receptor documentation						
Y vs. N	13.14	0.61	55.78	<0.001	12.00	14.38

Note: LHIN variable not shown due to confidentiality

Note: no interaction terms investigated in the system perspective.

Model D: only clinically important variables

This model is the same as Model A.

APPENDIX 3.3.3: CLINICAL PERSPECTIVE SENSITIVITY ANALYSIS MODELS

In the alternative clinical perspective we examined whether HER2 documentation was related to clinical measures of disease severity and comorbidity. This was operationalized within clinical perspective models by regressing HER2 documentation against clinically-defined categories for relevant variables (e.g. tumour stage I or II vs. III).

Sensitivity analyses (shown here) were conducted to test the robustness of model conclusions under alternative modeling scenarios:

(1) all clinically important and statistically significant predictors in the model,

(2) all clinically important and statistically significant predictors in the model, excluding variables with >25% missingness

(3) only statistically significant predictors in the model, and

(4) only clinically important predictors.

This approach sought to determine whether the exclusion of certain variables had an impact on the interpretation of other variables or the outcome. For example, results from modeling scenario (2) were compared to scenario (1) to understand whether the exclusion of variables with high missingness altered the nature of the relationships between the dependent variable (HER2 documentation) and the independent variables included in both models.

The preferred model from scenarios (1) through (4) was selected as the one that maximised the Chi-square likelihood ratio and the pseudo R^2 value while maintaining face validity across the independent and dependent variable relationships. The preferred model was then used to examine pre-specified interaction terms selected by clinical reasoning. In the clinical perspective we pre-specified the investigation of the following interactions:

(1) urbanicity and income,

(2) age at diagnosis and stage at diagnosis,

(3) age at diagnosis and comorbidity score,

(4) stage at diagnosis and tumour grade.

The final model was adjusted for the effect of LHIN as a grouped variable using the global likelihood test.

Model E: all clinically important and statistically significant variables, no interactions

Selected for final model; global test for LHIN variable results shown (individual LHIN results confidential)

Logistic regression — Number of obs = 6142
LR chi^2 = 626.91 — Prob > chi^2 = 0.0000
Log likelihood = -2854.0698 — Pseudo R^2 = 0.0990

HER2 Documentation Yes vs. No	Odds Ratio	Standard Error	z	P>z	95% Confidence Interval	
Urbanicity						
Rural vs. urban	0.93	0.09	-0.78	0.436	0.77	1.12
Age (years)						
50-69 vs. <50	0.85	0.07	-1.84	0.066	0.72	1.01
≥ 70 vs. <50	0.69	0.07	-3.78	<0.001	0.57	0.84
Clinical stage						
Stage II vs. I	1.44	0.11	4.99	<0.001	1.25	1.66
Stage III vs. I	1.65	0.19	4.43	<0.001	1.32	2.06
Income quintile						
(1 or 2 or 3 or 4) vs. 5	0.98	0.08	-0.32	0.748	0.84	1.14
Histologic tumour grade						
2 vs. 1	1.27	0.11	2.86	0.004	1.08	1.49
3 vs. 1	1.52	0.15	4.19	<0.001	1.25	1.84
Hormone receptor status						
negative vs. positive	1.41	0.18	2.65	0.008	1.09	1.82
unknown vs. positive	0.23	0.02	-21.08	<0.001	0.20	0.26
Charlson comorbidity score						
≥1 vs. 0	1.18	0.17	1.13	0.257	0.89	1.57

Block 1: Urbanicity
Block 2: Age
Block 3: Clinical stage
Block 4: Income quintile
Block 5: Histologic grade
Block 6: Hormone receptor status
Block 7: Charlson comorbidity score
Block 8: LHIN**

Block	LL	LR	df	Pr > LR	AIC	BIC
1	-3166.824	1.41	1	0.2356	6337.647	6351.093
2	-3157.054	19.54	1	0.0000	6320.108	6340.277
3	-3134.164	45.78	1	0.0000	6276.328	6303.22
4	-3134.07	0.19	1	0.6648	6278.14	6311.755
5	-3118.31	31.52	1	0.0000	6248.621	6288.958
6	-2902.446	431.73	1	0.0000	5818.891	5865.951
7	-2901.852	1.19	1	0.2759	5819.704	5873.487
8	-2879.844	44.02	1	0.0000	5777.689	5838.195

Model F: all clinically important and statistically significant variables, no interactions excluding variables missing >25%

This model examined whether the exclusion of variables with high missingness impacted the interpretation of other variables, such as the significance and directionality of the relationship between other variables and the outcome.

Logistic regression	Number of obs = 10362
LR chi^2 = 1665.75	Prob > chi^2 = 0.0000
Log likelihood = -5655.3069	Pseudo R^2 = 0.1284

HER2 Documentation Yes vs. No	Odds Ratio	Standard Error	z	P>z	95% Confidence Interval	
Urbanicity						
Rural vs. urban	0.97	0.07	-0.40	0.687	0.84	1.12
Age (years)						
50-69 vs. <50	0.89	0.05	-1.98	0.048	0.80	1.00
≥ 70 vs. <50	0.77	0.05	-3.90	<0.001	0.67	0.88
Clinical stage						
Stage II vs. I	1.39	0.07	6.66	<0.001	1.26	1.53
Stage III vs. I	1.68	0.13	6.89	<0.001	1.45	1.95
Income quintile						
(1 or 2 or 3 or 4) vs. 5	0.98	0.05	-0.39	0.698	0.88	1.09
Charlson comorbidity score						
≥1 vs. 0	0.99	0.10	-0.11	0.912	0.82	1.20

Note: LHIN variable not shown due to confidentiality

Model G: only statistically significant univariate clinical variables, no interactions

This model examined whether the exclusion of certain non-significant univariate variables, included originally to answer clinical questions, would impact the interpretation of other variables, such as the significance and directionality of the relationship between other variables and the outcome.

Logistic regression — Number of obs = 6153
LR chi^2 = 1035.56 — Prob > chi^2 = 0.0000
Log likelihood = -2654.9941 — Pseudo R^2 = 0.1632

HER2 Documentation Yes vs. No	Odds Ratio	Standard Error	z	P>z	95% Confidence Interval	
Urbanicity						
Rural vs. urban	0.92	0.10	-0.74	0.458	0.75	1.14
Age (years)						
50-69 vs. <50	0.78	0.07	-2.75	0.006	0.66	0.93
≥ 70 vs. <50	0.60	0.06	-4.96	<0.001	0.49	0.74
Clinical stage						
Stage II vs. I	1.46	0.11	4.96	<0.001	1.26	1.69
Stage III vs. I	1.83	0.22	5.12	<0.001	1.45	2.30
Histologic tumour grade						
2 vs. 1	1.26	0.11	2.63	0.009	1.06	1.49
3 vs. 1	1.52	0.16	4.05	<0.001	1.24	1.86
Hormone receptor status						
negative vs. positive	1.44	0.19	2.73	0.006	1.11	1.87
unknown vs. positive	0.21	0.02	-20.48	<0.001	0.18	0.24
Charlson comorbidity score						
≥1 vs. 0	1.19	0.18	1.14	0.253	0.88	1.60

Note: LHIN variable not shown due to confidentiality

Model H: Preferred clinical model (E), with interaction terms

The preferred model was selected as the one that maximised the Chi-square likelihood ratio and the pseudo R^2 value while maintaining face validity across the independent and dependent variable relationships.

Logistic regression — Number of obs = 6142
LR chi^2 = 1039.20 — Prob > chi^2 = 0.0000
Log likelihood = -2647.9253 — Pseudo R^2 = 0.1640

HER2 Documentation Yes vs. No	Odds Ratio	Standard Error	z	P>z	95% Confidence Interval	
Urbanicity						
Rural vs. urban	1.25	0.29	0.96	0.336	0.79	1.96
Age (years)						
50-69 vs. <50	0.87	0.12	-1.03	0.304	0.67	1.13
≥ 70 vs. <50	0.68	0.10	-2.64	0.008	0.50	0.90
Clinical stage						
Stage II vs. I	1.98	0.44	3.06	0.002	1.28	3.08
Stage III vs. I	1.71	0.67	1.35	0.175	0.79	3.69
Income quintile						
(1 or 2 or 3 or 4) vs. 5	1.05	0.09	0.54	0.590	0.88	1.25
Histologic tumour grade						
2 vs. 1	1.28	0.14	2.25	0.024	1.03	1.59
3 vs. 1	1.61	0.23	3.29	0.001	1.21	2.13
Hormone receptor status						
negative vs. positive	1.43	0.19	2.69	0.007	1.10	1.86
unknown vs. positive	0.21	0.02	-20.47	<0.001	0.18	0.24
Charlson comorbidity score						
≥1 vs. 0	3.61	3.83	1.21	0.227	0.45	28.96
Interactions						
Rural*Income quintile (1+2+3+4)	0.69	0.18	-1.46	0.145	0.42	1.14
Age 50-69 * stage II	0.78	0.15	-1.31	0.191	0.53	1.13
Age 50-69 * stage III	1.14	0.31	0.46	0.647	0.66	1.95
Age ≥70 * stage II	0.74	0.16	-1.41	0.159	0.48	1.13
Age ≥70 * stage III	1.04	0.33	0.11	0.909	0.56	1.92
Age 50-69 * Charlson ≥1	0.27	0.29	-1.22	0.222	0.03	2.23
Age ≥70 * Charlson ≥1	0.37	0.40	-0.91	0.362	0.04	3.12
Stage II * grade 2	0.89	0.17	-0.59	0.554	0.61	1.30
Stage II * grade 3	0.85	0.19	-0.76	0.446	0.55	1.30

HER2 Documentation Yes vs. No	Odds Ratio	Standard Error	z	P>z	95% Confidence Interval	
Stage III * grade 2	1.02	0.39	0.04	0.965	0.48	2.15
Stage III * grade 3	0.98	0.38	-0.06	0.953	0.45	2.11

Note: LHIN variable not shown due to confidentiality

3.4 SUPPLEMENTAL ANALYSES

These analyses were undertaken to describe the HER2 testing landscape in Ontario and to inform the decision analysis of testing strategies to direct targeted adjuvant trastuzumab treatment in Ontario (Chapter 4). Specifically, the supplemental materials answer research questions:

(2) What type of HER2 test (e.g. IHC, FISH) is used most commonly for HER2 testing?

(3) When is confirmatory testing for HER2 used?

   a. What type of test is used for confirmatory testing?

(4) What proportion of HER2 positive patients receive trastuzumab treatment?

3.4.1 METHODS

Extraction of Tumour Pathology Data

We developed a specific protocol detailing the definitions and processes for pathology report data extraction based on the data requirements for the study, and guided by the 2009 College of American Pathologists (CAP) Protocol.[5] The protocol and definitions were reviewed by our study co-investigators prior to finalizing. The Pathology Review Protocol was submitted and approved by the Research Ethics Board of St. Joseph's Healthcare Hamilton (RP#09-3229).

Table 3.1 lists the data elements extracted from each pathology report. Many of the data elements were coded into categories to improve the consistency

of data capture (e.g. HER2 test type - IHC, FISH, Unknown, Not reported). Stage was determined per the American Joint Committee on Cancer (AJCC) TNM staging system, 6th edition. Pathology report data extraction was completed as a patient-level analysis by prioritising the most aggressive tumour in multifocal cases, assuming that a clinician would treat the patient based on that tumour. If the pathology report detailed multiple primary tumours (multifocal disease) for a single patient, we used a hierarchical approach to select and prioritise the most aggressive tumour, and only the details of that tumour were recorded. This hierarchy was developed with the input of a study clinician (MT) and represents the rationale that a medical oncologist would apply when selecting treatment.

The hierarchical algorithm is presented in Figure 3.1. In the first step to prioritize the most aggressive tumour (Decision Tree 1), we considered HER2 status. Subsequently, histologic grade and hormone receptor status (triple negative status) were considered. For tumours not isolated under Decision Tree 1, in Decision Tree 2, pathologic tumour stage, tumour size (greatest dimension) and nodal status were considered to determine the most aggressive tumour. Pathology information was then collected for the most aggressive tumour. HER2 test type was recorded as reported, or according to the reported result where the type of test wasn't explicitly specified. We recorded HER2 status as determined by the pathologist, as well as actual test results where possible.

Data Extractor Training & Confidentiality

All extractors first completed the TriCouncil Policy Statement on the Ethical conduct for Research Involving Humans and McMaster Chart Review tutorials, specifically the Chart Review Research Ethics Tutorial: Tutorial for Researchers Conducting Retrospective Review of Health Records. These standardised tutorials provided training on the ethical handling and protection of confidential patient information for research purposes. All extractors signed confidentiality agreements. Data extractors were trained according to a pre-specified protocol that was reviewed by study clinicians (NL, MT) and a consultant pathologist (Dr. Sharon Nofech-Mozes). Data extractors were then trained on three progressive training sets comprised of 10, 20 and 50 pathology reports respectively.

Data Cleanup

The preliminary pathology database contained only extractor-verified data points, captured using semi-automated Synoptex® software, for 16,432 patients prior to application of the exclusion criteria (Paper 3). This database was provided to ICES for linkage to other administrative data and application of the exclusion criteria. Once received at ICES, the authorized analyst replaced all identifiers with a unique anonymised ICES key number (IKN) for each patient; this IKN was used to link the pathology data to individual data from the OCR and other administrative databases while protecting patient confidentiality.

A preliminary analysis of HER2 testing patterns was performed at ICES to identify patients with more than two tests for comprehensive review. Pathology reports were linked to the appropriate patient and sequenced assuming the date of the pathology report as a proxy for the date of the HER2 test to create a longitudinal record for the patient. Through this process we identified ~15% of patients with more than two HER2 tests; the complete pathology record for each of these patients was subsequently reviewed to ensure that repeated reporting of test results did not result in double counting. If present, repeated test results were removed and only the original report was maintained. If a single pathology report detailed more than one HER2 test, we assumed that the IHC test or the test with an equivocal result was conducted first based on the expert opinion of our study clinicians (MT, NL). In some cases, a patient received initial and confirmatory testing on both a biopsy and subsequent excisional tumour sample resulting in four total tests for the patient. This was not considered double counting when two different tissue sample types could be clearly distinguished. Any pending test results were also removed if the final result was noted in a later pathology report.

Treatment Data

Trastuzumab utilisation was derived from the New Drug Funding Program (NDFP) database of CCO. Confidential identifiers were used to determine whether each member of the cohort received trastuzumab treatment through the provincial cancer program. This approach was feasible given that the NDFP

program is the only provincial drug plan to reimburse the cost of trastuzumab treatment. This treatment data was then linked to the cohort dataset to permit analysis of treatment in routine clinical practice given documented HER2 status.

Analysis

HER2 status was determined based on reported test results in accordance with Canadian Guidelines.[2] For patients who had more than one test result, we assumed positive HER2 tumour status if any one result was (MT, NL). We calculated proportions to describe HER2 testing patterns: (1) all tests performed by type, (2) all first tests by type, and (3) all first tests by type and result. We also used simple proportions to describe the use of confirmatory testing based on first test result. We also described trastuzumab use by HER2 status using proportions. First, the proportion of trastuzumab treated patients within each HER2 result category (e.g. positive, equivocal, negative) was determined. Second, we examined the distribution of HER2 unknown, missing, positive, equivocal and negative patients within trastuzumab treated and untreated patient categories.

3.4.2 RESULTS

1st HER2 Test

HER2 testing by IHC accounted for 93% (n=8,249) of documented first tests, with 62% of IHC results being positive, 10% equivocal, 11% negative and 17% unknown (Table 3.2). FISH tests accounted for 2% (n=165) of first tests documented, with results distributed as 53% negative, 5% equivocal, 19%

positive and 23% unknown. A discordant category was created to differentiate patients with records indicating that a different test was requested or performed from the type of result that was recorded (e.g. IHC test requested, FISH result provided). Discordant tests accounted for 1% (n=89) of tests, while we were unable to determine a test type for the remaining 4% (n=344) of first tests documented. Overall, 60% of all first test results were negative, 10% equivocal, 11% positive and 19% were unknown, regardless of the test type.

Confirmatory Testing

Table 3.2 provides the distribution of IHC and FISH tests used to retest patients according to first test type and result. Overall, 27% (n=2,360) of tested patients received a second test; FISH accounted for 46% of second tests, while IHC accounted for 49%. Among all patients who received a confirmatory test, 48% first tested negative, 33% equivocal and 9% positive.

HER2 Status & Trastuzumab Use

Based on the documentation of any positive HER2 test result in pathology reports, 9% (n=1,186) of the cohort was HER2 positive, while 13% of patients with a documented test were HER2 positive. Equivocal and negative patients accounted for 4% (n=355) and 65% (n=5,753) of the tested cohort, respectively. We found documentation of a HER2 test, but no reported test result, for the remaining 18% (n=1,560) of tested patients.

Almost a third of trastuzumab-treated patients did not have record of a HER2 test (29%, n=394), as shown in Figure 3.2. Trastuzumab use in HER2 negative or equivocal patients accounted for 2% (n=30) and 5% (n=69) of treated patients, respectively. The majority of patients treated with trastuzumab were HER2 positive (50%, n= 676). However, only 57% of HER2 positive patients received trastuzumab treatment within the study period, as shown in Table 3.3.

3.4.3 DISCUSSION

We documented HER2 testing patterns that provide important insight actual practice given the unexpected differences from guideline recommendations. Widespread use of IHC for initial HER2 testing was consistent with expectations, as this test is less costly and easier to perform. However, IHC testing was also prevalent as a confirmatory test, which is inconsistent with Canadian guidelines. We expected FISH to account for the majority of confirmatory tests as guidelines strictly recommend it's its use for such purposes, and given the potential clinical and economic consequences of inappropriate diagnosis. However, retesting patterns become more consistent with guidelines if viewed according to first test type and result. Among patients who initially tested IHC equivocal, FISH accounted for 90% of secondary tests, a pattern that is consistent with recommendations. Inappropriate secondary testing with FISH was also noted in IHC negative and positive patients although at lower rates (4% and 6% respectively). We also observed inappropriate secondary testing with IHC in IHC

negative and positive patients (16% and 13% respecitively). These patterns provide important insight for the model of HER2 testing to guide trastuzumab treatment in Chapter 4. With these data, we can quantify the additional costs and clinical consequences resulting from inappropriate confirmatory testing and highlight the importance of guideline adherence. Similarly, we observed treatment patterns suggesting important deviations from guidelines in the early-stage BC population. Most notably, we have shown that a remarkably large proportion (43%) of HER2 positive patients did not undergo trastuzumab treatment in our cohort (years 2006 to 2007). It remains unclear whether these patients were not eligible for treatment for reasons not captured in our model (e.g. pre-existing cardiac condition, forgone anthracycline treatment). This also has important consequences when estimating the cost-effectiveness of targeted trastuzumab therapy (Chapter 4).

We observed many unexpected testing patterns that cannot be explained without additional data. For example, initial FISH testing followed by IHC would be consistent with guidelines in patients whose tumours were first biopsied (FISH) and then resected (IHC). This is likely to occur in patients with large tumours or advanced stage at diagnosis. However, the type of tissue sample used for testing was reported infrequently or often indiscernible. In this cohort, biopsy samples were clearly documented for 7% of patients and excisional tumour block samples in 47% of patients. Similarly, different testing practices at the laboratory

level might explain some of the guideline deviation observed, but this was very challenging to determine from the pathology reports. In total, we documented that 14% of the tested patients had at least one test performed at an approved HER2 laboratory, while 2% of the cohort had a HER2 test conducted at a specific laboratory that was identifiable, but not on the list of approved labs. The testing laboratory was not distinguishable for the remainder of the cohort. Such high levels of missing data, as observed for tumour sample and laboratory variables, should not be imputed and suggest a strong pattern or bias in pathology reporting. Pathology reports tended to be highly variable in layout and quality of reporting, with tumour sample type and laboratory setting reported less frequently than any other variable. A chart review was unlikely to provide further insight, as the same pathology reports were provided to medical oncologists to inform treatment decisions. This unfortunate deficiency in pathology documentation severely limits the ability of researchers to independently evaluate practice patterns at the level of the testing laboratory, which could have important consequences for policy recommendations.

We also observed surprising trastuzumab treatment patterns. Most notably, we did not expect 43% of HER2+ patients to go untreated with trastuzumab. HER2 testing may have taken place after development of the treatment plan, resulting in the omission of trastuzumab. Alternatively, patients may forgo treatment due to fatigue or may be contraindicated due to prior cardiac

conditions. Moreover, patients who refused chemotherapy, or anthracyclines specifically, would be ineligible for publicly funded trastuzumab. While it is likely that all of the aforementioned scenarios contributed to undertreatment, we still expected higher utilisation given the aggressive nature of HER2 positive disease. This surprising finding will undoubtedly affect the expected costs and benefits of targeted trastuzumab therapy modeled in Chapter 4. Our dataset also indicates that 50% of trastuzumab treated patients did not have documentation of a positive test result in our dataset. Almost one third of treated patients did not have record of any HER2 test in our dataset, indicating that HER2 reporting to the OCR was not complete at the time of this study. The NDFP reimburses trastuzumab only for HER2 positive cases, and thus reinforces our hypothesis of incomplete reporting to the registry. The remaining 21% had some record of a HER2 test but lacked a positive result in our dataset, further supporting the assertion that addendum pathology reports are not consistently submitted to the registry. It is also plausible that additional pathology reports were not submitted for the 2% of treated patients with documentation of HER2 negative status only; trastuzumab would not be funded by CCO without evidence of a positive HER2 test.

Despite the challenges created by incomplete and inconsistent tumour pathology reporting, this study revealed previously undocumented patterns of HER2 testing, test results and treatment given HER2 status in a Canadian cohort.

We have described test sequencing contingent on prior test type and results, and treatment patterns contingent on HER2 status. Description of these patterns will facilitate quantification of the effects of actual practice on expected costs and health outcomes within a decision-analytic model.

3.5 REFERENCES

1. Phillips KA, Marshall DA, Haas JS, Elkin EB, Liang S-Y, Hassett MJ, et al. Clinical practice patterns and cost effectiveness of human epidermal growth receptor 2 testing strategies in breast cancer patients. Cancer. 2009;115(22):5166–74.
2. Hanna W, O'Malley FP, Barnes P, Berendt R, Gaboury L, Magliocco A, et al. Updated recommendations from the Canadian National Consensus Meeting on HER2/neu testing in breast cancer. Curr. Oncol. 2007;14(4):149–53.
3. Snow S, Rayson D, Barnes P, Kara T, Younis T. Does HER2/neu status affect adjuvant chemotherapy utilization? A population-based analysis. Abstract 11549. ASCO 2008 Breast Cancer Symp. Washington: American Society of Clinical Oncology (ASCO); 2008.
4. Cancer Care Ontario. Surgical Oncology News. Ontario CC, editor. Surg. Oncol. Newsl. Toronto; 2009;4(1).
5. Lester SC, Bose S, Chen Y-Y, Connolly JL, de Baca ME, Fitzgibbons PL, et al. Protocol for the Examination of Specimens from Patients with Invasive Carcinoma of the Breast. Version 3.0.0.0. College of American Pathologists; 2009 p. 1–53.

3.6 TABLES

Table 3.1. Data elements extracted from pathology reports.

Data Element	Description
Date of pathology report:	The date of the pathology report was extracted to create a longitudinal record of HER2 tests provided to a single patient
Size of invasive tumour component:	The size of the invasive tumour component extracted to supplement tumour size or t-stage information missing from OCR data
T-stage:	The t-stage was extracted to supplement T-stage information missing from OCR data
Number of positive lymph nodes:	The number of positive lymph nodes was collected as reported to supplement N-stage information missing from OCR data
N-stage:	The N-stage was collected as reported to supplement N-stage information missing from OCR data
M-stage:	M-stage information was collected as reported to supplement metastasis or M-stage information missing from OCR data
TNM stage:	The AJCC 6th edition Tumour Node Metastasis (TNM) stage was collected as reported to supplement TNM staging missing from OCR data
Histologic tumour grade:	Histologic tumour grade (1, 2, or 3), either Scarff-Bloom-Richardson or Nottingham scales, was collected as reported to account for potential aggressive nature of the tumour; only overall tumour grades were collected and single nuclear or mitotic grades were excluded.
Laterality:	The laterality (left or right) of the primary tumour was collected as reported
Estrogen receptor (ER) status:	Estrogen receptor status (positive or negative) was collected as reported; data extractors did not determine ER status without an explicit status determination by the reporting pathologist
Progesterone receptor (PR) status:	Estrogen receptor status (positive or negative) was collected as reported; data extractors did not determine ER status without an explicit status determination by the reporting pathologist

Data Element	Description
Type of HER2 test(s) provided**:**	The type of HER2 test provided (IHC, FISH or unknown) was collected based on explicit observation of the test name in the pathology report, or based on the use of keywords associated with each specific HER2 test; if the report suggested that HER2 testing was performed but the type of test is not clear, the test type was reported as "unknown"; if the type of test was not explicitly reported, but HER2, ER and PR results were reported or requested together in immediate proximity within the report, an "IHC Assumed" category was assigned owing to the IHC methods routinely used for ER and PR measurement
HER2 test result(s)**:**	HER2 test results (positive, equivocal or negative) were recorded per the reporting pathologists' HER2 status determination as detailed in the pathology report; IHC and FISH results were distinguished, and unknown test type results were also collected
Actual HER2 test result(s)**:**	If the actual HER2 test result was reported, this information was collected as detailed in the pathology report. Actual results considered were: • IHC category (i.e. IHC 0, 1+, 2+ or 3+) • IHC percent cells stained (i.e. 0%) • HER2:CEP 17 ratio (i.e. 3.2:1.1, 3.0) • HER2 gene copy number (i.e. 5)
Type of tissue sample:	The type of tissue sample (excisional tumour block or biopsy) used for HER2 testing was recorded, if reported and distinguishable; if the tumour source was unclear, the field was marked as "unknown"
HER2 testing laboratory:	The laboratory where the HER2 test was conducted was recorded as reported in the pathology report if distinguishable; the HER2 testing lab was not assumed to be the same as the pathology report source because HER2 testing is often referred to other labs; a predefined list of approved Ontario HER2 testing pathology labs was provided for data capture; this field was marked unknown or left blank if the lab was not on the list or could not be determined

Table 3.2. Distribution of second (confirmatory) tests given first HER2 test type and result for all tested patients (n=8,854). Percentages indicate the proportion of patients retested according to the first test type and result. The unknown category represents records where a HER2 test result was clearly indicated, but the type of test (IHC or FISH) was not distinguishable. An unknown result represents records where a HER2 test was clearly indicated, but the result was not reported. The discordant category represents records where the type of HER2 test requested (e.g. IHC) differed from the type of result reported (e.g. FISH) across two separate pathology reports.

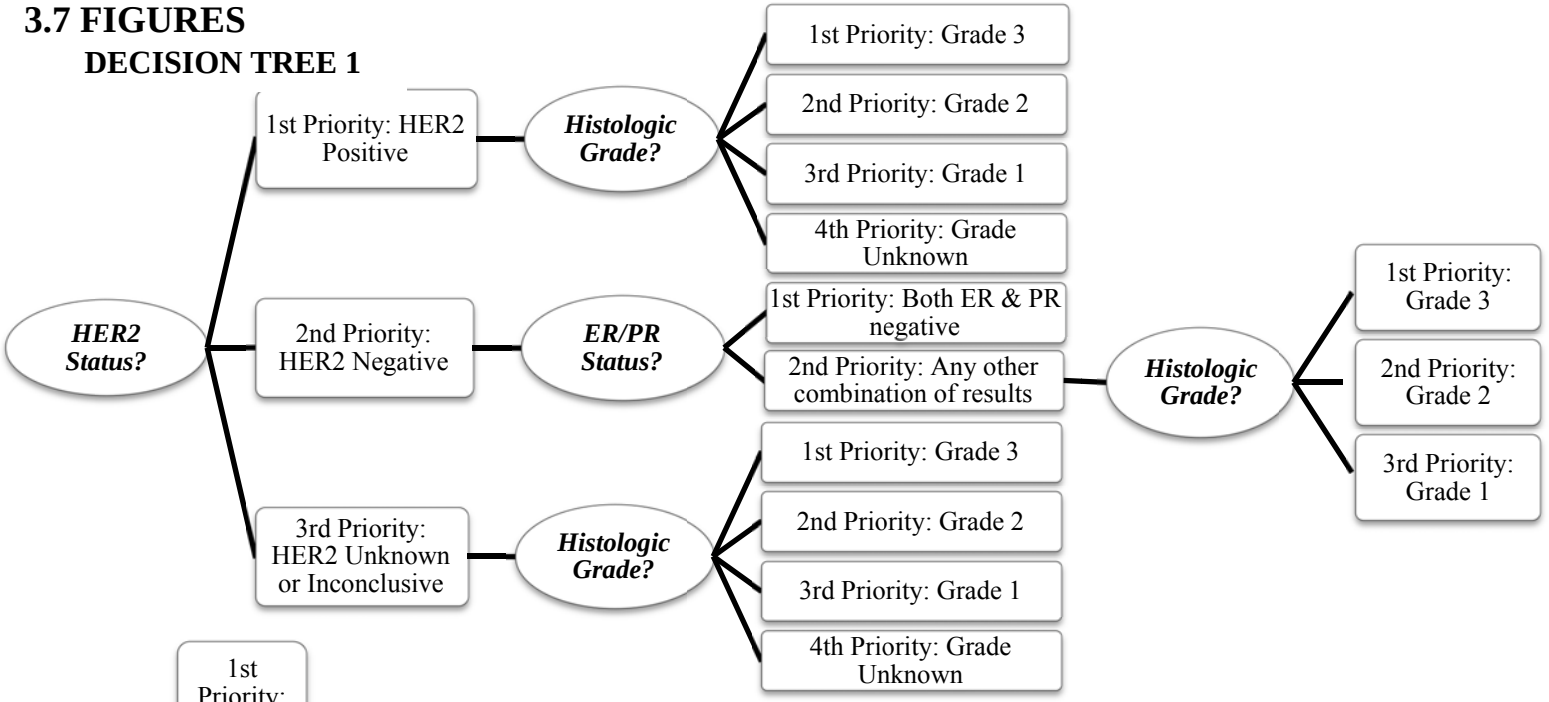
			2nd HER2 Test Type				
			FISH, % (n)	**IHC**, % (n)	**Unknown**, % (n)	**None,** % (n)	**Total**
1st HER2 Test Type & Result	**IHC**	negative (0, 1+)	4% (182)	16% (798)	1% (71)	79% (4,024)	5,075
		equivocal (2+)	90% (761)	1% (12)	0% (3)	9% (74)	850
		positive (3+)	6% (56)	13% (120)	2% (15)	79% (727)	918
		unknown	1% (20)	10% (135)	1% (16)	88% (1,235)	1,406
	FISH	negative	5% (4)	7% (6)	1% (1)	87% (76)	87
		equivocal	0% ()	11% (1)	0% ()	89% (8)	9
		positive	0% ()	10% (3)	0% ()	90% (28)	31
		unknown	0% ()	5% (2)	3% (1)	92% (35)	38
	Unknown	negative	0% ()	15% (12)	4% (3)	81% (63)	78
		equivocal	0% ()	0% ()	0% ()	100% (2)	2
		positive	0% ()	25% (1)	0% ()	75% (3)	4
		unknown	0% (1)	20% (53)	5% (12)	75% (194)	260
	Discordant	negative	80% (43)	11% (6)	0% ()	9% (5)	54
		equivocal	29% (6)	5% (1)	0% ()	67% (14)	21
		positive	79% (11)	7% (1)	0% ()	14% (2)	14
	Unsuitable sample		14% (1)	29% (2)	0% ()	57% (4)	7
	Total (with % of 2nd tests)		46% 1,085	49% 1,153	5% 122	- 6,494	8,854

Table 3.3. Distribution of trastuzumab use according to final HER2 status. Percentages indicate the proportion of patients within a status category that did or did not receive trastuzumab treatment within the study period (March 2009). The unknown category indicates patients with record of a HER2 test request without a result. The missing category indicates patients who did not have documentation of a HER2 test.

		Trastuzumab Use		
		No trastuzumab, % (n)	**Trastuzumab**, % (n)	**Total**
HER2 Status	**Equivocal**	81% (286)	19% (69)	355
	Negative	99% (5,723)	1% (30)	5,753
	Positive	43% (510)	57% (676)	1,186
	Unknown	88% (1,372)	12% (188)	1,560
	Missing	91% (4,148)	9% (394)	4,542
	Total	90% (12,039)	10% (1,357)	13,396

3.7 FIGURES

DECISION TREE 1



DECISION TREE 2

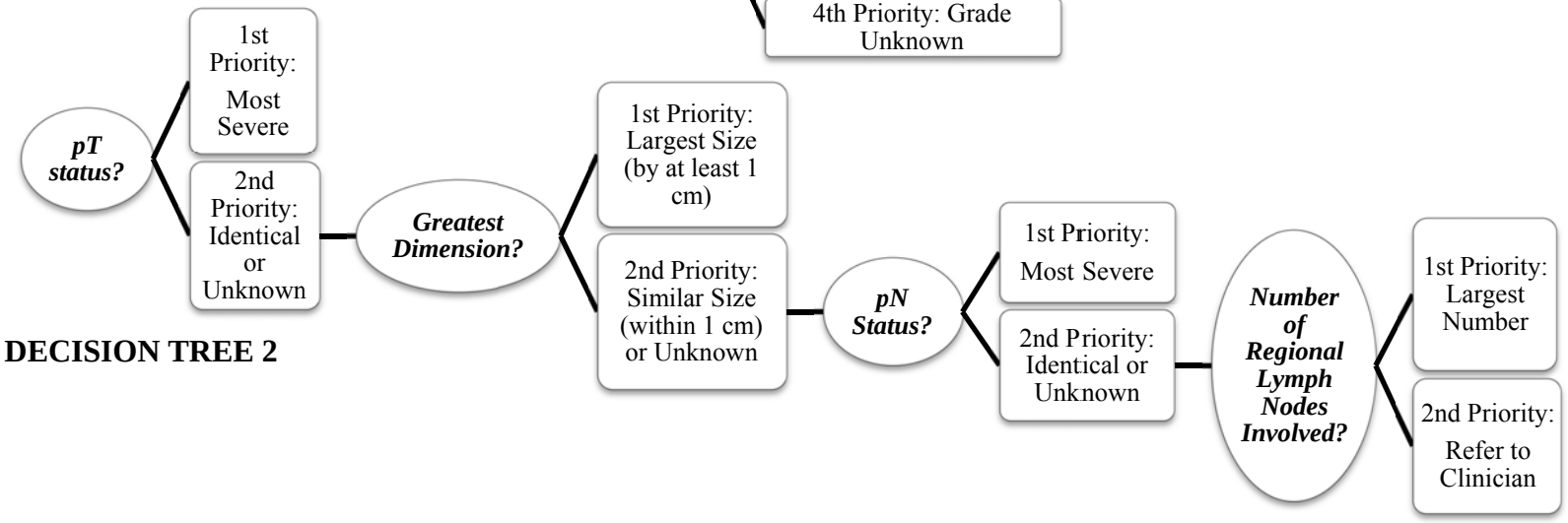


Figure 3.1. Decision trees defining the hierarchical algorithm used to select the most aggressive tumour in pathology reports with multifocal disease. Decision Tree 1 is used to classify aggressive tumours based on HER2 status, ER/PR status and histologic grade. Decision Tree 2 is used to prioritise tumours not isolated under Decision Tree 1. Tumours with the highest priority are reported. ER: estrogen receptor; pN: pathologic nodal stage; PR: progesterone receptor; pT: pathologic tumour stage.

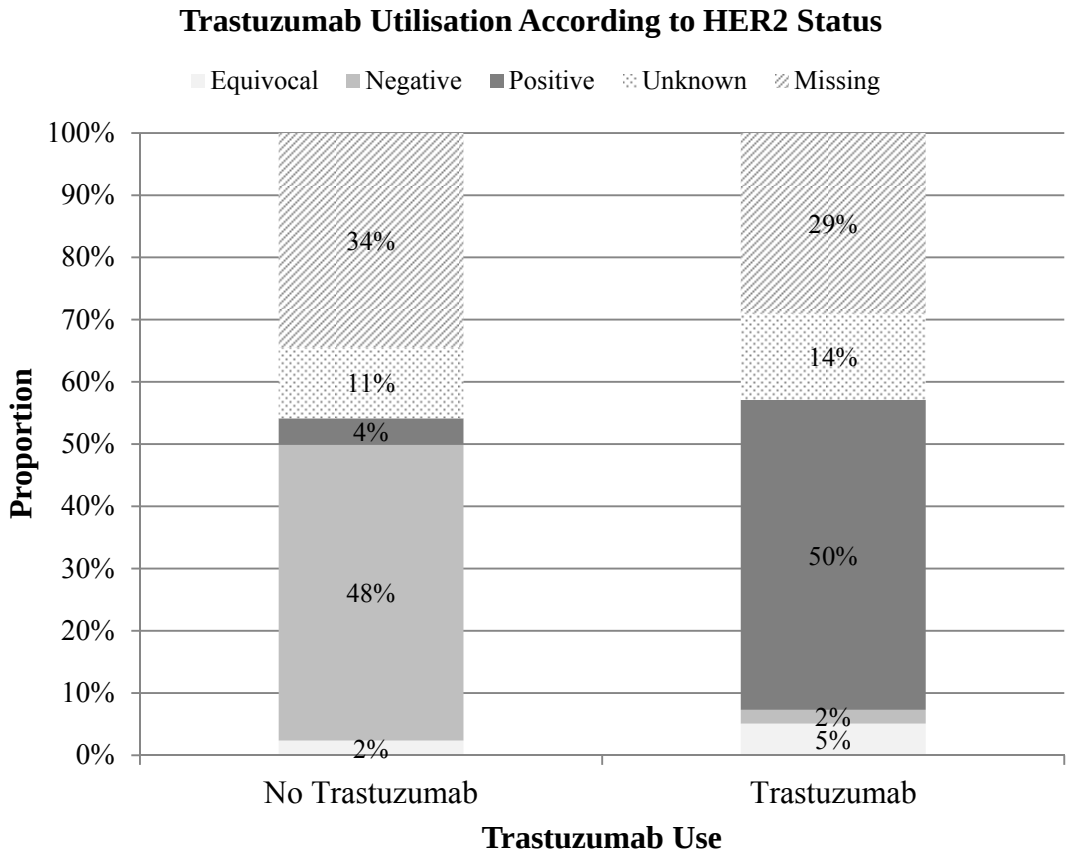


Figure 3.2. Documented HER2 status among patients who did or did not receive trastuzumab per New Drug Funding Program records up to March 2009. Missing patients did not have a documented HER2 test in reviewed pathology reports, and unknown patients had documentation of a HER2 test without a result.

CHAPTER 4

A COST-EFFECTIVENESS ANALYSIS OF ALTERNATIVE STRATEGIES TO DIAGNOSE HER2-POSITIVE EARLY-STAGE BREAST CANCER AND DIRECT TARGETED TRASTUZUMAB IN THE ADJUVANT SETTING USING CURRENT PRACTICE PATTERNS

PAPER 4†: Ferrusi, I. L., Kulin, N.A., Goeree, R, Leighl, N.B., Pullenayegum, E., Phillips, K.A., Marshall, D.A. Cost-utility analysis of alternative human epidermal growth factor receptor-2 (HER2) targeted test-treat strategies for breast cancer in routine practice. Breast Cancer Research and Treatment (*in submission)*

†Academic contributions by Ilia L. Ferrusi: concept and design, acquisition of data, analysis and interpretation of data, writing, critical revision and finalisation of the manuscript for important intellectual content.

SUPPLEMENTAL METHODS AND ANALYSES: Ferrusi, I. L., Kulin, N.A., Goeree, R., Leighl, N. B., Pullenayegum, E., Marshall, D. A.

4.1 PREFACE

The study of HER2 testing and trastuzumab treatment patterns (Chapter 3) revealed important differences from guidelines in Ontario practice, with some 12% of IHC equivocal cases that did not receive FISH clarification . We used these findings to examine the impact of current practice on the expected incremental cost-effectiveness of alternative HER2 testing strategies to direct targeted trastuzumab treatment in early-stage BC patients. We approached this evaluation by first estimating ICERs under a 'guideline adherence scenario' which was then compared to alternative scenarios that reflected current testing practice, current treatment practice, or both. Under the current treatment practice scenario, we estimated the impact of inappropriate retesting of some IHC+ and IHC- patients, and incomplete retesting of IHC equivocal patients. Under the current practice scenario, we estimated the impact of treating only 60% of HER2+ patients. We hypothesize that these alternative scenarios will have an important impact on ICER estimates, resulting in reduced LYs and QALYs gained as a result of increased misdiagnosis and inappropriate treatment based on those misdiagnoses. This hypothesis is based on the observed importance of test sequencing and test accuracy as key influences on the ICER of trastuzumab treatment in Chapter 2. Specifically, Elkin[1], Lidgren[2] and Blank[3] demonstrated that the ICER was most sensitive to changes in test sensitivity and specificity in the metastatic and adjuvant settings. Each analysis demonstrated the key role of FISH testing to confirm

IHC equivocal results or primary FISH testing, resulting in reduced false positive and false negative diagnoses, minimizing under and over-treatment and producing a net gain in LYs and QALYs.

The importance of incorporating current practice patterns in cost-effectiveness analysis is supported by recent shifts in health policy and research agendas. Canadian and American health outcomes researchers and policy makers have called for an increase in pragmatic 'comparative effectiveness research' reflecting routine practice and typical patient populations.[4] Thus, the use of Ontario testing and treatment behaviour in this model serves to improve the relevance of this study to policy-makers, inform current research priorities and provide insight beyond previously published models.[5]

Further to the need for more pragmatic evaluations, we identified a deficiency of calibrated disease models reported in the literature.[5] Calibration of disease models serves to ensure that modeled disease progression and mortality reflect local disease patterns, rather than any selected clinical trial population used to inform the model.[6,7] In the absence of calibration, the relevance of model results to the local setting is difficult to determine. Therefore, we validated and calibrated the disease model to improve generalisability to the Ontario setting and to distinguish it from existing published cohort models.

The manuscript and supplemental materials in this chapter detail the technical specifications, validation, calibration and results of a decision-analytic mod-

el used to compare alternative test-treat strategies to target adjuvant trastuzumab in a cohort of Canadian early-stage BC patients. We used a decision tree linked with a Markov model to represent alternative test-treat strategies and to extrapolate the long-term consequences of those strategies in the lifetime perspective. We lacked access to patient level data to inform the natural history of BC, and were therefore limited to the Markov model approach over discrete event simulation to represent the natural history of BC. Our model accounts for several characteristics determined to be clinically relevant and influential in Chapter 2, including:

(1) Practice Patterns:
    a. Alternative HER2 test sequencing strategies to examine the use of confirmatory testing;
(2) Flexibility to capture complete guideline adherence or variable use of testing and targeted treatment per current practice patterns;
(3) New Technologies: capture of CISH and SISH testing techniques;
(4) Toxicity: Capture of treatment-induced symptomatic cardiac toxicity leading to treatment discontinuation in a separate disease state;
(5) Probabilistic Sensitivity Analysis:
    a. Assessment of the uncertainty around the duration of treatment effect by modelling this parameter probabilistically; and

b. Probabilistic modelling of the uncertainty around HER2 test sensitivity and specificity.

This chapter presents the submitted manuscript, detailing the model structure, input parameters and findings. Supplemental material is provided to justify the model design, describe parameter derivation, illustrate model validation and calibration, and to expand upon findings.

4.2 PAPER 4

Ferrusi, I. L., Kulin, N.A., Goeree, R, Leighl, N.B., Pullenayegum, E., Phillips, K.A., Marshall, D.A. Cost-utility analysis of alternative human epidermal growth factor receptor-2 (HER2) targeted test-treat strategies for breast cancer in routine practice. Breast Cancer Research and Treatment (*in submission)*

ABSTRACT

PURPOSE: Adjuvant trastuzumab plus chemotherapy is standard treatment for patients with early stage human epidermal growth factor receptor-2 positive (HER2+) breast cancer (BC). HER2 overexpression is typically detected by immunohistochemistry (IHC) or HER2 amplification by fluorescence in situ hybridization (FISH), although newer chromogenic and silver-based tests (CISH, SISH) are available. The purpose of this study was to estimate the impact of current testing practice relative to guideline recommendations, and of newer alternatives.

METHODS: We used a decision tree and linked Markov model to estimate the incremental cost-utility of six test-treat strategies to identify HER2+ BC patients for adjuvant trastuzumab in Canada. The impact of local disease epidemiology and clinical practices were examined in probabilistic scenario analyses.

RESULTS: In the base case, assuming 20% HER2+ BC prevalence and perfect adherence to testing and treatment guidelines, testing with FISH alone was dominant; it decreased mean costs by $815 and improved outcomes by 0.0022 quality-adjusted life-years compared to the current approach of initial IHC testing with equivocal confirmation by FISH. All strategies performed similarly to the current testing approach producing uncertainty regarding the optimal strategy as evidence by the cost-effectiveness acceptability curves. Health outcomes were reduced in simulations of current testing and treatment practice compared to guideline adher-

ence. Results were sensitive to HER2+ prevalence, test accuracy, treatment effect carryover and practice behaviors.

CONCLUSIONS: This analysis suggests that initial FISH testing may improve outcomes and reduce lifetime costs from the payer's perspective by reducing over and under-treatment with trastuzumab in early stage BC.

Keywords: HER2, cost-effectiveness, personalized medicine

INTRODUCTION

Human epidermal growth factor receptor-2 (HER2) overexpression in breast cancer (BC) leads to more aggressive disease and responds poorly to standard cytotoxic therapy.[1–3] This BC subtype can be treated with trastuzumab, a targeted antibody that prevents tumor proliferation by blocking the HER-2 protein.[4] Several clinical studies have demonstrated adjuvant trastuzumab efficacy following or concurrent to cytotoxic chemotherapy for 9[5] or 52[6–9] weeks, leading to the approval of trastuzumab to treat early-stage BC. [10] A recent systematic review demonstrated the significantly improved ($p<0.05$) efficacy of trastuzumab plus chemotherapy in early-stage HER2+ BC with hazard ratios (HR) of 0.66 and 0.60 for overall and disease-free survival (DFS) vs. standard chemotherapy alone. [11] To date, HER2 testing to target trastuzumab treatment is among the most successful examples of personalized medicine in clinical practice.

Targeting 'the right treatment to the right person at the right time' is fundamental to personalized medicine, and such a strategy aims to deliver appropriate care - minimizing over- and under-treatment. Evidence of HER2 overexpression is required for publicly-funded trastuzumab treatment in Ontario. [12,13] Previous economic evaluations of trastuzumab in BC demonstrated that initial immunohistochemistry (IHC) testing for all, with fluorescence in situ hybridization (FISH) confirmation of equivocal (IHC2+) results, or FISH testing for all, were cost-effective strategies to target trastuzumab plus chemotherapy vs. standard

chemotherapy alone. [14–16] These studies also reported that HER2 test properties and test strategy were the most influential parameters on cost-effectiveness estimates. [17,18] However, utilization studies have repeatedly demonstrated inconsistent adherence to HER2 testing guidelines. [19–22] The introduction of novel chromogenic and silver in situ hybridization (CISH; SISH) tests has further complicated the issue. [23] Although utilization of these tests in practice is limited, their purported advantages include improved result consistency, easier use and cost savings. [24]

Given the potential advantages of CISH and SISH, we used a decision-analytic framework to evaluate the use of all ISH strategies compared to a strategy representing the most common HER2 testing approach to target adjuvant trastuzumab in BC. We also estimated the impact of various regional factors, including local HER2+ disease prevalence and deviations from testing and treatment guidelines in practice. Practice deviations from testing guidelines were expected to increase inaccurate diagnoses, resulting in fewer life-year gains. Similarly, under-treatment with trastuzumab was expected to reduce costs while also reducing health outcomes compared to treatment consistent with guidelines. The findings from this study sheds light on opportunities to improve targeted BC treatment, while also providing insight into the evaluation of future personalized medicines.

METHODS

We compared the incremental costs and health outcomes associated with alternative strategies to diagnose HER2+ early-stage BC and target adjuvant trastuzumab therapy to estimate the incremental cost-utility ratio (ICUR). We examined the benefit of trastuzumab sequential to standard chemotherapy in the base case, but considered concurrent treatment in sensitivity analyses. Future costs and outcomes were discounted at 5% annually per Canadian guidelines. [25]

Population & Setting

We modeled a cohort of women age 50 years at incident diagnosis of early-stage (American Joint Committee on Cancer stage I, II or III) invasive BC, having completed primary surgical treatment and adjuvant cytotoxic chemotherapy. We assumed 50% of patients had hormone receptor-positive tumors consistent with an Ontario population-based epidemiological study. [20] The analysis was carried out from the perspective of the payer (Ontario Ministry of Health and Long-Term Care [MOHLTC]) in the lifetime time horizon (50 years). In the base case, we assumed that 20% of the 100,000 patient cohort was HER2+.

Model Structure

We used a decision-tree to model 6 test-treat strategies, combined with a Markov model to capture the long-term consequences of testing and treatment decisions. All programming was performed in Microsoft Excel® & Visual Basic®

for Applications 2010. Table 1 presents the 6 strategies, assuming FISH is the gold standard. This is a common assumption in models of HER2 testing and is underscored by the promotion of FISH for confirmation in testing guidelines. Three strategies modeled initial IHC testing with confirmatory FISH, CISH or SISH for IHC2+ results only, respectively. Three more strategies modeled testing with FISH, CISH or SISH alone, assuming that either of the new ISH technologies would be used in place of FISH if incorporated into practice. We assumed all patients received a HER2 test at diagnosis, consistent with Canadian guidelines[17] and MOHLTC policy, and did not model a 'do nothing' strategy. IHC (FISH) served as the referent strategy (against which all others were compared) because it represented the most common testing practice. The sensitivity and specificity of each test was modeled within the decision tree, allowing us to represent disease progression and treatment response according to true HER2 status. We assumed that retesting of HER2 status at recurrence was not performed. Treatment was assigned per HER2 test result(s). In the base case, we assumed that only patients with a positive result (IHC 3+, ISH+) received 1 year of adjuvant trastuzumab sequential to standard chemotherapy per guidelines. [26]

Figure 1 shows the Markov model of early-stage BC. Patients transitioned between states annually with a half-cycle correction. The model considered locoregional recurrences (LRR) as well as distant (Stage IV; metastatic) recurrences (DR).

Transition Probabilities (Table 3)

The accuracy of each HER2 test was estimated from a systematic review (Appendix 1) of concordance studies consistent with the latest American and Canadian guideline thresholds. [17,18] Data were extracted from the 12 relevant studies that met inclusion criteria (Appendix 1) and pooled using the inverse-variance weighting method. [27] The corresponding transition probabilities (Table 2) were fit to beta or Dirichlet distributions.

Disease progression parameters (Table 3) were fit to beta distributions unless otherwise noted. All annual probabilities were derived as reported in the literature or extracted from Kaplan-Meier curves using GetData Graph Digitizer v2.25. All disease progression parameters were calibrated to Canadian BC mortality patterns using a stepwise random parameter search approach and best fitting parameter sets were selected by the least squares method (Appendix 2). The effect of trastuzumab treatment (Table 3) was derived from a recent Cochrane review of trastuzumab in the adjuvant setting,[11] assumed to carryover for 5 years in the base case. We assumed no mortality due to treatment-related congestive heart failure (CHF) per the Cochrane review. [11] We did, however, assume a 50% reduction in treatment effect for patients who discontinued due to treatment-related CHF.

Costs

All costs are in 2013 Canadian dollars (Table 3) and fit to gamma distributions for probabilistic analysis. [28] Only direct costs to the healthcare system were considered under the payer perspective. [29] All costs estimated from studies of resource utilisation were varied by 20% in probabilistic analysis, while trastuzumab costs were fixed. Costs for all ISH tests were equal under to the reimbursed rate for such tests under the public payer healthcare system.

Utilities

All utilities (Table 3) were derived from the literature and fit to gamma distributions. [28] We assumed that treatment discontinuation due to cardiac toxicity produced a utility decrement in the first year of the DC state. That decrement corresponded to NYHA Class III/ IV CHF[30] and was assumed to resolve within 1 year, after which patients returned to a full-health utility. [31] We also assumed a utility decrement for patients experiencing asymptomatic LVEF decline similar to NYHA class I/II CHF. [30] Patients with LRR received a utility decrement associated with its treatment[32] with utility returning to no-evidence of disease (NED) levels in the post-LRR state. [31] A utility decrement was also applied to the DR state, which remained constant until death. [31] We programmed a utility decrement with age.

Sensitivity & Scenario Analyses

Parameter uncertainty was represented through probabilistic sensitivity analysis of transition probabilities, costs and utilities in the base case. A series of

one-way deterministic sensitivity analyses tested the sensitivity of the model to assumptions about specific parameters or patterns of behavior (Table 3). We examined the sensitivity of the model to a range of values from the literature (Figure 3) for: HER2+ disease prevalence, test sensitivity and specificity, duration of trastuzumab benefit and discount rate. We also examined the impact of not re-treating with trastuzumab at LRR and DR. Treatment benefit for patients who discontinued therapy was varied to 0% and 100%. The impact of a 20% reduction in the unit price of trastuzumab was also considered. Finally, we examined the combined effect of sequential and concurrent treatment and corresponding rates of cardiac toxicity. [11]

We examined uncertainty in treatment benefit carryover by allowing it to vary between 1 and 10 years. We also conducted a series of probabilistic scenario analyses examining the consequences of current testing and treatment practice, and local patterns of HER2+ disease prevalence (Table 3). These scenarios focused on strategies A and D alone as only these strategies reflected the IHC and FISH technologies currently used in Ontario. Under current testing patterns, we modeled inappropriate retesting of some IHC3+ (positive) and IHC1+ or 0 (negative) results, and incomplete retesting of IHC2+ (Table 3). Under current treatment, we modeled some inappropriate treatment of HER2- or HER2 equivocal patients, and patterns of under-treatment of HER2+. We also modeled rates of treatment discontinuation due to cardiac toxicity and non-cardiac reasons in prac-

tice. [33] We examined the joint impact of current testing, treatment and local disease prevalence in a final scenario.

RESULTS

Compared to the reference strategy IHC (FISH) the strategy of FISH testing alone decreased costs by $815 and increased survival by 0.0109 LYs and 0.0022 QALYs in probabilistic analyses, producing a negative ICER (Table 4). All other strategies were dominated with utility weighting of health outcomes. Figure 2 presents the results of the probabilistic base case analysis on the cost-effectiveness plane. This demonstrates the significant overlap between all strategies with the referent, with the exception of IHC (SISH) which tended to produce higher incremental costs. CISH and SISH did not perform equivalent to FISH in this analysis as they tended to have poorer sensitivity, resulting in fewer QALY gains across the simulations. Each strategy was associated with significant uncertainty, with simulations spanning all quadrants. This uncertainty is attributed to the similar performance of each strategy relative to the referent in diagnostic accuracy and long-term outcomes, producing very small incremental differences in LYs and QALYs (Table 4). The results generally suggest that initial testing with FISH leads to improved outcomes at a reduced cost vs. IHC (FISH) under a scenario of perfect guideline adherence and a constant disease prevalence of 20%. When the duration of treatment benefit was varied probabilistically, FISH alone

strategy remained most efficient, producing more QALYs at lower cost vs. the referent.

All current practice scenarios are presented in Table 4. Under current testing practice, the strategy of FISH alone produced fewer cost savings but more QALY gains (0.035) vs. referent (ICUR=-$27,664/QALY) than under guideline adherence in the base case (ICUR=-$374,766/QALY). This is because current testing practice impacted the use of confirmatory FISH under the IHC (FISH) referent strategy only, causing an 11% increase in the number of inaccurate diagnoses in the reference strategy (Appendix 6). The impact of under-treatment of HER2+ and over-treatment of some HER2- patients was modeled under current treatment practice, where the FISH strategy produced lower incremental costs (-$188) but also fewer incremental QALYs (-0.008) compared to guideline adherence in the base case. This is because of the treatment of some false negative patients with trastuzumab under referent IHC (FISH), and the larger proportion of undertreated true positives with FISH alone. Under-treatment also had an impact on quality-of-life in earlier model cycles by avoiding treatment-related utility decrements. This impact received greater weighting due to discounting. The local disease prevalence scenario resulted in FISH being dominated by IHC (FISH). When Ontario HER2+ disease prevalence was modeled probabilistically, FISH alone cost less (-$1,155) than strategy IHC (FISH), but also produced fewer QALYs (-0.013). These strategies produced very similar outcomes under the local

HER2 prevalence rate, and the incremental differences between each were further diminished by discounting. When all local practice and disease prevalence parameters were modeled jointly, FISH alone was dominated by referent IHC (FISH), costing more ($43) and producing fewer QALYs (-0.003).

Univariate sensitivity analyses are summarized in a tornado diagram (Figure 3). The tornado diagram represents the difference in the ICER if the FISH alone strategy under an alternative set of parameter inputs as compared to the base case ICER. For example, under "Ontario testing and treatment practice" analysis, only the following parameters were changed from base case: (1) the probability of a second ISH test contingent on IHC result, and (2) the probability of treatment contingent on HER2 test result. Univariate analyses revealed local disease prevalence and current testing practices had the greatest impact on costs and outcomes than treatment practices (Figure 3). The model was also sensitive to test accuracy and assumptions about treatment benefit carryover. Changes in the discount rate had a larger impact than sequential and concurrent therapy, a 20% reduction in trastuzumab price, or assumptions about retreatment with trastuzumab at LRR and DR. Assumptions about the treatment effect in patients that discontinued trastuzumab had the smallest impact on outcomes.

Cost-effectiveness acceptability curves (CEACs; Figure 4) were estimated for all strategies assuming guideline adherence, and for FISH alone and IHC (FISH) under current practice (testing, treatment, HER2+ disease prevalence). We

varied treatment effect carryover in both cases given the important influence of this parameter. Under guideline adherence (base case, Figure 4A), FISH was dominant at all willingness-to-pay (WTP) thresholds >$2,000/QALY. All strategies involving primary IHC with confirmation of IHC2+ results with an ISH test, tended to have the lowest probabilities of being cost-effective. However, the CEAC demonstrates that as WTP increases, the probability of FISH being cost-effective decreases up to ~$200,000/QALY. Moreover, the probability of FISH being cost-effective relative to all other strategies never exceeds 40.5%. When current practices and disease prevalence were considered in the context of predominant testing technologies (Figure 4B), FISH was again dominant at all WTP thresholds >$7,000/QALY. The probability of FISH being cost-effective did not exceed 50.7% under current practice. Both analyses demonstrate that there is considerable uncertainty about which test-treat strategy is most cost-effective up to a WTP threshold of $50,000/QALY. This is attributed to the similar accuracy of each testing strategy.

DISCUSSION

Given the high cost of trastuzumab treatment and its life-saving potential in aggressive HER2+ BC, it is important to accurately target treatment to responders. [13,34] We demonstrate that initially testing all patients with FISH could improve long-term outcomes and reduce costs vs. what is currently the prevalent strategy of initial IHC with FISH confirmation of IHC2+. These improvements

were achieved by reducing the number of FP and FN diagnoses, resulting in a reduction in over- and under-treatment which offset the additional cost of testing all patients with FISH. However, these conclusions were highly sensitive to assumptions about the prevalence of HER2+ disease, test sensitivity and specificity, assumptions about the longevity of treatment benefit, and adherence to testing and treatment guidelines.

Our analyses revealed the importance of considering local disease epidemiology when modeling a targeted therapy and companion diagnostic. When the 13% Ontario HER2+ disease prevalence rate was modeled in lieu of the 20% assumed base case prevalence, the impact of inaccurate diagnoses and trastuzumab under-treatment was minimised such that IHC (FISH) was no longer dominated by strategy FISH alone. However, Ontario prevalence may be underestimated[20] as other North American cohorts report 18.6% prevalence. [35] When we modeled 30% prevalence, FISH dominated IHC (FISH). This suggests that a more expensive and accurate testing strategy tends to become cost-effective as HER prevalence increases. This was not apparent from previous studies of testing and treatment, where disease prevalence was not evaluated[14,15]or modeled as a function of test properties. [16] Our model was consistent with previous studies demonstrating sensitivity to test sensitivity and specificity. [14,15] We also demonstrated that ICUR value estimates were sensitive to uncertainty about treatment benefit carryover, but this uncertainty did not impact the rank order of the strategies.

A common theme through all analyses was the uncertainty regarding the most cost-effective test-treat strategy because all strategies produced very similar rates of accurate diagnoses. However, strategy FISH alone generally produced health outcome gains at a reduced cost in the lifetime perspective. Improvements in testing and treatment guideline adherence could improve outcomes, as illustrated in our scenario analyses of current test-treat strategies. Testing guideline deviations had a greater impact on costs and health gains than that of treatment guideline deviations. This lends further support to initial testing with the gold standard (FISH), particularly if IHC2+ verification is not always performed in practice. Our guideline adherence scenario findings were similar to two other models that considered testing and treatment[15,16], all of which demonstrated that initial FISH testing and treatment of FISH positives would increase QALYs in the adjuvant setting. Previous studies compared testing and treating against a 'do nothing' reference strategy without HER2 testing and targeted trastuzumab treatment. The inclusion of a 'do nothing' referent strategy is likely the reason why previous Swiss[16] and Swedish[15] analyses suggested initial FISH was associated with additional costs and QALYs compared to the cost savings and QALY gains of FISH demonstrated here. The choice of reference strategy also impacts the magnitude of the incremental costs and QALY gains, and accounts for the very small incremental gains in the present study. However, we could not justify evaluating a 'do nothing' strategy when local policy requires HER2 testing for all BC patients.

Adjuvant trastuzumab was evaluated in two previous Canadian studies. [36,37] Skedgel et al. considered sequential therapy using a similar disease model and demonstrated that cardiac toxicity did not have an important influence on the incremental cost-effectiveness ratio, despite modelling additional toxicity with trastuzumab retreatment at recurrence.[36] Both Skedgel and the current study demonstrated the impact of treatment benefit carryover on the ICER, and indicated that the optimal treatment regimen (sequential vs. concurrent) and duration (1 year vs. 9 weeks) remain important areas of research and continued follow-up. Hedden et al. [11] examined trastuzumab in the context of 'real world' Canadian resource utilisation, modeling concurrent therapy. That analysis demonstrated greater QALY gains, as studies assuming concurrent therapy have suggested a stronger treatment effect. This in addition to the 'do nothing' referent strategy likely accounted for the larger predicted QALY gains. Hedden also modeled the efficacy of trastuzumab at recurrence. We did not consider this given that the effectiveness of trastuzumab after progression is not well characterized.

We have examined newer ISH testing technologies as either initial testing or to verify IHC2+. CISH and SISH are both supported by current guidelines and anecdotal evidence indicates utilisation in other provinces. [23] These tests could facilitate wider ISH testing in practice because they require a standard light microscope (vs. the fluorescence microscope needed for FISH). [24] The literature also suggests that these tests may be less expensive and provide more stable re-

sults than conventional FISH, while improving concordance between needle biopsy and excision samples vs. IHC. [24,38] However, these tests must demonstrate equivalent sensitivity and specificity to FISH if the latter is to be replaced in practice. Both CISH and SISH demonstrated specificity very close to FISH (Table 2), resulting in little over-treatment. However, neither CISH nor SISH were as sensitive as FISH in this analysis, resulting in missed positive diagnoses and subsequent under-treatment. Our analysis indicated that CISH in particular may perform very closely to FISH as a diagnostic test (Table 4). However, these conclusions were based on the assumption that HER2 gene expression is the single best predictor of response to trastuzumab. Overexpression of the HER-2 protein is directly correlated with HER2 gene overexpression but as many as 10% of tumours, known as non-amplified overexpressors, may overexpress HER-2 protein without gene amplification. [39] Moreover, studies challenge convention with evidence suggesting that some IHC 2+ patients may benefit from trastuzumab. [40] This model did not account for the subset of IHC 3+ FISH- patients who may benefit from therapy. [39] It is also important to note that this analysis modeled a 12% false-positive rate for IHC that exceeds the maximum 10% set recommended by the College of American Pathologists guidance on quality assurance for HER2 testing. [18,41] However, our findings were robust and FISH remained the dominant strategy in sensitivity analysis when the IHC false-positive rate was set to 10% (data not shown).

This analysis faced challenges common to all decision-analytic models. As with many disease models a lack of detailed epidemiological data meant that our analyses combined all non-metastatic recurrences into a single state. [36,37] This limits our ability to answer detailed questions about the timing and types of recurrences occurring. We did endeavor to make this model representative of Canadian disease epidemiology through calibration. This process provided a better understanding of the shortcomings of the current model's predictions vs. other models without reported calibration. The current model tended to over-predict mortality in earlier cycles and was therefore likely to produce more conservative estimates of the impact of inaccurate diagnoses across all strategies. This has important implications for an analysis of HER2 targeted therapy, because aggressive HER2+ disease tends to lead to earlier and more frequent recurrence or progression. Thus, a conservative model would tend to minimise the gains produced by increased accuracy in diagnosis and treatment. While calibration to observed clinical trial progression rates was an option, we found that the selected HER2+ controls from trials experienced better survival than expected, resulting in a model that under-predicted mortality vs. Canadian data. This highlights an important limitation of modelling single trial findings, which tends to produce results with limited generalizability to clinical practice. Another important limitation related to trial findings is the effect of substantial control patient crossover. [11] Better than expected survival in HER2+ control patients can be partly attributed to the effect

of crossover. This would tend to underestimate the effect of trastuzumab, further minimizing any potential gains simulated using a conservative model.

CONCLUSIONS

In summary, we demonstrated that initial FISH testing to direct targeted adjuvant trastuzumab could cost less and result in QALY gains, on average, over the predominant practice of initial IHC testing and FISH confirmation of IHC2+ results. We've shown that this conclusion is highly sensitive to assumptions about the underlying prevalence of HER2+ disease, test accuracy, adherence to testing and treatment guidelines, and duration of treatment benefit carryover. These caveats have important implications for decision makers as they can produce significant fluctuations in the incremental costs and gains associated with trastuzumab in the adjuvant setting. We propose that personalized medicine analyses should not be conducted outside the context of the associated diagnostic test, [40-42] and that the impact of potential testing and treatment practices be considered in the process of policy development. We also suggest further rigorous research to validate CISH or SISH against FISH in the clinical setting, as these tests could reduce the need for specialized equipment and lead to increased health gains through wider use of more accurate ISH testing.

LIST OF ABBREVIATIONS

BC	breast cancer
CEAC	cost-effectiveness acceptability curve
CHF	congestive heart failure
CISH	chromogenic *in situ* hybridization
DFS	disease free survival
DR	distant recurrence
FISH	fluorescence *in situ* hybridization
HER2	human epidermal growth factor receptor-2
HR	hazard ratio
ICUR	incremental cost-utility ratio
IHC	immunohistochemistry
LRR	locoregional recurrence
LVEF	left ventricular ejection fraction
MOHLTC	Ministry of Health and Long-Term Care
NED	no evidence of disease
NYHA	New York Heart Association
QALY	quality adjusted life year
SISH	chromogenic *in situ* hybridization
WTP	willingness to pay

COMPETING INTERESTS

IF held an unrestricted studentship funded by Pfizer Canada at the time this research was conducted. All other co-authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

IF contributed to the concept and design, acquisition of data, analysis and interpretation of data, writing, critical revision and finalization of this manuscript for important intellectual content. All other co-authors contributed to the concept, interpretation of data, critical revision and finalization of this manuscript for important intellectual content.

ACKNOWLEDGEMENTS

We thank Dr. Nicole Mittmann and associates of the HOPE Research Group (Sunnybrook Health Sciences) for their assistance in identifying a cohort of metastatic patients in Ontario. Dr. Craig Earle provided invaluable insight into documenting current testing patterns. We also thank Drs. Elena Elkin and Michael Hassett for their input on model development, and Drs. JC Cutz, Leela Elavathil and Wedad Hanna for their insight on HER2 testing in Ontario.

Part of this material is based on data and information provided by Cancer Care Ontario. However, the analysis, conclusions, opinions and statements expressed herein are those of the author(s) and not necessarily those of Cancer Care Ontario.

REFERENCES

1. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL: **Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene.** *Science* 1987, **235**(4785):177–82.

2. Slamon DJ, Godolphin W, Jones LA, et al: **Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer.** *Science* 1989, **244**(4905):707–12.

3. Pietras RJ, Arboleda J, Reese DM, et al. **HER-2 tyrosine kinase pathway targets estrogen receptor and promotes hormone-independent growth in human breast cancer cells**. *Oncogene* 199, **10**(12):2435–46.

4. Hancock MC, Langton BC, Chan T, et al. **A monoclonal antibody against the c-erbB-2 protein enhances the cytotoxicity of cis-diamminedichloroplatinum against human breast and ovarian tumor cell lines.** *Cancer Res* 1991, **51**(17):4575–80.

5. Joensuu H, Kellokumpu-Lehtinen P-L, Bono P, et al. **Adjuvant docetaxel or vinorelbine with or without trastuzumab for breast cancer**. *N Engl J Med* 2006, **354**(8):809–20.

6. Smith I, Procter M, Gelber RD, et al. **2-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer: a randomised controlled trial.** *Lancet* 2007, **369**(9555):29–36.

7. Gianni L, Eiermann W, Semiglazov V, et al. **Neoadjuvant chemotherapy with trastuzumab followed by adjuvant trastuzumab versus neoadjuvant chemotherapy alone, in patients with HER2-positive locally advanced breast cancer (the NOAH trial): a randomised controlled superiority trial with a parallel HER2-negative cohort**. *Lancet* 2010, **375**(9712):377–84.

8. Romond EH, Perez E a, Bryant J, et al. **Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer.** *N Engl J Med* 2005, **353**(16):1673–84.

9. Slamon D, Eiermann W, Robert N, et al. **Adjuvant trastuzumab in HER2-positive breast cancer**. *N Engl J Med* 2011, **365**(14):1273–83.

10. Bordeleau L (University of Toronto). **Herceptin FAQ**. *Mt Sinai Hospital Marvelle Koffler Breast Centre*. 2012. Available at: http://www.mountsinai.on.ca/care/mkbc/faqs/herceptin . Accessed September 9, 2013.

11. Moja L, Tagliabue L, Balduzzi S, et al. **Trastuzumab containing regimens for early breast cancer.** Cochrane Breast Cancer Group, ed. *Cochrane Database Syst Rev* 2012, **4**:CD006243.

12. Ontario Ministry of Health and Long-Term Care. **Funding of Herceptin ® for the treatment of breast cancer.** 2001:1. Available at: www.health.gov.on.ca/english/providers/pub/cancer/ann040301.html. Accessed December 11, 2012.

13. Ferrusi IL, Marshall DA, Kulin NA, Leighl NB, Phillips KA. **Looking back at 10 years of trastuzumab therapy: what is the role of HER2 testing? A systematic review of health economic analyses.** *Per Med* 2009, **6**(2):193–215.

14. Elkin EB, Weinstein MC, Winer EP, Kuntz KM, Schnitt SJ, Weeks JC. **HER-2 testing and trastuzumab therapy for metastatic breast cancer: a cost-effectiveness analysis**. *J Clin Oncol* 2004, **22**(5):854–63.

15. Lidgren M, Jönsson B, Rehnberg C, Willking N, Bergh J. **Cost-effectiveness of HER2 testing and 1-year adjuvant trastuzumab therapy for early breast cancer.** *Ann Oncol* 2008, **19**(3):487–95.

16. Blank PR, Schwenkglenks M, Moch H, Szucs TD. **Human epidermal growth factor receptor 2 expression in early breast cancer patients: a Swiss cost-effectiveness analysis of different predictive assay strategies.** *Breast Cancer Res Treat* 2010, **124**(2):497–507.

17. Hanna W, O'Malley FP, Barnes P, et al. **Updated recommendations from the Canadian National Consensus Meeting on HER2/neu testing in breast cancer**. *Curr Oncol* 2007, **14**(4):149–53.

18. Wolff AC, Hammond MEH, Schwartz JN, et al. **American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer.** *J Clin Oncol* 2007, **25**(1):118–45.

19. Snow S, Rayson D, Barnes P, Kara T, Younis T. **Does HER2/neu status affect adjuvant chemotherapy utilization? A population-based analysis.** Abstract 11549. In: *ASCO 2008 Breast Cancer Symposium*. Washington: American Society of Clinical Oncology (ASCO); 2008.

20. Ferrusi IL, Earle CC, Trudeau M, et al. **Closing the personalized medicine information gap: HER2 test documentation practice.** *Am J Manag Care* 2013, **19**(1):838–44.

21. Phillips KA, Liang S-Y, Van Bebber S. **Challenges to the translation of genomic information into clinical practice and health policy: Utilization, preferences and economic value**. *Curr Opin Mol Ther* 2008, **10**(3):260–6.

22. Haas JS, Phillips K a, Liang S-Y, et al. **Genomic testing and therapies for breast cancer in clinical practice.** *J Oncol Pract* 2011, **7**(3 Suppl):e1s–7s.

23. Magliocco A. **HER2 testing in Alberta** (personal communication). 2010.

24. Penault-Llorca F, Bilous M, Dowsett M, et al. **Emerging technologies for assessing HER2 amplification**. *Am J Clin Pathol* 2009, **132**(4):539–48.

25. Canadian Agency for Drugs and Technologies in Health. **Guidelines for the Economic Evaluation of Health Technologies: Canada**. 3rd ed. Ottawa: *CADTH*; 2006:1–46.

26. Trudeau M, Madarnas Y, McCready DR, Pritchard KI, Messersmith H. **The Role of Trastuzumab in Adjuvant and Neoadjuvant Therapy in Women with HER2 / neu-overexpressing Breast Cancer : A Clinical Practice Guideline**. 2007:4–8.

27. Hartung J, Knapp G, Sinha BK: **Methods of Combining Effect Sizes.** In *Statistical Meta-Analysis with Applications*. Chicester: Wiley; 2008.

28. Briggs AH, Sculpher M, Claxton K. **Making decision models probabilistic.** In *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press; 2006:77–120.

29. Levy a R, Briggs a H, Demers C, O'Brien BJ. **Cost-effectiveness of beta-blocker therapy with metoprolol or with carvedilol for treatment of heart failure in Canada.** *Am Heart J* 2001;**142**(3):537–43.

30. Miller G, Randolph S, Forkner E, Smith B, Galbreath AD. **Long-term cost-effectiveness of disease management in systolic heart failure**. *Med Decis Mak* 2009, **29**(3):325–33.

31. Mittmann N, Verma S, Koo M, Alloul K, Trudeau M. **Cost effectiveness of TAC versus FAC in adjuvant treatment of node-positive breast cancer**. *Curr Oncol* 2010, **17**(1):7–16.

32. Skedgel C, Rayson D, Dewar R, Younis T. **Cost-utility of adjuvant hormone therapies with aromatase inhibitors in post-menopausal women with breast cancer: upfront anastrozole, sequential tamoxifen-exemestane and extended tamoxifen-letrozole**. *Breast* 2007, **16**(3):252–61.

33. Ng D, Ferrusi I, Khong H, et al. **Cardiac monitoring during adjuvant trastuzumab therapy for breast cancer** (poster). In *San Antonio Breast Cancer Symposium* (SABCS). San Antonio; 2012.

34. Ferrusi IL, Leighl NB, Kulin N a, Marshall D a. **Do economic evaluations of targeted therapy provide support for decision makers?** *J Oncol Pract* 2011, **7**(3 Suppl):36s–45s.

35. Yang XR, Chang-Claude J, Goode EL, et al. **Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies**. *J Natl Cancer Inst* 2011, **103**(3):250–63.

36. Skedgel C, Rayson D, Younis T. **The cost-utility of sequential adjuvant trastuzumab in women with Her2/Neu-positive breast cancer: an analysis based on updated results from the HERA Trial**. *Value Heal* 2009, **12**(5):641–8.

37. Hedden L, O'Reilly S, Lohrisch C, et al. **Assessing the Real-World Cost-Effectiveness of Adjuvant Trastuzumab in HER-2/neu Positive Breast Cancer.** *Oncologist* 2012, **17**(2):164–171.

38 Hanna W, Barnes P, Berendt R, et al. **Testing for HER2 in breast cancer: current pathology challenges faced in Canada.** *Curr Oncol* 2012, **19**(6):315–23.

39. Sauter G, Lee J, Bartlett JMS, Slamon DJ, Press MF. **Guidelines for human epidermal growth factor receptor 2 testing: biologic and methodologic considerations.** *J Clin Oncol* 2009, **27**(8):1323–33.

40. Vogel CL, Cobleigh MA, Tripathy D, et al. **Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer.** *J Clin Oncol* 2002, **20**(3):719–26.

41. Hicks DG, Kulkarni S. **HER2+ breast cancer: review of biologic relevance and optimal use of diagnostic tools.** *Am J Clin Pathol* 2008, **129**(2):263–73.

42. Elkin EB, Marshall DA, Kulin NA, et al. **Economic evaluation of targeted cancer interventions: critical review and recommendations.** *Genet Med* 2011, **13**(10):853–60.

43. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). **Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials.** *Lancet* 2005,**365**(9472):1687–717.

44. Onitilo AA, Engel JM, Greenlee RT, Mukesh BN. **Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival**. *Clin Med Res* 2009, **7**(1-2):4–13.

45. Montagna E, Bagnardi V, Rotmensz N, et al. **Breast cancer subtypes and outcome after local and regional relapse.** *Ann Oncol* 2012, **23**(2):324–31.

46. Ontario Cancer Registry. 2012.

47. Karnon J, Vanni T. **Calibrating models in economic evaluation: a comparison of alternative measures of goodness of fit, parameter search strategies and convergence criteria.** *Pharmacoeconomics* 2011, **29**(1):51–62.

48. Statistics Canada. Non-stage specific survival counts for breast cancer cases (1998-2006). 2011.

49. Crump M, Trudeau M, Sinclair S. **The Role of Trastuzumab (Herceptin) in the Treatment of Women with HER2 / neu-overexpressing**

Metastatic Breast Cancer Practice Guideline Report # 1-15 (Version 2 . 2004). *Cancer* 2005;**15**.

50. Hanna WM. **HER2 test costs** (personal communication). 2012.

51. Drucker A, Skedgel C, Virik K, Rayson D, Sellon M, Younis T. **The cost burden of trastuzumab and bevacizumab therapy for solid tumours in Canada.** *Curr Oncol* 2008, **15**(3):136–42.

52. Ontario Ministry of Health and Long-Term Care. **Schedule of Benefits for Physician Services**. 2012.

53. Leighl NB. **Trastuzumab unit price & vial wastage** (personal communication). 2012.

54. Will BP, Berthelot JM, Le Petit C, Tomiak EM, Verma S, Evans WK. **Estimates of the lifetime costs of breast cancer treatment in Canada.** *Eur J Cancer* 2000, **36**(6):724–35.

55. Sorensen S, Brown R, Benedict A, Flood E, Revicki D. **QL4 Patient-Rated Utilities in Postmenopausal Early Breast Cancer (EBC):A Cross-Country Comparison**. *Value Heal* 2004, **7**(6):641–642.

TABLES

Table 1. Tabular representation of the up-front HER2 testing and treatment decision-tree; base-case (per guidelines) treatment practice is shown. IHC (FISH) represents the most commonly practiced testing pattern in Ontario and was therefore the referent strategy. Abbreviations: +: positive; CISH: chromogenic *in situ* hybridization; FISH: fluorescence *in situ* hybridization; IHC: immunohistochemistry; SISH: silver *in situ* hybridization.

Strategy	1st Test	2nd Test[i]	Trastuzumab Treatment[ii]
IHC (FISH)	IHC	FISH	IHC 3+ or FISH+
IHC (CISH)	IHC	CISH	IHC 3+ or CISH+
IHC (SISH)	IHC	SISH	IHC 3+ or SISH+
FISH	FISH	-	FISH+
CISH	CISH	-	CISH+
SISH	SISH	-	SISH+

i: a second HER2 test was only modeled for tumors that first tested equivocal by IHC in the base case. The use of a second test varied in analyses of current testing practice.
ii: trastuzumab treatment patterns varied by IHC result and ISH test result in analyses of current treatment practice.

Table 2. Probability (and standard error) of agreement across IHC, CISH, SISH and FISH. All test parameters fitted to beta or Dirichlet (IHC only) distributions. FISH was assumed the gold-standard test. The cut-off for positivity with any ISH test was ≥2.2 gene copy ratio per Canadian and ASCO guidelines. ALL CISH results represent pooled estimates across studies of single and dual CISH. Abbreviations: TN: true negative; TP: true positive; +: positive; - negative

	CISH +	SISH+	FISH +
IHC 3+ if TP	0.99 (0.001)	0.99 (0.001)	0.88 (0.03)
IHC 3+ if TN	0.25 (0.06)	0.13 (0.03)	-
IHC 2+ if TP	0.98 (0.005)	0.99 (0.002)	0.45 (0.02)
IHC 2+ if TN	0.06 (0.02)	0.62 (0.09)	-
IHC 0/1+ if TP	0.88 (0.03)	0.02 (0.005)	0.04 (0.02)
IHC 0/1+ if TN	0.003 (0.0001)	0.001 (0.0002)	-
CISH +	-	-	0.93 (0.001)
CISH -	-	-	0.01 (0.0004)
SISH +	-	-	0.93 (0.001)
SISH -	-	-	0.001 (0.0002)

Table 3. Model inputs for a cohort of breast cancer patients age 50y at diagnosis. All breast cancer parameters represent post-calibration estimates. All disease and clinical practice parameters were fit to beta distributions, while costs and utilities were fit to gamma distributions (unless otherwise noted). Abbreviations: BC: breast cancer; CISH: chromogenic *in situ* hybridization DC: discontinuation; FISH: fluorescence *in situ* hybridization; HER2: human epidermal growth factor receptor-2; IHC: immunohistochemistry; LRR: locoregional recurrence; NED: no evidence of disease; SE: standard error; SISH: silver *in situ* hybridization y: year

Input Parameter	Mean	SE or Range [i]	Assumptions & Source
DISEASE PARAMETERS			
HER2+ disease prevalence			[20]; assumed base case probability modeled deterministically
Base case	0.20	[0.13, 0.30]	
Ontario	0.13	0.0001	
Hazard ratio, trastuzumab treatment effect			[11]; Hazard ratio applied only to recurrence from NED in true HER2+ assigned to treatment; reduced by 50% in base case if discontinued (varied 0% to 100%); assumed 5y carryover effect in base case
Sequential (base case)	0.71	[0.53, 0.95]	
Concurrent & sequential	0.60	[0.50, 0.71]	
Discontinuation *(symptomatic cardiac toxicity)*			
Sequential (base case)	0.02	0.005	[11]
Concurrent & sequential	0.03	0.006	[11]
Current Ontario practice	0.06	0.06	data on file
Asymptomatic cardiac toxicity			
Sequential (base case)	0.17	0.02	[11]
Concurrent & sequential	0.17	0.02	[11]
Current Ontario practice	0.15	0.07	data on file
Cancer recurrence from NED or DC *(year 10 values carried forward)*			
y 1-10, HER2+	Cycle-specific (Appendix 3)		[11]
y 1-10, HER2-	Cycle-specific (Appendix 3)		[43]
Probability DR from NED or DC *(contingent on experiencing a recurrence)*			
HER2+	0.67	0.02	[11]; weighted probability of DR as DFS outcome in HER2+ controls

Input Parameter	Mean	SE or Range [i]	Assumptions & Source
HER2-	0.62	0.05	[44]; probability of DR as DFS outcome
Progression to DR from LRR	Cycle-specific (Appendix 4)		[45]
BC Mortality from DR state	Age- & cycle-specific (Appendix 5)		[46]; validated by the least squares method (Appendix 2)[4]
General mortality	Age-specific	-	[48]; modeled deterministically
TESTING & TREATMENT PRACTICE			
Probability of ISH test to confirm IHC results			
Guidelines: confirm IHC2+ only (base case)	1.00	-	[17]
Current Ontario practice: IHC 0/1+	0.04	0.003	data on file
Current Ontario practice: IHC 2+	0.88	0.01	data on file
Current Ontario practice: IHC 3+	0.06	0.001	data on file
Probability of trastuzumab usage given test results			
Guidelines: only HER2+ (base case)	1.00	-	[49]
Current Ontario practice: HER2-	0.005	0.001	data on file
Current Ontario practice: HER2equivocal	0.19	0.02	data on file
Current Ontario practice: HER2+	0.57	0.01	data on file
COSTS (2013 CAD$)			
HER2 testing			[50]; assumed fixed reimbursement costs; assumed CISH, SISH reimbursed similarly to FISH
IHC	$80	fixed	
FISH, SISH, CISH	$388	fixed	
Adjuvant trastuzumab, 1y TOTAL	$50,689	fixed	[31,51–53]; assumes no vial wastage; includes drug acquisition, administration, supportive medication & cardiac monitoring; applied in y1 to all patients assigned to treatment
Recurrence monitoring in NED			[54]; variance assumed
Y1	$610	±20%	

Input Parameter	Mean	SE or Range [i]	Assumptions & Source
Y2	$551	±20%	
Y3	$492	±20%	
Y4	$433	±20%	
Y5+	$374	±20%	
Early discontinuation			[52,54]; half of trastuzumab cost applied for patients discontinuing due to cardiac toxicity; extra diagnostic workup, monitoring and supportive medication for cardiac toxicity; variance assumed
Y1	$811	±20%	
Y2	$751	±20%	
Y3	$692	±20%	
Y4	$633	±20%	
Y5+	$574	±20%	
LRR			[52–54]; assumed retreatment with trastuzumab 1y; includes workup, surgery, hospitalization, radiation, chemotherapy & hormonal therapy; variance assumed
HER2+	$58,404	±20%	
HER2-	$7,715	±20%	
Post-LRR recurrence monitoring			[54]; variance assumed
Y1	$1,081	±20%	
Y2-4	$800	±20%	
Y5+	$650	±20%	
Distant recurrence			[52–54]; assumed retreatment with trastuzumab and cardiac monitoring; includes diagnostic procedures, surgery, radiation, chemotherapy, hormonal therapy, ongoing and terminal care; variance assumed
HER2+	$27,713.64	±20%	
HER2-	$17,115.57	±20%	
UTILITY WEIGHTS			
Well, NED, on trastuzumab	0.92	±20%	[32]
Well, NED	0.96	±20%	[55]

Input Parameter	Mean	SE or Range [i]	Assumptions & Source
DC due symptomatic cardiac toxicity, Y1	0.58	±20%	[30]
DC, Y2+	0.96	±20%	[31]; assumed
Asymptomatic cardiac toxicity	0.74	±20%	[30]; assumed
LRR	0.7	±20%	[32]
Well, post-LRR	0.96	±20%	assumed
DR	0.57	±20%	[67]

i: 95% confidence interval reported for treatment effect estimates only

Table 4. Results of the probabilistic base case analysis case and scenario analyses. Only IHC 2+ test results were confirmed with a second test in the base case. IHC (FISH) was the referent strategy. The base case assumed sequential therapy, a 5-year treatment effect duration, and 20% disease prevalence. Abbreviations: ICER: incremental cost-effectiveness ratio; ICUR: incremental cost-utility ratio; Incr: incremental; LY: life years; QALY: quality adjusted life year;

Strategy	Costs	LYs	QALYs	Incr. costs	Incr. LYs	Incr. QALYs	ICER	ICUR
BASE CASE (GUIDELINE ADHERENCE)								
IHC (FISH)	$44,308	11.61	10.60					
FISH	$43,493	11.62	10.60	-$815	0.0110	0.0022	-$74,272	-$374,766
CISH	$43,449	11.62	10.60	-$44	-0.0008	-0.0025	dominated	dominated
SISH	$44,794	11.62	10.59	$1,345	0.00254	-0.00888	dominated	dominated
IHC (CISH)	$44,765	11.61	10.57	-$4,835	-0.00150	-0.00749	dominated	dominated
IHC (SISH)	$49,600	11.61	10.58	$4,807	-0.01043	-0.00686	dominated	dominated
BASE CASE + PROBABILISTIC TREATMENT BENEFIT CARRYOVER								
IHC (FISH)	$43,571	11.60	10.57					
FISH	$42,861	11.61	10.60	-$709	0.01065	0.02564	-$66,632	-$27,664
SISH	$44,088	11.61	10.59	$1,227	-0.00426	-0.00446	dominated	dominated
CISH	$42,853	11.61	10.58	-$1,235	-0.01130	-0.01130	dominated	dominated
IHC (CISH)	$43,998	11.59	10.57	-$4,472	-0.00419	0.00197	dominated	-$2,264,179
IHC (SISH)	$48,469	11.60	10.57	$5,616	-0.00775	-0.01377	dominated	dominated
BASE CASE + CURRENT ONTARIO TESTING PRACTICE								
IHC (FISH)	$42,876	11.61	10.58					
FISH	$42,721	11.62	10.61	-$155	0.01137	0.03456	-$13,634	-$4,485
BASE CASE + CURRENT ONTARIO TREATMENT PRACTICE								
IHC (FISH)	$45,555	11.59	10.57					
FISH	$45,367	11.59	10.56	-$188	0.00706	-0.00819	-$26,630	$22,956
BASE CASE + ONTARIO HER2+ DISEASE PREVALENCE								

Strategy	Costs	LYs	QALYs	Incr. costs	Incr. LYs	Incr. QALYs	ICER	ICUR
IHC (FISH)	$38,878	11.75	10.74					
FISH	$37,723	11.75	10.73	-$1,155	0.00607	-0.01321	-$190,138	$87,400
ONTARIO TESTING, TREATMENT PRACTICE + HER2+ DISEASE PREVALENCE								
IHC (FISH)	$44,060	11.72	10.69					
FISH	$44,103	11.72	10.69	$43	0.00123	-0.00304	$34,679	dominated

FIGURES

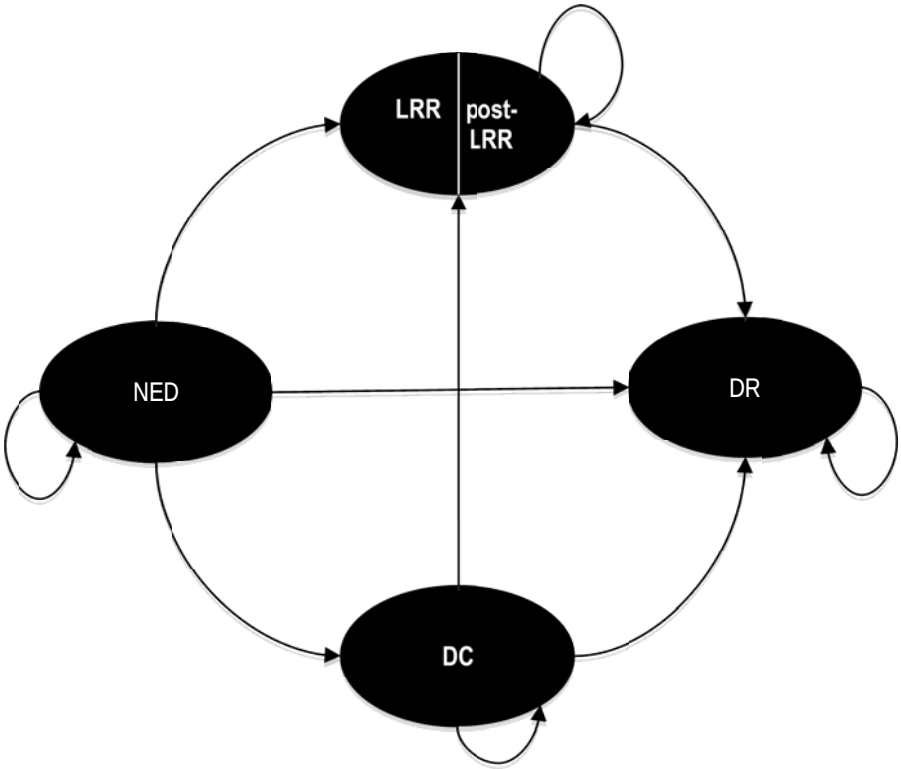


Figure 1. Markov natural history model of early-stage breast cancer. Patients entered the model in the no evidence of disease (NED) state after completing primary surgical and adjuvant cytotoxic chemotherapy. Costs and disutilities associated with trastuzumab were modeled in the first cycle of NED for patients assigned to therapy. We assumed that symptomatic cardiac toxicity due to trastuzumab therapy (New York Heart Association [NYHA] class III, IV congestive heart failure [CHF]) resulted in treatment discontinuation by 6 months (8 doses). Patients in the discontinuation (DC) state were also considered free of disease, but experienced a temporary disutility due to cardiac toxicity. Patients could remain disease-free or experience a recurrence, either locoregional or distant, from NED or DC. Those who experienced a local, regional or contralateral recurrence transitioned to the locoregional recurrence (LRR) state, where we assumed that standard surgery, radiation and chemotherapy were provided in the first and only cycle in that state. We assumed that all patients completed LRR therapy and transitioned directly into the post-LRR state where they remained disease-free or progressed to distant recurrence (DR). Patients could also transition directly to DR without experiencing LRR first. We assumed that all DR patients eventually experienced disease-specific mortality. Patients in the DR state received palliative treatment consisting of radiation or chemotherapy for each year in the cycle until death. Patients may die due to general mortality from any state.

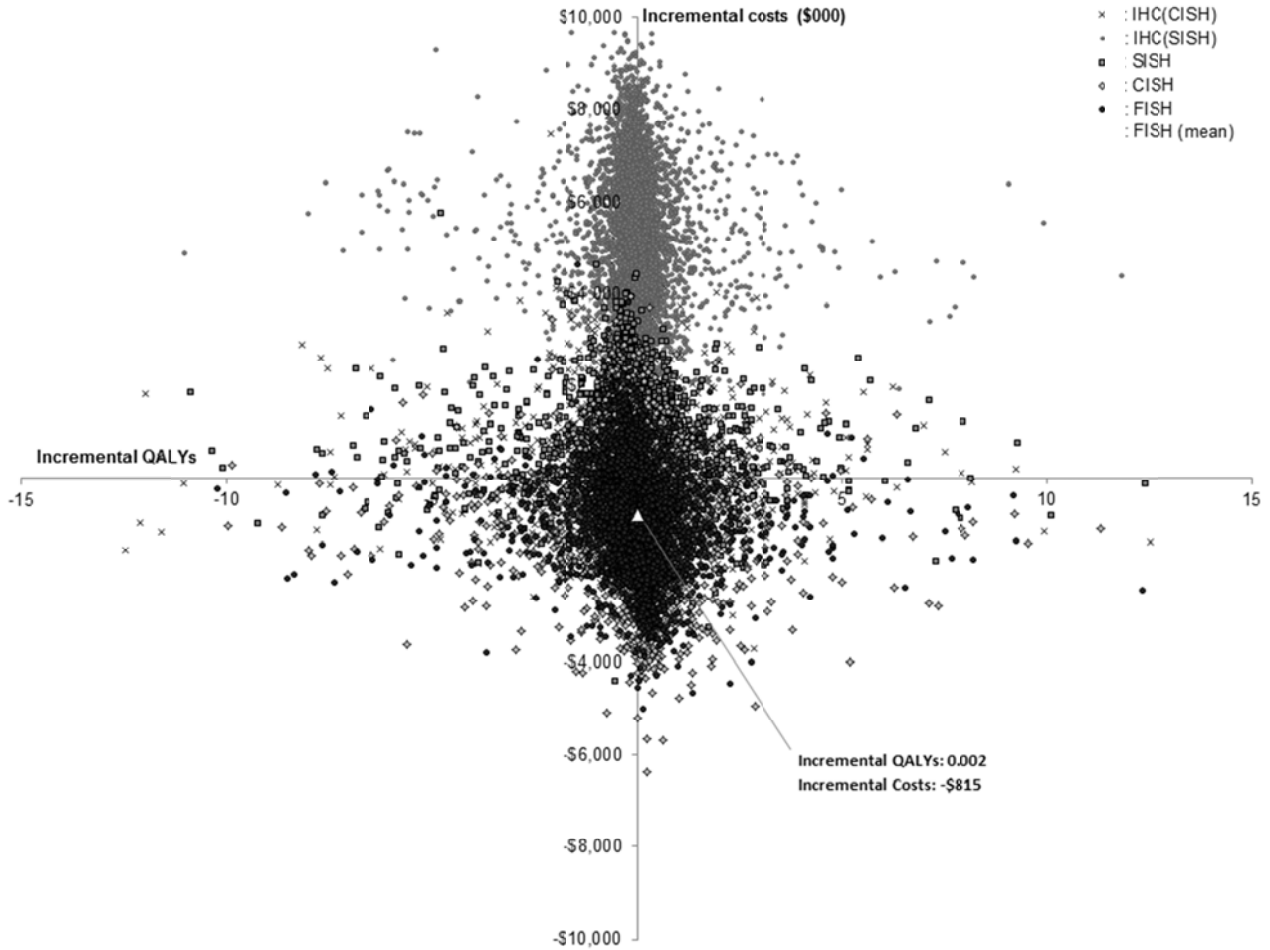


Figure 2. Results of the base case probabilistic analysis on the cost-effectiveness plane for alternative HER2 testing strategies vs. referent strategy IHC(FISH). All strategies illustrate the uncertainty in the relative cost-effectiveness of each strategy vs. the referent strategy. The triangle represents the mean of the probabilistic results for the only non-dominated strategy, FISH.

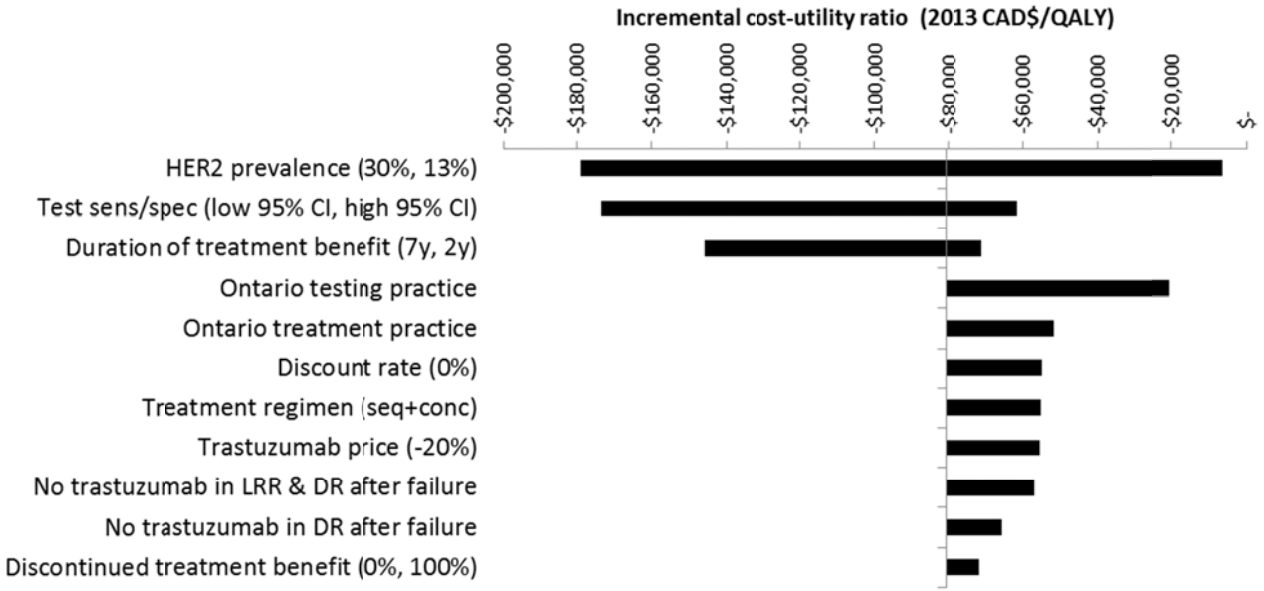


Figure 3. Tornado diagram illustrating model sensitivity to specific parameters and assumptions per univariate deterministic sensitivity analyses. Results are shown for the strategy of FISH testing alone vs. referent IHC(FISH). The value at which the vertical axis crosses the horizontal axis represents the univariate estimate for FISH alone in the base case. Ontario testing practice indicates that the observed rates of FISH confirmation were modeled, rather than assuming guideline adherence. Similarly, Ontario treatment practice indicates that the observed rates of trastuzumab utilization, contingent on HER2 test result, were modeled, rather than assuming guideline adherence.

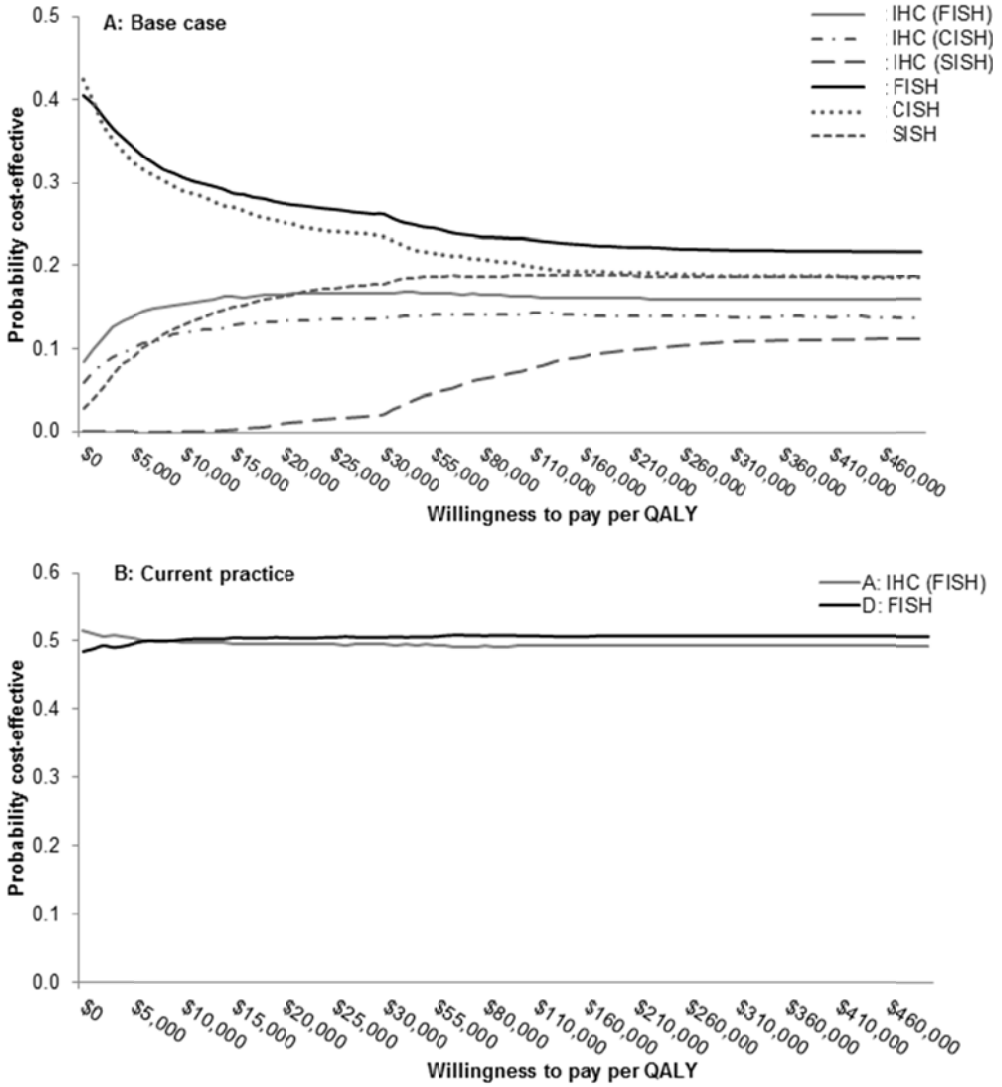


Figure 4. Cost-effectiveness acceptability curves with probabilistic treatment benefit duration under the base case (A) and current practice and local disease prevalence scenario (B). The duration of treatment effect was varied probabilistically in both cases.

4.2 SUPPLEMENTAL MATERIAL

Appendix 1: Systematic review & meta-analysis methods

We conducted a systematic review of the literature to determine the sensitivity and specificity of immunohistochemistry (IHC), chromogenic *in situ* hybridization (CISH), and silver *in situ* hybridization (SISH) tests relative to fluorescence *in situ* hybridization (FISH) as the gold standard test for detecting human epidermal growth factor receptor-1 (HER2) overexpression in breast cancer.

We developed MEDLINE and EMBASE search strategies (available upon request) that combined disease filters (breast cancer and HER2) with diagnostic test filters (IHC, SISH, FISH or CISH) and a diagnostic accuracy outcome filter. We then applied a primary human study filter under consultation with a librarian. All abstracts were reviewed by a single reviewer, while full text review was conducted in duplicate. Studies were included if they examined agreement between ≥2 HER2 tests (IHC, FISH, CISH or SISH only) including the gold standard (FISH). The latter requirement was necessary to inform the contingent probabilities of the decision tree. Only studies of test agreement in formalin-fixed, paraffin embedded invasive BC tissue samples were included. We also limited our analysis to studies that reported original research and those published in English language due to resource limitations. Finally, included studies were limited to those that reflected the most recent Canadian HER2 positivity cut-offs for IHC and ISH testing[8]:

(1) IHC positive (3+): strong, complete homogenous membrane staining in >30% of cells; and

(2) ISH positive: HER2:CEP17 ration >2.2 *or* average gene copy number >6

We extracted information positivity cut-offs used, ISH scoring system, reporting of uninterpretable results and agreement across tests. The outcome of interest was the proportion of results in agreement on HER2 status. We assumed that the proportion outcome had a binomial distribution. The probability of agreement was determined for the following comparisons:

(1) IHC agreement with FISH,

(2) CISH agreement with FISH,

(3) SISH agreement with FISH,

(4) IHC agreement with (CISH agreement with FISH), and

(5) IHC agreement with (SISH agreement with FISH).

Taking comparison (4) as an example, we calculated the probability that an IHC test result (positive, equivocal or negative) would be in agreement with the corresponding CISH result (positive or negative) for the same sample, and in turn, in agreement with the corresponding FISH result for the same sample (positive or negative). We treated any ISH equivocal as negative for calculation purposes, although this was reported infrequently. Some studies reported zero events, often the result of perfect agreement between the 2 tests in the study sample. The arcsin

transformation was performed prior to weighting to allow for the inclusion of studies with zero events.[9]

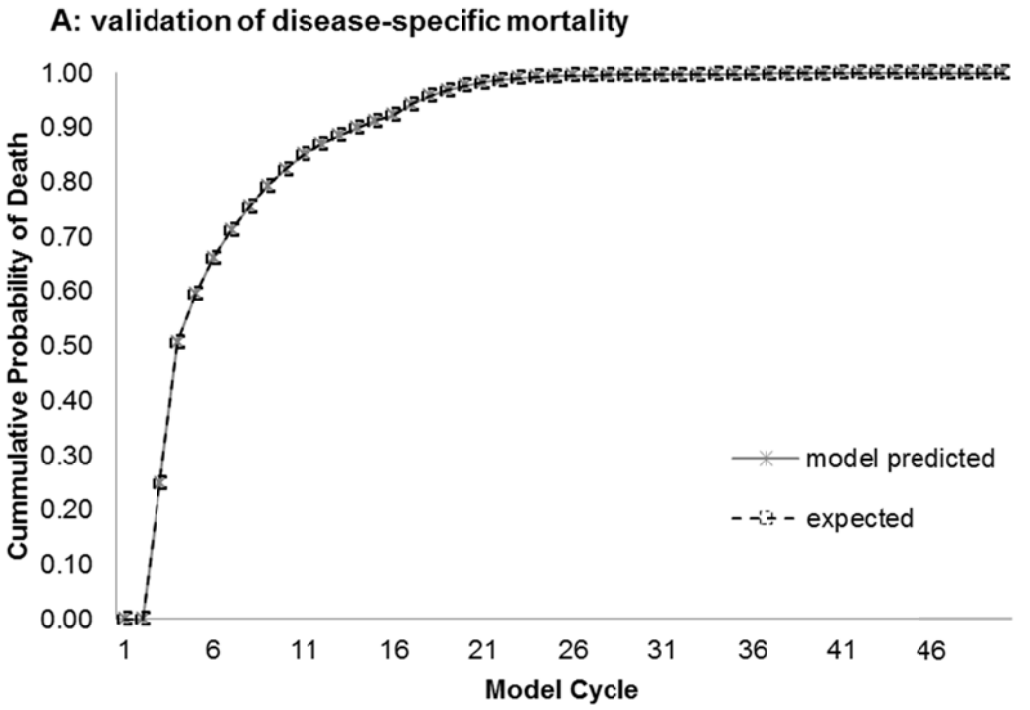
The agreement probabilities were determined for each study and then pooled to determine an overall estimate of agreement between tests. Data was synthesized across studies using the inverse variance method of weighting under the fixed effects model.[10] These probabilities were then used to inform the model. The weighted standard error (for each pooled estimate of agreement) was also determined, allowing us to fit the probabilities of agreement to beta distributions for modelling purposes. The probabilities for IHC results were fit to a Dirichlet distribution to ensure that the probabilities of each possible test result (positive, equivocal or negative) did not exceed 1 when modeled stochastically.

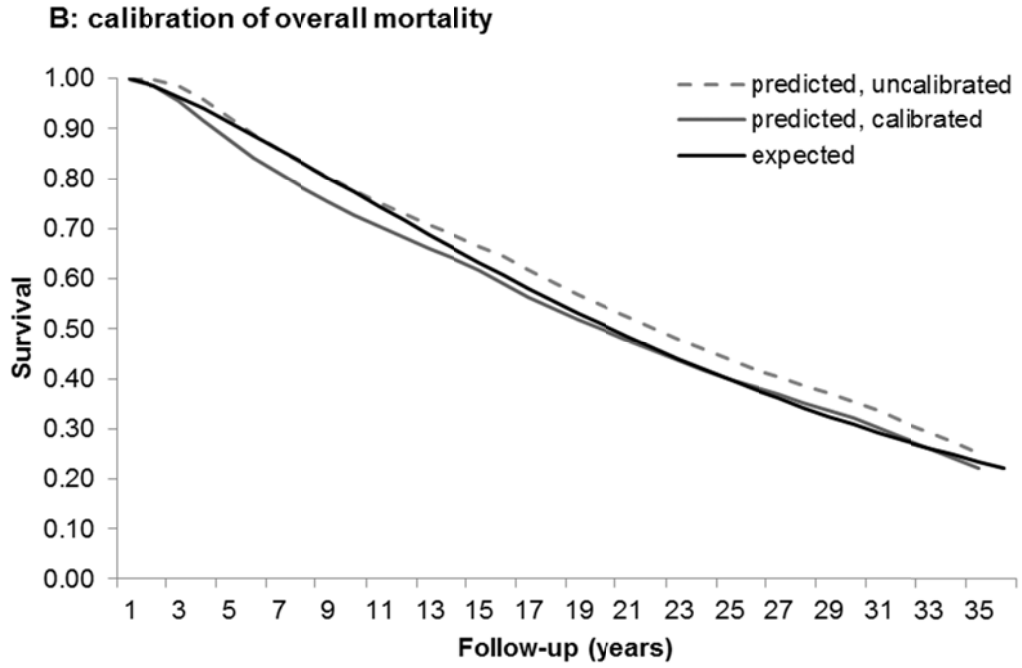
Appendix 2: Model Validation & Calibration

The Markov model was validated and calibrated by the least-squares method7to estimate the goodness-of-fit of model predictions against expectations. We validated disease-specific mortality predictions at various ages.[11–13] The model predicted mortality in close agreement with registry data. We also calibrated the model to fine-tune several disease progression parameters that were not derived from a Canadian population. Model-predicted all-cause mortality was compared to expected mortality from on an external dataset reporting the overall mortality for a Canadian cohort diagnosed with BC at 50-55y.[14] We optimized the probabilities of recurrence from no evidence of disease (NED), probability of first recurrence being distant, and the probabilities of distant recurrence (DR) after locoregional recurrence (LRR) in a step-wise fashion. Each parameter was varied probabilistically over 1,000 simulations. The parameter sets associated with the minimum goodness of fit estimate over all cycles was selected as the optimal parameter set and used for all subsequent simulations (Appendix 3 to Appendix 5).

Prior to calibration, the model tended to under-predict overall mortality. This was attributed to lower than expected recurrence rates in HER2+ controls in the trastuzumab trials, likely due to the selected trial populations and significant proportion of crossover to active treatment.[15] Through the calibration process we adjusted disease recurrence for HER2+ patients, probabilities of DR, and progres-

sion from post-LRR to DR. The resulting model tended to over predict mortality vs. expectations, but reduced error from 14% to 8%(Appendix 2b).


A: validation of disease-specific mortality
Cummulative Probability of Death
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
1
6
11
16
21
26
31
36
41
46
Model Cycle
model predicted
expected


B: calibration of overall mortality
Survival
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31
33
35
Follow-up (years)
predicted, uncalibrated
predicted, calibrated
expected

Appendix 2. Validation and calibration of the BC natural history model. Panel A shows disease-specific mortality resulting from progression to metastatic disease for a cohort of women diagnosed with BC at 50y. Panel B shows modeled predicted overall survival vs. expected survival in a cohort of Canadian women diagnosed with BC at 50y. Expected survival is based on data from Statistics Canada. The observed dataset represents the set of recurrence parameters producing the closest fit to expected survival derived through model calibration. The probabilities of recurrence, distant recurrence, and progression from local to distant recurrence were the only parameters that varied in the calibration process.

Appendix 3 – Appendix 5:Cycle- and age-specific disease parameters

Appendix3. Conditional annual probabilities of recurrence (local or distant) from no evidence of disease (NED) state for HER2+ and HER2- patients age 50 at diagnosis. Uncalibrated estimates represent conditional annual probabilities as calculated from the literature. Only the calibrated estimates were used to estimate incremental cost-utility ratios. Estimates shown for cycle 10 were used for all remaining cycles. All probabilities (μ) were fit to a beta distribution with the standard error (SE) shown.

	HER2+				HER2-			
	Uncalibrated [16–20]		Calibrated		Uncalibrated[21]		Calibrated	
cycle	μ	SE	μ	SE	μ	SE	μ	SE
1	0.0746	0.0113	0.1053	0.0113	0.0520	0.0011	0.0516	0.0011
2	0.1205	0.0149	0.1678	0.0149	0.0837	0.0014	0.0823	0.0014
3	0.0796	0.0144	0.1579	0.0144	0.0791	0.0014	0.0794	0.0014
4	0.0667	0.0150	0.1421	0.0150	0.0702	0.0013	0.0698	0.0013
5	0.0407	0.0227	0.1018	0.0227	0.0507	0.0016	0.0536	0.0016
6	0.0278	0.0070	0.0957	0.0070	0.0481	0.0015	0.0469	0.0015
7	0.0514	0.0155	0.0984	0.0155	0.0489	0.0016	0.0479	0.0016
8	0.0514	0.0155	0.1021	0.0155	0.0512	0.0016	0.0511	0.0016
9	0.0514	0.0155	0.0857	0.0155	0.0376	0.0014	0.0425	0.0014
10+	0.0514	0.0155	0.0814	0.0155	0.0422	0.0028	0.0429	0.0028

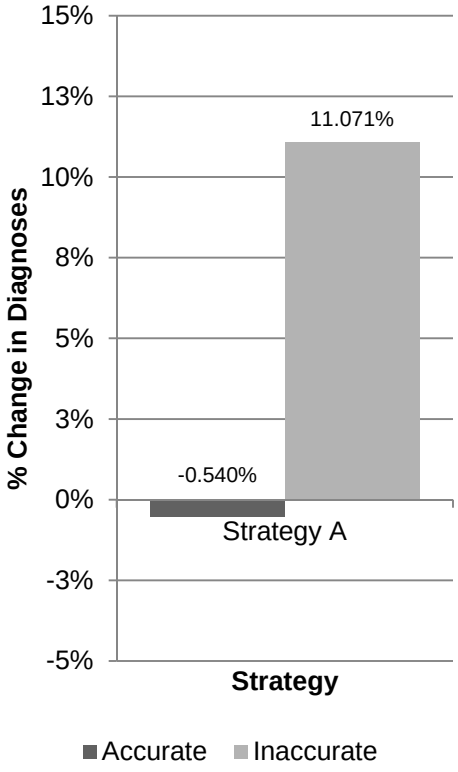
Appendix 4. Conditional annual probabilities of distant recurrence (DR) from the locoregional recurrence (LRR) state for HER2+ and HER2- patients. Uncalibrated estimates represent conditional annual probabilities as derived from the literature. Only the calibrated estimates were used to estimate incremental cost-effectiveness ratios. Estimates shown for cycle 10 were used for all remaining cycles. All probabilities (μ) were fit to a beta distribution with the standard error (SE) shown.

	All Uncalibrated[22]		**HER2+ Calibrated**		**HER2- Calibrated**	
cycle	μ	SE	μ	SE	μ	SE
1	0.2249	0.0250	0.2258	0.0250	0.2164	0.0250
2	0.2155	0.0282	0.2145	0.0282	0.2127	0.0282
3	0.1612	0.0284	0.1632	0.0284	0.1568	0.0284
4	0.1485	0.0298	0.1491	0.0298	0.1494	0.0298
5	0.0154	0.0113	0.0784	0.0113	0.0785	0.0113
6	0.0781	0.0304	0.0152	0.0304	0.0150	0.0304
7	0.0056	0.0117	0.0055	0.0117	0.0055	0.0117
8	0.0056	0.0117	0.0057	0.0117	0.0055	0.0117
9	0.0056	0.0117	0.0054	0.0117	0.0055	0.0117
10+	0.0056	0.0117	0.0058	0.0117	0.0058	0.0117

Appendix 5. Age-specific, conditional annual probabilities of death due to distant disease (stage IV, metastatic breast cancer) as derived from the Ontario Cancer Registry.[23] The same probabilities were used for HER2+ and HER2- patients, and were not calibrated. Estimates shown for cycle 5 were used for all remaining cycles.

	Age at Distant Recurrence							
cycle	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85+
1	0.2500	0.3352	0.2839	0.3309	0.4578	0.5068	0.5152	0.5581
2	0.3431	0.2105	0.2613	0.2258	0.3444	0.3425	0.3125	0.4167
3	0.1778	0.2778	0.1951	0.2113	0.1724	0.1667	0.2045	0.3333
4	0.1667	0.1111	0.1522	0.1163	0.1818	0.2143	0.1579	0.1250
5+	0.0833	0.1500	0.1250	0.2609	0.2667	0.0010	0.1667	0.5000

Appendix 6: Changes in diagnostic accuracy under current testing practice



Strategy	A: IHC (FISH)
Accurate diagnoses	
Base case	95.4%
Current testing	94.8%
Inaccurate diagnoses	
Base case	4.6%
Current testing	5.2%

Appendix 6. Percentage change (relative) in accurate and inaccurate diagnoses with current testing practice vs. guideline recommendations. results shown represent the average of probabilistic analysis results.

4.3 SUPPLEMENTAL METHODS

4.3.1 MODEL SPECIFICATIONS & ANALYSIS

4.3.1.1 Cohort, Analysis & Perspective

The model base case was a cohort of post-menopausal women, aged 50 years at incident diagnosis of early-stage BC. Analyses took the perspective of the public payer in Ontario (Ministry of Health and Long-Term Care). We assumed a cohort of patients similar the Ontario BC population across molecular subtypes and stage at diagnosis. The effectiveness of each test-treat strategy was estimated as life years (LYs) and quality-adjusted life years (QALYs). We calculated the incremental costs per LY and QALY gained for any strategy that was not dominated as follows:

Equation 1:

$$\boldsymbol{Incremental\ Cost\ Effectiveness\ Ratio} = \frac{\boldsymbol{C_2 - C_1}}{\boldsymbol{E_2 - E_1}}$$

where C_2 and E_2 represent the costs and effects associated with the alternative strategy and C_1 and E_1 represent the current or reference strategy. All future costs and effects were discounted at 5% per year in the base case as per Canadian economic evaluation guidelines [24,7] with rates of 3% and 0% considered in sensitivity analysis. We used a decision-analytic model to represent the sequelae of alternative HER2-targeted test-treat strategies over the cohort's lifetime (50 years). A decision tree represented alternative strategies to test for HER2 status and assign patients to trastuzumab therapy contingent on test results. A Markov model then

estimated the long-term consequences of test results and treatment. Each decision tree and Markov model parameter was programmed stochastically, with beta distributions fit to all transition probabilities unless otherwise noted.

4.3.1.2 Decision Tree: HER2 Prevalence, Testing & Treatment Assignment

A decision tree was used to represent the 6 alternative test-treat strategies:

A. IHC (FISH): Initial IHC with FISH confirmation of IHC 2+, treatment for IHC 3+ or FISH+;
B. IHC (CISH): Initial IHC with CISH confirmation of IHC 2+, treatment for IHC 3+ or CISH+;
C. IHC (SISH): Initial IHC with SISH confirmation of IHC 2+, treatment for IHC 3+ or SISH+;
D. FISH: Initial FISH testing, treatment for FISH+;
E. CISH: Initial CISH testing, treatment for CISH+; and
F. SISH: Initial SISH testing, treatment for SISH+.

In the base case we considered testing strategies that were consistent with Canadian testing guidelines.[8] We modeled test sensitivity and specificity accounting for HER2+ disease prevalence. Patients were identified as being either true positive, false positive, true negative or false negative for HER2 status within the decision tree. We introduced additional complexity in sensitivity analyses by allowing testing behaviour to vary from guideline recommendations. The model accommodated alternative patterns of confirmatory testing contingent on the first

test result. This allowed us to account for current practice patterns, (Chapter 3) which demonstrated inappropriate testing of some IHC- and IHC+ patients.

We assumed that a fixed 20% of BC patients were HER2+ in the base case and examined alternative rates of HER2 + disease prevalence in probabilistic and deterministic sensitivity analyses. A prevalence rate ranging from 20%-30% is widely cited in the literature. Therefore, we also examined the implications of HER2+ disease at a fixed prevalence of 30%.[9,25] However, more recent population studies suggest that actual prevalence may be lower.[26] We addressed this uncertainty by modelling HER2 prevalence probabilistically in sensitivity analysis according to the 13.4% prevalence estimate from our Ontario study (Chapter 3).

The decision tree also accounted for treatment behaviour contingent on HER2 test results. In the base case we assumed that all patients with a positive test result would receive trastuzumab plus standard chemotherapy according to CCO guidelines.[27] In probabilistic sensitivity analyses, we allowed treatment behaviour to vary contingent on test results. This also allowed us to account for the large proportion of HER2+ patients that did not receive trastuzumab treatment in Ontario practice (Chapter 3).

The various test-treat strategies resulted in the following patient groups distinguished by their true HER2 status and treatment:

(1) Base case:

i. True HER2+ patients treated with trastuzumab plus standard chemotherapy;
ii. False HER2+ patients treated with trastuzumab plus standard chemotherapy;
iii. True HER2- patients treated with standard chemotherapy alone; and
iv. False HER2- patients treated with standard chemotherapy alone.

(2) Sensitivity analyses, in addition to the groups above:

i. True HER2+ patients treated with standard chemotherapy alone;
ii. False HER2+ patients treated with standard chemotherapy alone;
iii. True HER2- patients treated with trastuzumab plus standard chemotherapy; and
iv. False HER2- patients treated with trastuzumab plus standard chemotherapy.

The lifetime costs and consequences of treatment contingent on the patient's underlying HER2 status were then modeled with a Markov model of the natural history of early-stage BC in Canada.

4.3.1.3 BC Natural History Markov Model

A Markov model with a 1 year cycle length approximated the natural history of early-stage BC and its economic consequences in Canada. Patients transi-

tioned between the health states at the half cycle. The model was comprised of 6 health states:

(1) No evidence of disease (NED)

(2) Early therapy discontinuation (DC)

(3) Locoregional recurrence (LRR)

(4) Post-locoregional recurrence (PLRR)

(5) Distant recurrence (DR)

(6) Death

A diagram of the BC natural history model is provided in Paper 4 Figure 1. The model accommodated differential rates of disease progression from the NED state according to true HER2 status. We defined disease progression as any recurrence from the NED or PLRR states. Death due to general mortality was possible from any state.

All cohort patients entered the NED state from the decision tree. We assumed all patients were free of disease after completing primary surgical treatment (mastectomy or lumpectomy) with or without radiotherapy. We also assumed that all patients had completed standard chemotherapy upon entering the NED state. Those assigned to trastuzumab therapy received treatment in the first cycle of the NED state only. Patients could experience a recurrence [p(recur|NED)] from the NED state, either locoregional or distant [p(DR|recur)] in nature. The probability of having a recurrence depended on HER2 status and time

spent in the NED state for the first 10 years. This permitted us to more closely model peaks in the hazard of disease progression and its rate of decline over time, as first demonstrated by Saphner *et al*.[28]

The DC state accommodated patients who discontinued trastuzumab treatment early. We assumed that these patients completed half of a normal course of trastuzumab (9 treatments). Patients could discontinue due to cardiac toxicity as a complication of therapy, or for other reasons. Only patients who experienced symptomatic cardiac toxicity could discontinue in the base case. We defined symptomatic cardiac toxicity as New York Heart Association (NYHA) class III or IV heart failure, consistent with the adjuvant trials.[15] Recurrences from DC were modeled similarly to the NED state. We assumed no mortality due to symptomatic cardiac toxicity from the DC state.[15]

1313 The LRR state is a 1 year tunnel state for patients who experience a local, regional or contralateral BC recurrence. Patients remained in the LRR state while undergoing surgical treatment, radiotherapy or chemotherapy. After 1 year, patients could return to disease-free status in the PLRR state, or progress to DR [p(DR|LRR)]. Patients who experienced both symptomatic cardiac toxicity and LRR were preferentially assigned to the LRR state as this state was associated with greater morbidity, mortality, costs and disutility. Mortality due to LRR was not modeled.

Patients transitioned directly into the PLRR state from LRR unless they experienced a distant recurrence or general mortality. Transitions from PLRR to DR [p(DR|LRR)] were dependent on HER2 status and time spent in the state. We assumed that the risk of recurrence after LRR was not constant, consistent with assumptions about recurrence from NED. The DR state accommodated patients whose disease had recurred and spread to distant tissues (metastatic, Stage IV). Patients could transition to DR from NED, DC or PLRR. The DR state is the only state from which patients could experience BC disease-specific mortality, consistent with current clinical understanding of the disease. [14,29]

4.3.2 MODEL INPUTS

A summary of all model inputs is provided in Appendix 4.3 through Appendix 4.6 and in Paper 4 Table 3.The selection of model inputs is described below. Subsequent calibration of key progression parameters is described in section 4.3.3 Model Validation & Calibration.

4.3.2.1 HER2 Test Properties

We conducted a systematic review of the literature to determine the sensitivity and specificity of IHC, CISH and SISH tests relative to FISH as the gold standard test. We also reviewed the concordance between IHC, CISH and SISH to inform the decision tree. Systematic reviews to date have focused on the sensitivity of IHC relative to FISH alone.[30,31] Our review focused on 2 research questions:

(1) What is the concordance of CISH or SISH testing techniques with IHC for detecting HER2 overexpression in BC patients?

(2) What is the sensitivity and specificity of IHC, CISH and SISH testing techniques compared with gold standard FISH for detecting HER2 over-expression in BC patients?

We did not limit our research questions to female BC patients as a preliminary literature scan indicated that study population gender was rarely reported.

We developed MEDLINE (Appendix 4.1) and EMBASE (Appendix 4.2) search strategies that combined disease filters (breast cancer and HER2) with diagnostic test filters (IHC, SISH, FISH or CISH) and a diagnostic accuracy outcome filter. We then applied a primary human study filter in consultation with a medical librarian. All abstracts were reviewed by a single reviewer, while full text review was conducted in duplicate. Individual forms for abstract review and full text review were developed and pilot tested with samples of 20 and 10 articles, respectively.

Inclusion/ Exclusion Criteria

Studies could be either included or excluded at abstract review, but we permitted reviewers to indicate uncertainty at full text review. Any disagreement at full text review was resolved through discussion and subsequent consultation (IF) if necessary. Studies were included if they examined agreement between ≥2 HER2 tests (IHC, FISH, CISH or SISH only) including the gold standard (FISH).

The latter requirement was necessary to inform the contingent probabilities of the decision tree. Only studies of test agreement in formalin-fixed, paraffin embedded invasive BC tissue samples were included. Tissue fixation was an important criterion as guideline recommendations advocate testing only on breast tissue samples fixed in formalin and embedded in paraffin.[8,32] We further limited the study sample to invasive disease as trastuzumab is recommended to treat invasive tumours ≥1 cm. Moreover, the morphology of non-invasive tumour tissue is different and the validation of HER2 testing in such samples is not well established.[33] We also limited our analysis to studies that reported original research and those published in English language due to resource limitations.[18] Finally, included studies were limited to those that reflected the most recent Canadian HER2 positivity cut-offs for IHC and ISH testing[8]:

(3) IHC+ (3+): strong, complete homogenous membrane staining in >30% of cells; and

(4) ISH+: HER2:CEP17 ration >2.2 *or* average gene copy number >6

A number of exclusion criteria were implemented to ensure a uniform sample of studies evaluating test accuracy and concordance under conditions similar to Canadian practice. We excluded studies that:

(1) examined the primary outcome using different threshold or cut-off values for HER2 positivity;

(2) examined concordance between ≥2 laboratory settings (e.g. regional hospital vs. cancer centre);

(3) examined interobserver agreement (either between humans or automated platforms);

(4) focused on selected sample populations (e.g. exclusive use of ISH to clarify IHC2+ or to confirm IHC3+);

(5) focused on a CEP17 polysomy population; and

(6) examined agreement between ≥2 tests as a secondary research question (e.g., studies of prognostic factors related to survival wherein HER2 was assessed by 2 methods).

A Microsoft Access database was designed to facilitate duplicate full text review and data extraction. We extracted information about study location (country, type of laboratory), perspective, sample (period of recruitment, age, sex, size of sample), tissue sample type, positivity cut-offs used, ISH scoring system, reporting of interpretable results and agreement across tests. The full text review and data extraction forms captured select aspects of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) study reporting quality tool, specifically description of selection criteria, blinded assessment, verification of whole sample and reporting of interpretable results.[34] The QUADAS tool is a widely-used and validated series of questions intended to assess the quality of patient selection,

index test, reference standard and patient flow domains of a primary diagnostic study.

Outcomes

The outcome of interest was the proportion of results in agreement on HER2 status assuming FISH as the gold standard test. We assumed that the proportion outcome had a binomial distribution consistent with the binary nature of the agreement outcome. The probability of agreement was determined for the following comparisons:

(1) IHC agreement with FISH,

(2) CISH agreement with FISH,

(3) SISH agreement with FISH,

(4) IHC agreement with (CISH agreement with FISH), and

(5) IHC agreement with (SISH agreement with FISH).

Taking comparison (4) as an example, we calculated the probability that an IHC test result (positive, equivocal or negative) would be in agreement with the corresponding CISH result (positive or negative) for the same sample, and in turn, in agreement with the corresponding FISH result for the same sample (positive or negative). We treated any ISH equivocal as negative for calculation purposes, although this was reported infrequently. Some studies reported zero events, often the result of perfect agreement between the 2 tests in the study sample. The arcsin

transformation was performed prior to weighting as this transformation allowed for the inclusion of studies with zero events.[9]

The agreement probabilities were determined for each study and then pooled to determine an overall estimate of agreement between tests. Data were synthesized across studies using the inverse variance method of weighting.[10] This method is used to combine two or more random variables (studies), whereby the weight given to the random variable is inversely proportional to the variable's variance. The weighting for individual studies was determined as:

Equation 2:

$$\boldsymbol{Individual\ Study\ Weight} = \frac{\mathbf{1}}{\boldsymbol{\sigma_i^2}} \quad [10]$$

whereσ^2 is the variance of the agreement probability of the i^{th} study. The pooled overall probability of agreement was then determined as:

Equation 3:

$$\boldsymbol{Pooled\ \theta} = \frac{\sum_i \boldsymbol{\theta_i}/\boldsymbol{\sigma_i^2}}{\sum_i \mathbf{1}/\boldsymbol{\sigma_i^2}} \quad [10]$$

where θ_i is the probability of agreement of the i^{th} study. These equations reflect the fixed effects variance model. We assumed the fixed effects model as the variability in study settings and patient populations identified in the review was reflective of the ethnic diversity and variability in practice throughout Ontario. These probabilities were then used to inform the model. The weighted standard error (for each pooled estimate of agreement) was also determined, allowing us to fit the probabilities of agreement to beta distributions for modelling purposes.

The probabilities for IHC results were fit to a Dirichlet distribution to ensure that the sum of the probabilities of each possible test result (positive, equivocal or negative) did not exceed 1 when modeled stochastically.[35]

4.3.2.2 Testing & Treatment Behaviour

We modeled alternative scenarios with varying degrees of adherence with HER2 testing and trastuzumab treatment guidelines. In the base case, we modeled perfect adherence to testing and treatment guidelines. In other words, *only* IHC2+ test results were verified with a second CISH, SISH or FISH test, and *only* HER2+ test results (IHC3+ or FISH+) received trastuzumab. The base case also assumed that testing and treatment behaviours were applied uniformly, meaning that *all* IHC2+ results were retested, and *all* patients with a HER2+ test result were treated with trastuzumab.

Scenario analyses were used to investigate the impact of deviations from guidelines documented in Ontario practice (Chapter 3). The first scenario analysis modeled current testing practice patterns while assuming adherence to treatment guidelines. Under scenario 1, the probability of inappropriate retesting of IHC- and IHC+ patients was fit to a beta distribution and modeled probabilistically for all testing strategies including the IHC test. We also modeled (Scenario 2) the observed probability of retesting of IHC equivocal (2+) results. Under the second scenario, we assumed that all patients were tested according to guidelines, but that treatment based on test results reflected current Ontario practice. There-

fore, some HER2- patients received trastuzumab while some HER2+ patients did not. Finally, under the scenario 3, we modeled testing and treatment practice to examine the joint impact of current practice patterns. The probabilities corresponding to current testing and treatment practice are provided in Chapter 3, Tables 3.2 and 3.3, respectively.

4.3.2.3 Disease Progression

Disease progression is significantly different according to HER2 status [36–38] and it was therefore crucial to derive estimates of progression according to this molecular subtype. This feature distinguishes the current model from many other evaluations of adjuvant trastuzumab. The majority of models considered a HER2+ cohort exclusively.[5,39–41] The current analysis required additional information about the probability of disease progression in HER2- patients to properly characterise the natural history of early-stage BC in true negative and false positive patients tested for HER2 status. Each model input for disease progression was identified through a targeted PubMed literature search which prioritised potential sources according to the hierarchy of evidence[42], with estimates from a systematic review and meta-analysis receiving top priority. We also prioritised studies that analysed outcomes using time to event analysis, given that this approach accounts for both the occurrence of events as well as the period of time that each patient remained event-free. Time to event analyses can also account for the sur-

vival time of censored patients who are lost to follow-up.[43], given that this approach accounts

In all previously published cost-effectiveness analyses of trastuzumab, disease recurrence was modeled from a single clinical trial.[5,39–41] We considered the OCR and Canadian Cancer Registry as potential data sources for Canadian patterns of disease recurrence and progression. However neither registry documented recurrences or HER2 status. We also considered the Early Breast Cancer Trialists' Collaborative Group (EBCTCG) ongoing meta-analyses of the effects of chemotherapy on BC recurrence and survival.[21,44] These reviews document the survival experience of over 140,000 women with 15 years of follow-up, but do not capture HER2 status. We therefore estimated the probability of disease recurrence based on the studies included in the most recently published meta-analysis of trastuzumab in early-stage BC.[15] This approach provided consistency with the treatment effect estimate in the model. The 4,971 control patients captured in this meta-analysis had a median age of 49 years and had LN+ or high risk LN- disease. All control patients had prior surgical resection of their tumours (≥ 1 cm) except for those enrolled in 2 neoadjuvant trials (n= 135, 11.3% overall weighting). Finally, all controls had normal heart function. Further discussion of the patients included in the meta-analysis is provided in section I4.3.2.5 Treatment-Related Parameters.

We extracted data from the control arms of all published adjuvant and neoadjuvant trastuzumab trials per the Cochrane review to establish a more robust estimate of $p(recur|NED)_{HER2+}$ from a larger sample population. Recurrence rates in these patients provided an estimate of how HER2+ disease progresses in the presence of standard adjuvant chemotherapy without trastuzumab. This allowed us to model the consequences of undertreatment due to false negative test results. The rates of disease recurrence were applied to all true HER2+ patients irrespective of test result. The probability of disease recurrence in HER2+ patients was derived from the disease-free survival (DFS) outcome. This is a composite outcome that measures the time to disease recurrence or death, whichever event comes first, and therefore includes local or distant recurrences. We also preferred the DFS outcome because it was analysed by time-to-event analysis. We extracted the cumulative probability of DFS, and the number of patients at risk, in HER2+ control patients annually from study start to the end of follow-up as reported for each adjuvant trastuzumab trial. Confidence intervals could not be extracted as they were not reported. Software (GetData Graph Digitizer v2.25) was used to digitally extract the probability of DFS at each annual interval from published Kaplan-Meier curves where the exact probability was not explicitly reported. For example, the Breast Cancer International Research Group (BCIRG)-006 trial explicitly reported the probability of DFS for the AC-T control arm at 2, 3, 4 and 5 years follow-up.[18] Therefore, we digitally extracted the approximate probabilities

of DFS from the published Kaplan-Meier graph at 1, 6 and 7 years follow-up. We then calculated the cumulative annual probability of recurrence [p(recur|NED)] as follows:

Equation 4:

$$\boldsymbol{Cummulative\ p(recur|NED)_t = 1 - p(DFS)_t}$$

where *t* represents time, and $p(DFS)_t$ represents the cumulative probability of being DFS at *t* as extracted from Kaplan-Meier curves for HER2+ controls. Then, each yearly cumulative probability of recurrence was converted to an annual probability of recurrence conditional on surviving to the start of the year as follows:

Equation 5:

$$\boldsymbol{Conditional\ ann.\ p(recur|NED)_t = 1 - \frac{1 - cumulative\ p(recur|NED)_t}{1 - cumulative\ p(recur|NED)_{t-1}}}$$

The conditional annual probabilities of recurrence derived from each trial were used to calculate a weighted mean annual probability of recurrence according to the weights assigned to each study in the Cochrane review.[15]

Standard error was estimated for the probabilities derived from each trial at each time point using the following formula:

Equation 6:

$$\boldsymbol{Standard\ error = \sqrt{\frac{p_t\,(1 - p_t)}{n_t}}}$$

where *p* is the conditional annual probability of recurrence at *t* for an individual trial and *n* is the number of patients at risk at *t* for that trial. A weighted mean estimate of standard error was also calculated. We carried forward the values observed at seven years of follow-up for all subsequent years in the absence of further follow-up data. The estimates and their standard errors as calculated from the literature are provided in Appendix 4.3.

The p(recur|NED) estimates for HER2- patients were derived from a recent EBCTCG meta-analysis of the effect of polychemotherapy on BC recurrences and survival.[21] The analysis did not specify rates of progression for HER2+ vs. HER2- patients treated with polychemotherapy. This was likely due to the fact that HER2 status was not tested in all clinical trials, particularly those initiated before 1998. We assumed that the prevalence of HER2+ disease in the EBCTCG sample was similar to published studies of HER2+ disease, approximately 20% of BC cases. We subsequently calibrated these parameters to address the lack of HER2 specificity (see 4.3.3 Model Validation & Calibration). The EBCTCG meta-analysis included 194 clinical trials of adjuvant hormonal or chemotherapy from 1995 onwards. Analyses were carried out at 10 and 15-years of follow-up using patient level data. We preferred the 2005 meta-analysis to the more recent 2012 update[44] because the former provided an analysis of all polychemotherapy regimens (vs. controls), whereas the latter compared specific polychemotherapy regimens (e.g. anthracycline alone vs. anthracycline + taxane). The 2005 analy-

sis provided strong evidence for the rate of recurrence with polychemotherapy as it involved a total of 60 trials, 28,764 women and 10,173 deaths observed. We digitally extracted the cumulative probabilities of recurrence (either local, distant or death) for years 1 through 10 from the analysis of recurrence with polychemotherapy (vs. no adjuvant chemotherapy) for patients aged 50-69 years. The cumulative probabilities of recurrence were then used to calculate conditional annual probabilities of recurrence per Equation 5. Standard error was also estimated similarly (Equation 6), with the number of patients at risk derived from appendices to the EBCTCG meta-analysis.[45] The literature-based estimates and their standard errors are provided in Appendix 4.3.

The probability of the first recurrence being distant, conditional on having a recurrence [p(DR|recur)] was calculated as the proportion of first distant events among all first events reported in the literature. The HER2+ data were derived from each individual trial included in the p(recur|NED) estimate. We included recurrences in the brain, bone, central nervous system, liver, lung, skin, pleural effusion, soft tissue or recurrences termed 'distant' in our calculation of distant recurrences. We also included death events in our calculation of distant recurrences. For the Buzdar neoadjuvant trial, we assumed that the single recurrent event leading to death due to distant disease was a distant recurrence for calculation purposes.[20] The FinHer study reported all events without specifying first events.[16] We contacted the author for clarification without success. Therefore, we assumed that

the number of first distant events was equal to the total number of events less locoregional and contralateral recurrences and deaths.[46] Finally, the NOAH study specified only distant recurrences and death events.[17] We therefore assumed that all other events were non-distant for calculation purposes. The standard error for each trial-based p(DR|recur) estimate was calculated per Equation 6 with the number of patients at risk equal to the number of patients that experienced a recurrence in each respective control arm. A weighted mean estimate for p(DR|recur) and the corresponding standard error were calculated as previously described. The resulting $p(DR|recur)_{HER2+}$ was 0.69 with a standard error of 0.04 (see Paper 4 Table 3).

We were unable to derive the p(DR|recur) from the EBCTCG meta-analysis because the types of first recurrences were not reported. A targeted literature scan of references related to the sources used in other models[2,3] was conducted to identify more recent studies of disease recurrence in HER2- patients. We excluded the data source used by Lidgren as this relied on a small (n=116), single clinical trial population based in Finland.[16] We also excluded the source from Blank's analysis as it included patients diagnosed before standard chemotherapy practice changed in 1998 and did not report data in sufficient detail to distinguish the type of first recurrence.[47] A European study of recurrences after breast-conserving surgery was eliminated as it included metastatic patients and those diagnosed before 1998.[48] The preferred study to inform $p(DR|recur)_{HER2-}$ was a ret-

rospective cohort study of 1,134 early-stage BC patients diagnosed between 1998 and 2005 in Wisconsin.[49] The study sample was similar to Ontario BC patients based on age at diagnosis (mean 62.7 years), surgical treatment patterns (98.5% surgically treated). The proportion of HR+ (79.1%) and HER2+ disease (17.7%) was slightly higher than Canadian [50] and Ontario populations, respectively. The proportion of HR+, HER2- patients was notably higher (83.7%) than Ontario or Canadian population estimates. This would tend to reduce the p(DR|recur) estimate in the HER2- arm because HR+ disease is less aggressive and better managed with endocrine treatment. In terms of treatment compatibility, the cohort study included patients diagnosed between 1998 and 2005, and was therefore more likely to reflect practice consistent with current protocols. Patients were followed for a median of 4.8 years (range 3.2 to 9.4) during which 99 recurrences were observed. The p(DR|recur) estimate was calculated as per HER2+ patients. Standard error was estimated from the number of HER2- patients that experienced a recurrence per Equation 6. The resulting $p(DR|recur)_{HER2-}$ was 0.66 with a standard error of 0.05. (see Paper 4 Table 3)

Although HER2+ disease is known to be more aggressive than HER2-, studies of the progression of HER2+ disease after LRR are not available. Furthermore, the natural history of BC following LRR is generally poorly understood and not indexed in the major medical literature databases. Therefore we approached our literature review using a seed article [51] to identify studies of DR af-

ter LRR in BC. The seed article described a multistate model that investigated the impact of LRR on progression to metastatic disease, but the authors assumed a constant probability over time. Moreover, the generalisability of the study to current practice was limited as the patient sample was recruited from 1980 to 1999. Finally, we excluded the seed study because it did not analyse the DR outcome using time to event analysis. We screened all citations related to the seed article to identify a recent study reporting the cumulative annual incidence of distant recurrence in a cohort of patients who had experienced a local or regional recurrence.[22] The patients were diagnosed between 1994 and 2005, with 15% of the 279 patient cohort diagnosed prior to 2001. Therefore, we reasoned that the impact of pre-1998 chemotherapy practices was likely to have a limited effect on recurrence rates. Patients were followed for a median of 5.9 years (range 0.9-12.2). We digitally extracted the cumulative annual probability of DR after local and regional recurrences from the reported Kaplan-Meier curves for the available 7 years of follow-up. The cumulative probabilities of distant recurrence after local and regional recurrences (reported separately) were summed to estimate the cumulative probability distant recurrence after locoregional recurrences. We then estimated the conditional annual probabilities of p(DR|LRR) and standard error estimates as described previously. The value for year 7 was carried forward for all subsequent years in the absence of additional follow-up (Appendix 4.4). We subsequently

optimised this parameter during the calibration process to address the lack of HER2 specific progression data.

4.3.2.4 Mortality

General mortality was derived from Statistics Canada life tables for Canadian women.[52] This provided age-specific annual mortality rates for a sample of the Canadian population. This variable was not fit to a beta distribution, but modeled deterministically. We also considered the use of Statistics Canada life tables to estimate disease-specific mortality due to DR [p(mort|DR)]. However, the Canadian Cancer Registry maintained by Statistics Canada does not contain information about stage at diagnosis, making it impossible to limit to those with distant disease. Instead, disease-specific mortality rates were derived from the OCR.[23] This dataset described the mortality of patients diagnosed at clinical stage IV (metastatic) disease. The mortality of patients after progressing to distant disease following an incident early-stage diagnosis is not documented in the OCR. This dataset described the survival of women diagnosed between 2005 and 2009, age at diagnosis and provided 5 years of follow-up. We modeled age-specific mortality rates deterministically and assumed the probability of mortality observed at 5 years remained constant for all subsequent years.

4.3.2.5 Treatment-Related Parameters

We expected the treatment effect to significantly impact ICER estimates and therefore prioritised evidence derived from systematic reviews. The outcome

of interest was the effect of 1 year of adjuvant trastuzumab on DFS analysed using time to event analysis. A targeted literature review was conducted to identify all systematic reviews of trastuzumab therapy published up to August 2012 in PubMed (see Appendix 4.7). We identified seven reviews of trastuzumab treatment. [15,53–58] Among these, 1 focused on metastatic treatment [58], 1 focused on trastuzumab-related cardiac toxicity [55] and the remaining 5 evaluated adjuvant therapy [15,53,54,56,57]. We considered only adjuvant therapy reviews to estimate the effect of trastuzumab on recurrence in the model. We eliminated 3 adjuvant reviews that did not use time to event analysis for the DFS outcome.[54,56,57] Finally, we excluded an adjuvant therapy meta-analysis as the methods were not reported in the publication or appendices.[53]

Treatment effect was estimated from a Cochrane review of adjuvant and neoadjuvant therapy.[15] The inclusion of neoadjuvant patients in the sample was not considered a threat to the model, as neoadjuvant treatment with trastuzumab is practiced in Ontario, and the neoadjuvant studies [17,20] made up 11% of the DFS treatment effect estimate. Moreover, neoadjuvant therapy is typically offered to patients who present with locally advanced disease at incident diagnosis. Our Ontario study of trastuzumab utilisation demonstrated similarity to the patients in these trials; trastuzumab-treated patients were more likely to have higher stage and grade at diagnosis, suggesting that trastuzumab-treated patients were more likely to have locally advanced disease than untreated (data on file). Therefore,

the combined neoadjuvant and adjuvant treatment effect represented a more heterogeneous patient population. The meta-analysis was conducted with patient level data within an intention-to-treat (ITT) framework and outcomes were expressed as hazard ratios. The ITT perspective likely biased against trastuzumab because numerous studies stopped early or allowed patients to crossover to trastuzumab due to patient benefit (see 4.3.2.3 Disease Progression).

For the base case analysis, we assumed that all patients completed 1 year of trastuzumab treatment subsequent to anthracycline-based chemotherapy per Ontario guidelines.[27] We modeled the treatment effect from the DFS outcome for sequential adjuvant trastuzumab in the base case. The corresponding HR of 0.71 (95% CI: 0.53, 0.95) was applied to the probability of recurrence from the NED state for true HER2+ patients assigned to trastuzumab treatment. This estimate was derived from the HERA and PACS-04 trials. The international HERA trial [59] sample had a median age of 49 years and was restricted to HER2+ (IHC 3+ or FISH+) tumours ≥1 cm. The sample included patients who received prior neoadjuvant (11%) and adjuvant (95%) chemotherapy. Prior treatment consisted of anthracyclines alone (68%), anthracyclines in combination with taxanes (26%), or a regimen without taxanes (6%). Patients with HR+ disease (50%) were permitted endocrine therapy. The meta-analysis included 2 year follow-up data instead of the 4 year follow-up data [15] to limit the attenuation of the treatment effect due to crossover. The meta-analysis included only the patients randomised to

trastuzumab or observation from the PACS-04 trial.[19] That patient sample had a median age of 48 years and was restricted to LN+ patients with HER2 overexpression (IHC 3+ or FISH+). All patients received prior treatment with an anthracycline, either in combination with alkylating and anti-metabolite agents, or with a taxane. Endocrine therapies were permitted for HR+ patients (60%). We consulted with local clinical experts (N.L, Maureen Trudeau) to confirm that these studies were similar to Ontario patient populations and practices. Patient age in both trials was younger than the mean age at incident BC diagnosis in Ontario, but patients age ≤ 59 years made up 70% of trastuzumab treated patients in our study of Ontario practice. (data on file) Thus, we determined that the estimate of sequential treatment effect was likely generalisable to the Ontario population. (data on file)[13]

We used the probability of CHF with sequential therapy, defined as NYHA class III or IV CHF, to estimate the probability of treatment discontinuation due to symptomatic cardiac toxicity in the base case. The pooled proportion of sequentially treated patients that experienced CHF among all treated patients was 0.02 (SE=0.006) This informed the transition from the NED state to the DC state for all patients receiving trastuzumab independent of true HER2 status. The treatment discontinuation assumption is consistent with Ontario practice and Canadian guidelines, which advocate treatment discontinuation if absolute left ventricular ejection fraction (LVEF) decreases by ≥15% from pre-treatment levels, or

if symptomatic CHF develops. [60] We also assumed the corresponding sequential treatment estimate of LVEF decline to estimate the rate of asymptomatic cardiac toxicity in the model (0.18, SE=0.02). This probability was used to estimate the additional costs of monitoring and pharmacological therapy for asymptomatic patients. We assumed that all asymptomatic patients discontinued treatment temporarily while resolving cardiac dysfunction, but returned to complete a full course of treatment.(NL, MT) Finally, we assumed no mortality due to symptomatic or asymptomatic cardiac toxicity, consistent with published evidence in the adjuvant setting.[15]

We considered the pooled effect of sequential and concurrent trastuzumab on DFS in a scenario analysis. The Cochrane review pooled the treatment effect across an additional 5 trials and found a DFS HR of 0.60 (95% CI:0.50, 0.71) . Although Ontario treatment guidelines do not currently advocate sequential trastuzumab therapy, we found that 9% of Ontario patients received >18 trastuzumab doses suggesting some sequential treatment in Ontario practice.[61] Moreover, consultation with clinical experts (NL, MT) suggested that current practice does include patients treated concurrently with a taxane or carboplatin at certain institutions (~20% of current patients). Therefore, we determined that it was reasonable and informative to estimate the impact of sequential and concurrent treatment despite the lack of guideline endorsement. The BCIRG-006 trial examined trastuzumab concurrent with carboplatin, which represents 23% of the

overall sequential and concurrent DFS treatment effect. The international BCIRG-006 study [18] included patients with a median age of 49 years who had LN+ or high risk LN- disease with HER2 overexpression (FISH+). Patients were treated with a taxane concurrent to carboplatin and trastuzumab or an anthracycline concurrent to an alkylating agent, followed by trastuzumab concurrent to a taxane. The estimate also included the National Surgical Adjuvant Breast & Bowel Project (NSABP)-B31and N9831 combined trials [62] of trastuzumab concurrent to a taxane following combination therapy with an anthracycline and alkylating agent. The patient sample included LN+ patients with HER2 overexpression (IHC 3+ or FISH+), 60-67% of which were aged 40-59. The FinHer, Buzdar and NOAH trials were the only trials that did not evaluate 1 year of trastuzumab therapy accounting for 16.5% of the overall estimate. Adjuvant treatment with 9 weeks of trastuzumab in combination with a taxane or vinorelbine was investigated in the FinHer trial, while the Buzdar and NOAH trials examined neoadjuvant treatment. The shorter duration and higher risk patient populations included in these studies was seen as introducing heterogeneity that would make the treatment effect estimate resemble current practice more closely. Indeed, our study of Ontario practice patterns found that 26.0% of patients completed less than 17 trastuzumab treatments.[61] We also modeled the pooled sequential and concurrent treatment estimates of CHF (0.03, SE:0.006) and LVEF decline (0.17, SE:0.02) for consistency. This was an important consideration, as concurrent trastuzumab therapy is thought

to increase the risk of cardiac toxicity. We anticipated higher rates of cardiac toxicity in actual practice. Canadian guidelines suggest a minimum LVEF of 50% - 55% before starting therapy[60], while the lower limit for the majority of adjuvant trials was 55%.

The duration of the treatment effect was fixed at 5 years in the base case for patients who completed therapy (NED) and those who discontinued (DC). This is consistent with several other models of adjuvant trastuzumab[39] and therefore allowed us to compare our results to others.[26] However, long-term follow-up is needed to confirm the duration of treatment effect and this assumption has not been verified. We therefore examined alternative treatment duration scenarios in sensitivity analysis. Treatment effect durations of two and seven years were modeled to coincide with the smallest median follow-up and longest follow-up duration of the adjuvant trials. Finally, we allowed the treatment effect duration to vary probabilistically between 1 and 10 years in sensitivity analysis to examine the impact of our uncertainty about treatment duration on ICER estimates. A separate set of assumptions was applied to the effect of treatment in patients who discontinued. In the base case we assumed a 50% reduction in the trastuzumab HR. Alternative reduction scenarios (0%, 100%) were examined in sensitivity analysis.

4.3.2.6 Current Treatment Practice

The impact of current practice patterns was estimated through a series of scenario analyses informed by our retrospective cohort study Ontario practice

(Chapter 3). In addition to the test and treatment practices that were modeled in the HER2 decision-tree, we modeled Ontario-specific estimates of symptomatic (0.06, SE:0.06) and asymptomatic (0.16, SE:0.07) cardiac toxicity. We also modeled treatment discontinuation for non-cardiac reasons in sensitivity analysis, which occurred in 10.4% (SE:0.01) of trastuzumab-treated patients.

4.3.2.7 Costs

All costs were inflated to 2012 Canadian dollars per the Health and Personal Care component of the Consumer Price Index where necessary.[63] We included only direct healthcare costs from the perspective of the Ontario public payer. All costs are summarised in (see Paper 4 Table 3). The cost of testing for HER2 status was estimated per 2011 provincial reimbursement rates.[64]

We estimated the cost of trastuzumab treatment assuming an average patient mass of 70kg. We calculated the drug cost and estimated administration costs per product monograph [65] recommendations:

- Loading dose: 90 minute infusion at 8 mg/kg
- Maintenance dose: 30 minute infusion at 6 mg/kg for 17 subsequent doses (3 weekly dosing)

Drug acquisition costs were estimated per mg and assumed no vial wastage.[66] Drug administration included the cost of physician supervision for a complex single biological agent at each visit [67], chair time and nursing costs.[68] Cardiac monitoring costs included the cost of 1 baseline assessment and 4 subsequent assess-

ments at 3 month intervals per Canadian guidelines.[60] The cost of cardiac monitoring was a weighted average of echocardiography (ECHO) and multi-gated acquisition scanning (MUGA), and physician interpretation of each [67], per Ontario utilisation patterns.[61] The cost of supportive medicines during trastuzumab therapy were estimated from a Canadian cost burden study of trastuzumab in the adjuvant setting.[69] Finally, trastuzumab treatment costs also included the cost of additional cardiac assessments and 8 weeks of pharmacological treatment of asymptomatic (NYHA class I, II) cardiac toxicity.[60,70] This cost was applied for only the proportion of patients experiencing asymptomatic cardiac toxicity.

Patients in the NED and DC states incurred costs associated with monitoring for disease recurrence. These costs were derived from a study of the lifetime costs of BC treatment in Canada.[71] The costs of monitoring included physician assessments, blood work, biochemistry tests, mammograms, bone scans and liver ultrasounds. We assumed that the costs would decrease steadily between years 1 and 5. Patients in the DC state incurred additional costs associated with cardiac monitoring [60] and pharmacological treatment [70] of symptomatic cardiac toxicity on top of recurrence monitoring.

The costs of treatment in the LRR state were determined according to HER2 status. For HER2- patients, we estimated costs including diagnostic workup, surgical treatment, hospitalisations, radiation therapy, chemotherapy and hormonal therapy per a prior Canadian disease burden study.[71] We assumed that HER2+

patients would receive the additional cost of sequential trastuzumab therapy in the LRR state per clinical consultation (NL, MT). Monitoring for recurrences in the PLRR state included physician assessments, clinic costs, blood work, biochemistry tests, mammograms, chest x-rays, bone scans and liver ultrasounds.[71]

The costs of treating distant recurrence were also estimated from an earlier study of the burden of BC in Canada. [60,61,71] That study reported initial treatment, ongoing and terminal care costs for metastatic disease. Initial treatment costs included diagnostic workup (general assessment, biochemistry , bone scan, haematology, abdominal ultrasound, chest x-ray, bilateral mammogram, CT scan, and skeletal survey) and local treatment of the breast tumour and metastases (therapeutic and diagnostic surgery, radiation, systemic chemotherapy and hormonal treatment). Ongoing care and terminal care costs were of hospitalisations, palliative radiation and physician consultations. We estimated the annual average cost of combined terminal, initial and ongoing care based on the average survival of 2.6 years reported by the study authors. This annual cost was applied as the cost of DR for HER2- patients. We assumed that HER2+ patients would incur additional costs associated with trastuzumab therapy and monitoring for cardiac complications. We estimated the additional cost of trastuzumab for 7.2 months of therapy as previously reported in the Canadian setting. [69]

We varied any costs based on resource utilisation estimates from the literature by ±20% in probabilistic analysis. All stochastic cost parameters were fit

to gamma distributions, which allow for modelling of skewed cost data and can be constrained to restrict costs below $0. The costs of trastuzumab therapy and HER2 testing were not varied as the reimbursed cost of this drug per unit is fixed.

4.3.2.8 Utilities

A vast array of health utilities are available to describe the quality of life associated with various BC-related health states. Studies using a single methodological approach (i.e. time trade-off) in a Canadian sample were not available. Therefore, we selected quality of life studies conducted in North American female BC populations that matched the health states of this model. All health utilities are shown in Paper 4 Table 3. Utility weights were applied to the life-years in each health state to estimate QALYs. All utilities were derived from the literature, and were varied ±20% stochastically using the gamma distribution. The gamma distribution allows for modelling of utility weights that are less than zero when a utility decrement transformation is applied, and therefore is a reasonable selection for utilities.[72]

We assumed a utility decrement of 0.08 (estimate of 0.92) for patients in the NED state during trastuzumab therapy.[73] All recurrence-free patients remaining in the NED state after completing therapy received a utility weight of 0.96 derived from a US sample of post-menopausal women.[74] We applied a utility decrement to patients as they aged in the NED state. Patients who discontinued treatment early were assumed to experience a utility decrement associated with

the loss of therapeutic opportunity and the symptoms of CHF. Because the utility decrement associated with symptomatic CHF was larger than expected for treatment discontinuation, we applied the former utility weight for the DC state. We estimated the utility weight for the DC state as the median between NYHA class III and IV CHF measured in a US patient sample.[75] The utility decrement was applied to the first year in the DC state, after which we assumed resolution of acute cardiac toxicity and a return to the NED utility weight. A utility decrement for asymptomatic CHF (NYHA class I and II) is applied to the proportion of patients that experienced asymptomatic cardiac toxicity in the first year in the NED state.[75]

Patients who progressed to LRR received a utility weight (0.70) that reflected typical treatment (surgical, radiological and chemotherapeutic) for first local recurrences.[76] Patients in the post-LRR state were assumed to return to full health and received the same utility as the NED state. Finally, patients who progressed to the DR state experienced a utility of 0.57.[74]

4.3.3 MODEL VALIDATION & CALIBRATION

Model validation and calibration are 2 separate approaches to improving the accuracy and generalisability of a disease model to the local setting. Validation is a process by which the performance of the model is evaluated. It ensures the internal validity of model predictions by confirming that the model calculations appropriately replicate the datasets used in its construction.[6,12] Calibration is

a process of external validation which ensures that model predictions are consistent with external datasets describing disease natural history.[6,77] The calibration process can be used to fine-tune parameters that were unobserved in the target population.

4.3.3.1 Selection of Parameters to Validate & Calibrate, & Target Endpoints

The BC natural history model estimates incremental costs per life-year (LY) and quality-adjusted life-year (QALY) gained in the Canadian population. Furthermore, the model is designed to evaluate the cost-effectiveness of diagnosing and targeting HER2-directed therapy in BC with targeted therapy is to extend life and improve its quality for patients with a particularly aggressive form of the disease. Therefore, the parameters with the greatest influence on outcomes are those that influence LY and QALYs. This includes transition probabilities that inform disease progression (to either LRR or DR) and disease-specific mortality [e.g. p(mort|DR)]. A further factor that influences the decision of which parameters to calibrate or validate is whether the parameter was observed for the population and setting of the evaluation.[77,78] In the case of the BC Markov model, disease-specific death due to distant recurrence [p(mort|DR)] was informed by a cohort of Canadian patients identified through the OCR. All other disease-specific parameters were identified from the literature, either from a mix of international patient samples or a US-based sample. Therefore, the best candidate for validation is the p(mort|DR) parameter, as it was (1) directly observed in the population of

interest, (2) is most likely to reflect local practice, and (3) expected to have a direct impact on costs, LY and QALY outcomes based on findings from other models.[1,2] Conversely, the p(recur|NED), p(DR|recur) and p(DR|LRR) disease recurrence parameters were not observed in a Canadian sample with some differences in clinicopathologic features of the study populations noted previously. We therefore decided to calibrate disease recurrence and progression parameters against external empirical data. The probability of treatment discontinuation due to symptomatic heart failure [p(dc)] was not considered for calibration because it was not a direct source of mortality in the model, and has demonstrated little impact on LY and QALY outcomes in other models. General mortality was not considered for calibration or validation as it was observed directly from the Canadian population, was not transformed in model programming and impacted patients equally regardless of HER2 status.

Validation and calibration target selection is largely influenced by the intervention under examination and the availability a large empirical dataset with limited biases.[77] Mortality is therefore an ideal candidate for both validation and calibration. Disease-specific mortality from the DR state was the target for validation, while all-cause mortality was the target for calibration. Further details of the approach to measuring each target and the selection of empirical data are provided below.

4.3.3.2 Validation Process

The validation process compared BC mortality predicted by the model to the expected BC mortality according to the stage IV mortality dataset [23,45]used to inform the model. Specifically, the cumulative predicted mortality at each model cycle was compared to the cumulative expected mortality at the same cycle, for all cycles until the cohort reached age 100. This comparison was made for each arm of the model (A through F) to ensure that programming was consistent for the model arms accommodating each test-treat strategy.

All transition probabilities that could confound BC mortality predictions were eliminated by: (1) adjusting all disease progression parameters such that all patients in the cohort transitioned directly into the DR state within 1 model cycle, (2) setting the probability of treatment discontinuation to zero, (3) setting all other sources of mortality (e.g. general mortality) to zero, and (4) setting the discount rate to zero. Independent validation of BC mortality rates by HER2 status was not conducted as the rates weren't modeled separately for each subgroup. Additionally, the effect of trastuzumab treatment was not applied during validation to permit reflect the natural disease process in the absence of targeted therapy.

The cumulative mortality for a 50y cohort was calculated as follows:

$$\boldsymbol{Cumulative\ Mortality} = \sum_{n=1}^{n} \boldsymbol{\delta_n}$$

where n represents the model cycle and δ represents the number of deaths (events) occurring up to and including that cycle.

4.3.3.3 Calibration Process

The calibration process compared model predictions for the target endpoint (all-cause mortality) to expectations based on estimates from an external empirical dataset. Cumulative mortality was calculated per the validation approach. An external dataset describing the survival of Canadian women diagnosed with BC at any stage was obtained from Statistics Canada. This data could not be used to inform the model as it did not describe transitions between the various disease states preceding death and it did not distinguish between patients diagnosed at early-stage vs. metastatic BC. However, it provided a good basis of comparison for the BC model predicted mortality because it represented the natural history of BC, including diagnostic and management practices, in the target Canadian population. In addition, the dataset was appropriate because it: (1) incorporated patients diagnosed at any stage of disease, (2) accounted for the majority of Canadian patients, and (3) captured the target outcome of all-cause mortality (e.g. death due to BC and general mortality).

The Statistics Canada dataset included a maximum of eight years of follow-up for patients diagnosed with BC between 1998 and 2006. Patients diagnosed before 1998 were excluded due to differential chemotherapy practices which were associated with lower survival rates vs. current practice.[2914] The da-

taset included variables describing: (1) time to death or right-censored loss to follow-up from incident BC diagnosis, and (2) age at diagnosis in years. The time to event data was fit to a Weibull model with age as the sole predictor to estimate mortality expectations beyond the eight years of available follow-up. Age was represented as a categorical variable comprised of 11 five-year age groups. Prior to model fitting, the log-rank test[4373] was used to test the null hypothesis that survival was the same between the 11 age groups. This test indicated that survival was significantly different by age ($p<0.0000$) and justified fitting to a model with age as a predictor. The appropriateness of the Weibull model was also tested by graphing the ln (-ln) of the Kaplan-Meier survival estimates against the ln of follow-up time for each age category. The log-log plots for the age categories were all relatively straight and parallel (Appendix 4.3). These suggest that hazards were proportional between the age categories over time (parallel lines), and that the effect of age on survival was multiplicative with respect to survival time (straight lines on a log-time scale).[79] Therefore, the data exhibited consistency with Weibull assumptions and fitting to this parametric model was appropriate. For the purposes of model calibration, expected cumulative mortality was estimated from the group of patients diagnosed between 50-54 years of age per the fitted Weibull model (specifications provided in Appendix 4.4).

A random search method [77] was used to identify recurrence input parameter values that produced all-cause mortality predictions which best matched ex-

pectations. This optimisation approach made use of the pre-existing model programming in Excel by randomly selecting parameter values from the stochastic distributions assigned to each parameter undergoing calibration. The parameters were calibrated in a stepwise fashion by allowing a single parameter or parameter group to vary probabilistically while holding all others constant at the expected (or calibrated) value. The recurrence parameters were optimised in sequence of their natural occurrence in the course of BC. The p(recur|NED) parameter was calibrated first, followed by p(DR|recur), and finally p(DR|LRR) was optimised. Once an optimal value was identified for a given parameter, that value was used in all subsequent steps of the calibration process.

The p(recur|NED) parameter for HER2+ and HER2- patients was given special attention during calibration. Clinical input provided by the committee (NL) indicated that the pre-calibration probabilities of recurrence (Figure 4.3) for HER2+ and HER- patients lacked face and clinical validity for several reasons:

(1) The conditional annual probability of recurrence for HER2+ patients was not approximately twice that of HER2- patients (Figure 4.3); this is generally accepted in the clinical community [16] and is a common modelling assumption [2,80];

(2) The HER2+ probability of recurrence decreased and increased between years six and seven (Figure 4.3), while its generally accepted that BC recurrence rates steadily decline from year two onwards [28]; and

(3) [7,512]The probability of recurrence in HER2+ patients dipped below that of HER2- patients between four and seven years of follow-up (Figure 4.3); this violates (1) and also contradicts evidence that HER2+ disease is more aggressive than HER2-.

These inconsistencies highlighted concerns about the validity of the HER2+ data specifically. The HER2+ recurrence rates were derived from the control arms of seven adjuvant and neoadjuvant trials included in a Cochrane review and meta-analysis.[15] Several of these trials either stopped early [20,62] or allowed control patients to cross over to the treatment arm due to benefit at prescheduled interim analyses.[16–18,59] In the most extreme case, 50.7% of control patients switched over to trastuzumab treatment in the HERA trial.[59] The HERA trial accounted for 25% of the weighted estimates of the conditional annual probabilities of recurrence in years one through three. Moreover, each trial reported results per the intention to treat principle, making it impossible to differentiate controls that switched when extracting data to estimate the rate of recurrence in untreated HER2+ patients. Therefore, we could not conclude that the HER2+ estimates were free of bias from trastuzumab crossover. We recognised that the HER2+ rates of recurrence could be artificially low in later years as a result of crossover, and therefore investigated alternative scenarios with increased HER2+ recurrence rates during calibration. Because we also had uncertainty about the relevance of the pre-calibration HER2- recurrence estimates to the Canadian setting, we considered

scenarios where we reduced HER2- recurrence relative to the HER2+ data. Thus, we investigated 2 alternative hypotheses about the relationship between HER2+ and HER2- recurrence. We considered 3 deterministic scenarios where assumptions about the HER2+ or HER2- recurrence rates were tested (see Table 4.1). The scenario that produced all-cause mortality that was most consistent with expectations was then carried forward for probabilistic calibration of the recurrence from NED parameter.

The random search method was constrained during calibration by fitting narrow beta distributions to the varying parameter(s). This ensured that the randomly selected series of probabilities maintained the expected general shape of recurrence patterns with respect to time. In other words, we constrained the stochastic process to eliminate parameter sets with drastic and clinically implausible increases and decreases in recurrence over time. A summary of each calibration scenario is provided in Table 4.1.

4.3.3.4 Model Performance Evaluation

The goodness of fit of (GoF) model predictions with expectations in both the validation and calibration processes was measured using the least squares method to estimate goodness of fit.[6,78] This approach was appropriate for calibration to a single endpoint (mortality) and was chosen for its intuitiveness and low data demands.[6,77] The sum of the squared differences (errors) between predicted and expected mortality was calculated for both validation and calibration purpos-

es. Cycle-specific cumulative mortality was used to calculate GoF over all model cycles using the following formula:

$$\boldsymbol{Goodness\ of\ Fit} = \sum_{n=1}^{n}(\boldsymbol{\delta}_{predicted_n} - \boldsymbol{\delta}_{expected_n})^2 \quad 6$$

where δ represents mortality.

Expected mortality for model validation was determined directly from age-specific five-year, stage IV mortality data obtained from CCO.[23] The model predicted mortality also represented mortality due to progression to distant disease (stage IV). Therefore, this comparison satisfied the needs of the validation process by ensuring that the model was accurately predicting disease-specific mortality due to metastasis. The GoF estimate was determined across all model cycles until the cohort reached age 100. The mortality observed at five years follow-up was assumed for all subsequent years when determining expected mortality. This approach to model fit estimation treated all years (or cycles) equally as the errors were not weighted by cycle. Our expectation was that model fit would be close to or near perfect, resulting in a GoF estimate of zero, if the model were producing valid estimates of disease-specific mortality. This expectation was reasonable given that all other sources of mortality were eliminated during the validation process. Thus, the validation process served to ensure that the model could accurately replicate the data that was used to inform disease-specific BC mortality. We estimated the GoF for each arm of the model during validation to ensure that the programming of each arm correctly replicated disease-specific mortality. This

confirmed that programming errors did not result in biases for or against the test-treat strategy assigned to any given model arm. Model fit was also evaluated by graphing predicted cumulative survival vs. expected cumulative survival to provide a visual comparison of the magnitude of error in earlier vs. later model cycle predictions.

A similar approach to estimating GoF was used for model calibration, however, each parameter varied probabilistically and predicted mortality represented the sum of disease-specific and general mortality. Expected mortality was determined from an external dataset of all-stage mortality in Canadian women diagnosed with BC (previously described in section 4.3.3). We evaluated model fit across (1) all model cycles, (2) up to cycle 10, and (3) up to cycle 25. The purpose of calibration was to identify the trial (and corresponding parameter set) that produced the best (smallest) GoF estimate across all model cycles, but in cases where two or more trials resulted in very close GoF estimates, we examined fit over cycles 10 and 25 as secondary measures. The GoF estimate was determined for each stochastic iteration (trial) of the disease model for a total of 1,000 iterations. Model fit was also evaluated by graphing predicted cumulative mortality vs. expected cumulative mortality for the 10 best fitting trials for each parameter under calibration. This provided a visual representation of whether errors in estimated mortality were greatest at earlier or later model cycles.

We used different convergence criteria for validation vs. calibration. The convergence criteria define the conditions under which a new set of parameter inputs were accepted.[77] In the case of model validation, we wanted to ensure that the model was predicting BC-specific mortality that was exactly consistent with the data used to inform the BC mortality parameter. In other words, we wouldn't accept any error in model predictions vs. expectations and sought a GoF estimate of 0.00. The criteria for calibration were not as strict, as we were searching for the parameter sets that best minimised the GoF estimate over all cycles or a subset of cycles as described previously. This is a common criterion for reaching convergence when calibrating a parameter using the random search method.[77]

4.4 SUPPLEMENTAL RESULTS

4.4.1 STRUCTURED REVIEW

4.4.1.1 Study Selection

The structured search of MEDLINE and EMBASE to February 2, 2012 identified 1,761 and 2,363 citations for review, respectively. We reviewed 1,621 titles and abstracts after removing non-English and duplicate citations. We selected 210 citations for full text review. Four citations could not be obtained in full text and were excluded. The quorum diagram (Figure 4.1) details the number of studies that were excluded for each criterion. Fifteen percent (n=30) of excluded studies examined agreement in selected populations such as those sent to a referral laboratory in routine practice, emphasis on clarification of IHC equivocal or failed FISH results. We noted a similar 18% (n=35) of studies that reported agreement between HER2 tests as a secondary outcome to primary research questions about prognosis or pre-post treatment HER2 expression. Finally, 19% (n=38) of excluded studies did not verify IHC, CISH or SISH results against gold standard FISH. Overall, 197 citations were excluded during full text review, leaving 17 studies for data extraction and meta-analysis. Agreement on study selection (Cohen's Kappa) was 0.86. An additional 5 studies did not report data in an extractable format, most frequently reported as a proportion of concordant or discordant results (Appendix 4.10). In total, data was extracted from 12 studies for meta-analysis.

4.4.1.2 Characteristics of Included Studies

All included studies were published between 1999 and 2012. Table 4.2 describes the characteristics of each included study. A geographically diverse collection of studies were included in this review; two (17%) were conducted in North America, one in (8%) Europe, five in (42%) Asia-Australia and two (17%) in India and one (8%) in South America. A number of North American studies were excluded due to the use of the lower HER2:CEP17 positivity ratio of >2.0 for ISH. Many study characteristics were reported inconsistently and descriptors of the study population were challenging to characterise. Overall, the sex of the study population was reported least frequently, only in two studies (17%). One study reported including a single male patient.[81] Disease stage was also reported in only two studies, although all studies did indicate invasive disease and were therefore eligible for inclusion. Metastatic tissues were noted in one study, but these comprised <10% of the entire sample.[82] All studies specified formalin fixation and paraffin tissue preservation, but only two provided further detail as to the type of tumour specimen. Details about the study population age were not reported for 69% (n=8) of included studies. The laboratory setting was not reported for just over half (n=7) of included studies. Among those that did report the setting, the majority indicated a hospital or university laboratory.

4.4.1.3 Quality of Included Studies

Characteristics of study design examined as markers of study quality were also poorly reported, suggesting a high risk of bias in the studies in this review (Table 4.3). Sample selection methods were not reported in three quarters of studies, while the remainder reported consecutive (n=2), or randomised (n=1) selection. The recruitment period was best documented among study quality indicators, with only two studies failing to report this characteristic. The blinding of sample assessors was not reported in five (42%) included studies; 6 (50%) reported blinded assessment. The majority of studies applied the gold standard to verify the entire sample population (n=8, 75%). All studies that did not verify the entire sample (n=4) provided an explanation for the attrition of samples. Based on the quality indicators assessed, we found that the sample of studies included in this review had either a moderate of high risk of bias.

4.4.1.4 Meta-Analysis & Model Inputs

The meta-analysis of IHC agreement with the gold standard FISH test included 10 studies and a total of 3,064 tissue samples (Table 4.4). True positives made up a pooled mean of 54.1% of IHC equivocal results. The best agreement was observed for IHC-, of which a pooled 95.7% were confirmed HER2- by FISH. The analysis of agreement among IHC- results included 2 studies without discordant events. These studies were smaller, with 15[83] and 30[84] IHC- results, and thus received smaller weightings of 0.01 and 0.02 (respectively) in the overall

estimate. Concordance was slightly lower for IHC+ test results, with 87.8% being true positive by FISH. The pooled analysis of IHC+ results included 1 study that observed no discordant events.[85] This study sample was relatively large in comparison to others in the pooled estimate, and thus contributed to 0.18 of the overall estimate. The IHC+/FISH+ agreement rate is lower than the minimum (90%) required for IHC+ results under the College of American Pathologists' guidelines for quality assurance of HER2 testing laboratories.[33] Although this estimate was derived through a rigorous systematic review of primary diagnostic studies, the lower agreement rate may raise questions of validity in Canadian practice. Indeed, HER2 testing in Ontario is subject to a quality assurance program that reviews a random selection of tissue samples from approved laboratories annually to ensure that quality standards are met.[86] Nonetheless, we have demonstrated important deviations from practice in confirmatory testing patterns, and therefore assert that it is reasonable to model similarly sub-optimal rates of test agreement. We propose that the higher disagreement may be a closer reflection of actual practice given other observed deviations, and hypothesize that the increased rates of IHC+ false positives will make any strategy including IHC less favourable than an ISH testing strategy alone. This is because IHC+ results are not typically confirmed with FISH in practice, resulting in higher rates of overtreatment it is assumed that IHC+/FISH- patients do not benefit from trastuzumab

therapy. We also tested this assumption by modelling minimum IHC+ and FISH+ agreement rates set out by Canadian guidelines.[8]

True positives made up a pooled mean of 54.1% of IHC equivocal results. The pooled analysis of IHC+ results included 1 study that observed no discordant events.[80] This study sample was relatively large in comparison to others in the pooled estimate, and thus contributed to 0.18 of the overall estimate. The IHC+/FISH+ agreement rate is lower than the minimum (90%) required for IHC+ results under the College of American Pathologists' guidelines for quality assurance of HER2 testing laboratories.[18] Although this estimate was derived through a rigorous systematic review of primary diagnostic studies, the lower agreement rate may raise questions of validity in Canadian practice. Indeed, HER2 testing in Ontario is subject to a quality assurance program that reviews a random selection of tissue samples from approved laboratories annually to ensure that quality standards are met. Nonetheless, we have demonstrated important deviations from practice in confirmatory testing patterns, and therefore assert that it is reasonable to model similarly sub-optimal rates of test agreement. We propose that the higher disagreement may be a closer reflection of actual practice given other observed deviations, and hypothesize that the increased rates of IHC+ false positives will make any strategy including IHC less favourable than an ISH testing strategy alone. This is because IHC+ results are not typically confirmed with FISH in

practice, resulting in higher rates of over-treatment it is assumed that IHC+/FISH- patients do not benefit from trastuzumab therapy.

Agreement between IHC and CISH or SISH both included the results of a single study that also confirmed results against the gold standard. These pooled analyses were further limited by small sample sizes and no observed discordance in some cases. As a result, we consistently applied the arcsin transformation for all probabilities where perfect agreement was observed. Agreement between IHC and CISH is shown in Table 4.7, while agreement with SISH is shown in Table 4.8. Results were divided showing IHC agreement within the subgroup of true positive and true negative samples. Concordance of IHC test results with CISH and SISH results tended to be high, although this might've been the result of small study samples. For example, the analysis of CISH demonstrated perfect agreement with IHC- and IHC+ results. Within the IHC equivocal category, CISH results were either all positive or all negative owing to the division of the sample by true status. The same was observed for the comparison between IHC and SISH, although there was some discordance in the IHC equivocal category.

Agreement between SISH (Table 4.5) or CISH (Table 4.6) with the gold standard FISH was estimated by the pooling 3 and 2 studies respectively. Although sample sizes were larger for this analysis, we observed perfect agreement in some studies. This is not surprising given the similarity among ISH testing techniques. For example, 98.6% of CISH- test results were true negatives by

FISH, while 93.5% of CISH+ were true positive by FISH. The agreement was stronger with CISH, as 99.9% of SISH+ being true positive and 92.6% being true negative by FISH.

The pooled agreement findings were used to inform the decision-tree model. We modeled the probabilities representing agreement (i.e. probability CISH+ among IHC+) by fitting a beta distribution to the corresponding pooled mean and standard error described in Table 4.4 through Table 4.8 (see also Paper 4 Table 2). The beta estimate was ideal for modelling the probability agreement of agreement as the distribution is bounded by zero and one.3

4.4.2 MODEL VALIDATION & CALIBRATION

4.4.2.1 Model Validation

The model predicted disease-specific BC mortality in perfect agreement with the stage IV BC mortality provided by CCO (Figure 4.2). This was verified by the GoF estimate of 0.00 across all model cycles for each arm of the model, which suggests no difference (error) between model predictions and expectations. This confirms that model programming of the influential disease-specific parameter is consistently programmed across each arm, eliminating the potential for programming errors to bias against any one test-treat strategy.

4.4.2.2 Model Calibration

The results of the stepwise calibration process are summarised in Table 4.1 and Figure 4.4. Prior to calibration, the model GoF was 0.14 over all model

cycles (Figure 4.4). The largest deviations observed using parameters as derived from the literature were from cycles 1 to 6 and 11 to 38, wherein the model under-predicted mortality vs. expectations.

We tested 3 deterministic scenarios (Table 4.1) prior to probabilistic optimisation of the p(recur|NED) parameter. Scenarios 1 and 2, which examined a reduction in literature-based $p(recur|NED)_{HER2-}$ probabilities relative to $p(recur|NED)_{HER2+}$, resulted in large differences from expectations as a result of under-estimated mortality in later cycles (not shown). However, the 3rd scenario produced mortality estimates much closer to expectation by adjusting the $p(recur|NED)_{HER2+}$ to twice that of $p(recur|NED)_{HER2+}$. This indicates that DFS in the control arms of the trastuzumab trials was lower than is typical of the Canadian population, assuming a disease prevalence of 20%. This is likely due to the selected inclusion of healthier patients, in combination with the bias produced by patient crossover, in the trials. However, the 3rd p(recur|NED) scenario was still associated with measurable deviation from expected all-cause mortality in earlier cycles, where mortality was lower than expectations (not shown). We therefore fit the 3rd scenario probabilities to narrow beta distributions for probabilistic calibration. This process identified a parameter set that improved model GoF over cycles 10 and 25 (Figure 4.3).

We optimised the p(DR|recur) and p(DR|LRR) parameters in subsequent steps of the model calibration process. Optimisation of these parameters produced

marginal improvements or reductions in GoF over cycles 10 and 25, while gradually improving model fit over all cycles (Table 4.1). The final set of optimised p(recur|NED), p(DR|recur) and p(DR|LRR) parameter inputs is provided in Appendix 4.3, Appendix 4.6 and Paper 4 Table 3. Overall model fit was optimised to a GoF of 0.0801 (Figure 4.4). The calibrated model produced better fit over cycles 10, 25 and all vs. the uncalibrated model. The model could not perfectly replicate expected survival, with mortality under-predicted over earlier cycles. However, the calibration process optimised key disease progression parameters to produce closer agreement with expectations. Both calibrated and uncalibrated model predictions differed from expectations notably after cycle 35, when patients reached age 85. This pattern was consistent in every scenario of the calibration process. This is the product of 3 factors: (1) an increase in general mortality rates associated with age (Appendix 4.8); (2) an increase in disease-specific mortality rates with age (Appendix 4.6); and, (3) parametric fitting of 8 year follow-up data to a Weibull model to extend expected survival data into the lifetime perspective.

The calibration process also revealed the strengths and inherent limitations of the BC Markov model and the data used to inform its disease progression parameters. We made structural assumptions about the disease states to include in the model and the way in which disease progression was modeled based on the best available evidence of the natural history of BC. The most influential of these assumptions was likely structural programming of the NED state as a 10-year

tunnel state. This approach was supported by the clinical literature, which characterises patients who survived 10 or more years without a recurrence as being in 'remission.'[29] This 10 year structural assumption was also seen as a reasonable compromise between the longer follow-up available for HER2- patients vs. the limited follow-up available for HER2+ patients. Finally, the 10-year tunnel state was seen as a compromise between achieving clinical accuracy and limiting model complexity. However, the assumption also enforces a constant probability of disease recurrence for all patients in the NED state beyond cycle 11. This is likely the cause of the model under- and over-predicting mortality during the calibration process. Deviations from expectations changed noticeably relative to cycle 11, with the uncalibrated model beginning to deviate substantially at cycle 11 while the calibrated model adhered more closely to expectations after cycle 11 (Figure 4.4). This structural assumption is therefore a key limitation of the model, and provides context to inform the interpretation of model results. We would caution the use of this model to make predictions about the course of BC disease progression, but instead suggest that the model can be used to test the impact of a series of clinical or policy assumptions in a cohort of patients with the select characteristics that were assumed during calibration.

4.4.3 ADDITIONAL MODEL FINDINGS & INTERPRETATION

This section focuses on results, and interpretations thereof, not addressed in Paper 4. For example, Paper 4 discusses the consequences of modelling lower

IHC+/FISH+ agreement than set out by quality assurance guidelines, while concerns about treatment effect estimates derived from adjuvant clinical trials with early-crossover are addressed here.

The effect of sequential trastuzumab treatment was considered in the base case and most sensitivity analyses, since this reflects current treatment guidelines. However, we also modeled the combined effect of sequential+concurrent treatment given recent evidence suggesting that as many as 30% of patients may receive concurrent therapy.[61] Sequential+concurrent treatment tended to increase the QALY gains associated with all test-treat strategies vs.[55] sequential+concurrent treatment tended to increase the QALY gains associated with all test-treat strategies vs. sequential alone (base case), though more so for strategies that produced the most accurate diagnoses (Table 4.9). Therefore, the QALY gains were larger across both deterministic and probabilistic analyses (d:0.016, p:0.011) for Strategy D (FISH) than in the base case (d:0.011, p:0.002).), respectively. Sequential+concurrent therapy was associated with smaller cost savings in the lifetime perspective vs. sequential therapy alone for all strategies. This can be attributed to more frequent transfusion costs and increased rates of cardiac toxicity associated with concurrent therapy. Despite the increased rate of cardiac toxicity, mortality due to cardiac complications were not observed in the trials, and thus suggest that improved outcomes could be achieved with concurrent therapy in patients with no history of cardiac disease. Unfortunately, we were unable to esti-

mate the consequences of trastuzumab effectiveness based on clinical practice. Although real-world trastuzumab effectiveness would be ideal for a model of current practice, these estimates are not available and may not be available for many years given that HER2 status was only recently added to the Ontario Cancer Registry.[87]

We made a number of testing and treatment-related assumptions in this model which bear consideration.[18] Guidelines recommend re-testing or reflex testing with an alternative modality in cases of equivocal results, but we did not model this scenario. Firstly, we treated FISH as the gold standard test that would determine either true positive or true negative status. We did not model equivocal FISH results, which account for approximately 2% of cases.[33] Guidelines recommend re-testing or reflex testing with an alternative modality in cases of equivocal results, but we did not model this scenario. Thus, this model does not capture all sequelae of FISH testing, and thus would tend to bias slightly in favour of FISH. Secondly, we assumed that only true positive (FISH+) patients would receive benefit from therapy, consistent with other models of HER2 testing and trastuzumab treatment.[1–3] Although clinical trials have demonstrated trastuzumab benefit in IHC 3+ patients, re-analysis of treatment response according to IHC and FISH status in the metastatic setting suggests that FISH was superior in predicting trastuzumab response. [33,82,88] We have therefore assumed that a similar pattern of gene over-expression was also the best predictor of treatment response in the ad-

juvant setting.[18] We have therefore assumed that a similar pattern of gene over-expression was also the best predictor of treatment response in the adjuvant setting. Indeed, many of the adjuvant trails included only FISH+ patients. Never the less, our assumption would tend to bias towards primary FISH testing. Thirdly, we did not account for retesting for HER2 at recurrence, nor did we address the consistency of HER2 status between incident diagnosis and recurrence. Current HER2 testing practice at disease progression is unknown, although a growing body of research suggests that as many as 10-15% patients may lose or gain HER2 overexpression at recurrence.[89,90] Current opinions on whether patients should be retested for HER2 at recurrence are mixed, and the cost-effectiveness of retesting is further complicated by questions about the responsiveness of tumours to trastuzumab after progression. A separate model would be better suited to answer questions about retesting and retreatment in a hypothetical framework. Finally, we did not address questions about HER2 testing in different sample types (e.g. core needle biopsy vs. excision) or failed FISH tests. The question of test accuracy in core needle biopsies is important in the context of neoadjuvant therapy, where trastuzumab is used to improve tumour resectability. This is another avenue of analysis that could be pursued pending the long-term follow-up of neo-adjuvant therapy studies.[17,19] Failed FISH tests due to sample abnormalities (e.g. polysomy, aneusomy) cannot always be clarified by IHC protein detection[33] and therefore represent a challenge for pathologists and clinicians.

However, This is the first economic evaluation of trastuzumab to be based on the highest level of treatment effect evidence, a Cochrane systematic review. Effect estimates from this particular review were subject to known clinical trial biases, particularly the effect of substantial patient crossover in select trials. Thus, estimates of sequential and sequential+concurrent treatment effect were influence by biases acting in opposition:

(1) Clinical trials tend to include a highly selected patient population, free of comorbidities and, more likely to respond to therapy with fewer side effects, leading to a bias in favour of the treatment effect; and

(2) Patient crossover from control to active treatment would tend to attenuate the treatment effect.

It is impossible to determine how strong either bias influenced the treatment effect without long-term follow-up of trastuzumab patients in the real-world setting. Therefore, it will be crucial to include pathologic tumour characteristics and genetic or genomic information in cancer registries to facilitate comparative effectiveness research in the future. Thirdly, we did not account for retesting for HER2 at recurrence, nor did we address the consistency of HER2 status between incident diagnosis and recurrence. Current HER2 testing practice at disease progression is unknown, although a growing body of research suggests that as many as 10-15% patients may lose or gain HER2 overexpression at recurrence.[83,84] Current opinions on whether patients should be retested for HER2 at recurrence are

mixed, and the cost-effectiveness of retesting is further complicated by questions about the responsiveness of tumors to trastuzumab after progression.

4.5 REFERENCES

1. Elkin EB, Weinstein MC, Winer EP, Kuntz KM, Schnitt SJ, Weeks JC. HER-2 testing and trastuzumab therapy for metastatic breast cancer: a cost-effectiveness analysis. J. Clin. Oncol. 2004 Mar 1;22(5):854–63.

2. Lidgren M, Jönsson B, Rehnberg C, Willking N, Bergh J. Cost-effectiveness of HER2 testing and 1-year adjuvant trastuzumab therapy for early breast cancer. Ann. Oncol. 2008 Mar;19(3):487–95.

3. Blank PR, Schwenkglenks M, Moch H, Szucs TD. Human epidermal growth factor receptor 2 expression in early breast cancer patients: a Swiss cost-effectiveness analysis of different predictive assay strategies. Breast Cancer Res. Treat. 2010 Nov;124(2):497–507.

4. Ferrusi IL, Leighl NB, Kulin N a, Marshall D a. Do economic evaluations of targeted therapy provide support for decision makers? J. Oncol. Pract. 2011 May;7(3 Suppl):36s–45s.

5. Weinstein MC. Recent developments in decision-analytic modelling for economic evaluation. Pharmacoeconomics. 2006 Jan;24(11):1043–53.

6. Karnon J, Vanni T. Calibrating models in economic evaluation: a comparison of alternative measures of goodness of fit, parameter search strategies and convergence criteria. Pharmacoeconomics. 2011 Jan;29(1):51–62.

7. Canadian Agency for Drugs and Technologies in Health. Guidelines for the Economic Evaluation of Health Technologies: Canada. 3rd ed. Ottawa: CADTH; 2006. p. 1–46.

8. Hanna W, O'Malley FP, Barnes P, Berendt R, Gaboury L, Magliocco A, et al. Updated recommendations from the Canadian National Consensus Meeting on HER2/neu testing in breast cancer. Curr. Oncol. 2007 Aug;14(4):149–53.

9. Roche PC, Ingle JN. Increased HER2 with U.S. Food and Drug Administration-approved antibody. J. Clin. Oncol. 1999 Jan;17(1):434.

10. Yang XR, Chang-Claude J, Goode EL, Couch FJ, Nevanlinna H, Milne RL, et al. Associations of breast cancer risk factors with tumor subtypes: a

pooled analysis from the Breast Cancer Association Consortium studies. J. Natl. Cancer Inst. 2011 Feb 2;103(3):250–63.

11. Trudeau M, Madarnas Y, McCready DR, Pritchard KI, Messersmith H. The Role of Trastuzumab in Adjuvant and Neoadjuvant Therapy in Women with HER2 / neu-overexpressing Breast Cancer : A Clinical Practice Guideline. Toronto: Cancer Care Ontario; 2007. p. 4–8.

12. Saphner T, Tormey DC, Gray R. Annual hazard rates of recurrence for breast cancer after primary therapy. J. Clin. Oncol. 1996 Oct 1;14(10):2738–46.

13. Moja L, Tagliabue L, Balduzzi S, Parmelli E, Pistotti V, Guarneri V, et al. Trastuzumab containing regimens for early breast cancer. Cochrane Breast Cancer Group, editor. Cochrane Database Syst. Rev. 4th ed. Chichester, UK: Cochrane Database of Systematic Reviews; 2012 Jan 1;4:CD006243.

14. Hecht SS, Hawley Berger A, Pandolfi PP, Howley PM, Ganem D, Kieff E, et al. Etiology of Cancer. In: Rosenberg SA, editor. DeVita, Hellman, Rosenberg's Cancer Princ. Pract. Oncol. 9th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2011. p. xlvii, 2638 p. :150–232.

15. Dendukuri N, Khetani K, McIsaac M, Brophy J. Testing for HER2-positive breast cancer: a systematic review and cost-effectiveness analysis. Can. Med. Assoc. J. 2007 May 8;176(10):1429–34.

16. Cuadros M, Villegas R. Systematic review of HER2 breast cancer testing. Appl. Immunohistochem. Mol. Morphol. 2009 Jan;17(1):1–7.

17. Hanna W, Barnes P, Berendt R, Chang M, Magliocco A, Mulligan AM, et al. Testing for her2 in breast cancer: current pathology challenges faced in Canada. Curr. Oncol. 2012 Dec;19(6):315–23.

18. Sauter G, Lee J, Bartlett JMS, Slamon DJ, Press MF. Guidelines for human epidermal growth factor receptor 2 testing: biologic and methodologic considerations. J. Clin. Oncol. 2009 Mar 10;27(8):1323–33.

19. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of

diagnostic accuracy included in systematic reviews. BMC Med. Res. Methodol. 2003 Nov 10;3:25.

20. Zar JH. Biostatistical analysis. 4th ed. Upper Saddle River, N.J.: Prentice Hall; 1999. p. 278–80.

21. Hartung J, Knapp G, Sinha BK. Methods of Combining Effect Sizes. Stat. Meta-Analysis With Appl. Chicester: Wiley; 2008.

22. Briggs AH, Ades AE, Price MJ. Probabilistic sensitivity analysis for decision trees with multiple branches: use of the Dirichlet distribution in a Bayesian framework. Med. Decis. Mak. 2003;23(4):341–50.

23. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science (80-.). 1987 Jan 9;235(4785):177–82.

24. Slamon DJ, Godolphin W, Jones L a, Holt J a, Wong SG, Keith DE, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science (80-.). 1989 May 12;244(4905):707–12.

25. Pietras RJ, Arboleda J, Reese DM, Wongvipat N, Pegram MD, Ramos L, et al. HER-2 tyrosine kinase pathway targets estrogen receptor and promotes hormone-independent growth in human breast cancer cells. Oncogene. 1995 Jun 15;10(12):2435–46.

26. Ferrusi IL, Marshall D a, Kulin N a, Leighl NB, Phillips K a. Looking back at 10 years of trastuzumab therapy: what is the role of HER2 testing? A systematic review of health economic analyses. Per. Med. 2009 Mar;6(2):193–215.

27. Hall PS, Hulme C, McCabe C, Oluboyede Y, Round J, Cameron DA. Updated cost-effectiveness analysis of trastuzumab for early breast cancer: a UK perspective considering duration of benefit, long-term toxicity and pattern of recurrence. Pharmacoeconomics. 2011 May;29(5):415–32.

28. Hedden L, O'Reilly S, Lohrisch C, Chia S, Speers C, Kovacic L, et al. Assessing the Real-World Cost-Effectiveness of Adjuvant Trastuzumab in HER-2/neu Positive Breast Cancer. Oncologist. 2012;17(2):164–71.

29. Guyatt GH, Rennie D, Evidence-Based Medicine Working Group. Users' guides to the medical literature : a manual for evidence-based clinical practice. Chicago, IL: AMA Press; 2002.

30. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. Lancet. Elsevier Ltd; 2012 Feb 4;379(9814):432–44.

31. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet. 2005;365(9472):1687–717.

32. Slamon D, Eiermann W, Robert N, Pienkowski T, Martin M, Press M, et al. Adjuvant trastuzumab in HER2-positive breast cancer. N. Engl. J. Med. 2011 Oct 6;365(14):1273–83.

33. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Web Appendix 1. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet. 2005;365(9472):(online).

34. Buzdar AU, Valero V, Ibrahim NK, Francis D, Broglio KR, Theriault RL, et al. Neoadjuvant therapy with paclitaxel followed by 5-fluorouracil, epirubicin, and cyclophosphamide chemotherapy and concurrent trastuzumab in human epidermal growth factor receptor 2-positive operable breast cancer: an update of the initial randomized study. Clin. Cancer Res. 2007 Jan 1;13(1):228–33.

35. Joensuu H, Kellokumpu-Lehtinen P-L, Bono P, Alanko T, Kataja V, Asola R, et al. Adjuvant docetaxel or vinorelbine with or without trastuzumab for breast cancer. N. Engl. J. Med. 2006 Feb 23;354(8):809–20.

36. Joensuu H, Kellokumpu-Lehtinen P-L, Bono P. Supplementary Appendix. Adjuvant docetaxel or vinorelbine with or without trastuzumab for breast cancer. N. Engl. J. Med. 2006;354(8):(online) http://www.nejm.org/doi/suppl/10.1056/NEJ.

37. Gianni L, Eiermann W, Semiglazov V, Manikhas A, Lluch A, Tjulandin S, et al. Neoadjuvant chemotherapy with trastuzumab followed by adjuvant trastuzumab versus neoadjuvant chemotherapy alone, in patients with HER2-positive locally advanced breast cancer (the NOAH trial): a randomised controlled superiority trial with a parallel HER. Lancet. Elsevier Ltd; 2010 Jan 30;375(9712):377–84.

38. Harris EER, Hwang W-T, Lee E a, Cengel K a, Feldman MD, Demichele A, et al. The impact of HER-2 status on local recurrence in women with stage I-II breast cancer treated with breast-conserving therapy. Breast J. 2006;12(5):431–6.

39. Mechera R, Viehl CT, Oertli D. Factors predicting in-breast tumor recurrence after breast-conserving surgery. Breast Cancer Res. Treat. 2009 Jul;116(1):171–7.

40. Onitilo AA, Engel JM, Greenlee RT, Mukesh BN. Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. Clin. Med. Res. 2009 Jun;7(1-2):4–13.

41. Pritchard KI, Gelmon KA, Rayson D, Provencher L, Webster M, McLeod D, et al. Endocrine therapy for postmenopausal women with hormone receptor-positive her2-negative advanced breast cancer after progression or recurrence on nonsteroidal aromatase inhibitor therapy: a Canadian consensus statement. Curr. Oncol. 2013 Feb;20(1):48–61.

42. De Bock GH, Putter H, Bonnema J, van der Hage JA, Bartelink H, van de Velde CJ. The impact of loco-regional recurrences on metastatic progression in early-stage breast cancer: a multistate model. Breast Cancer Res. Treat. 2009 Sep;117(2):401–8.

43. Montagna E, Bagnardi V, Rotmensz N, Viale G, Renne G, Cancello G, et al. Breast cancer subtypes and outcome after local and regional relapse. Ann. Oncol. 2012 Feb;23(2):324–31.

44. Statistics Canada. Table 2b Complete life table, Canada, 2000 to 2002: females. Ottawa: Statistics Canada; 2012.

45. Ontario Cancer Registry. Toronto: Cancer Care Ontario; 2012.

46. Petrelli F, Barni S. Meta-analysis of concomitant compared to sequential adjuvant trastuzumab in breast cancer: the sooner the better. Med. Oncol. 2012 Jun;29(2):503–10.

47. Yin W, Jiang Y, Shen Z, Shao Z, Lu J. Trastuzumab in the adjuvant treatment of HER2-positive early breast cancer patients: a meta-analysis of published randomized controlled trials. PLoS One. 2011 Jan;6(6):e21030.

48. Chen T, Xu T, Li Y, Liang C, Chen J, Lu Y, et al. Risk of cardiac dysfunction with trastuzumab in breast cancer patients: a meta-analysis. Cancer Treat. Rev. Elsevier Ltd; 2011 Jun;37(4):312–20.

49. Dahabreh IJ, Linardou H, Siannis F, Fountzilas G, Murray S. Trastuzumab in the adjuvant treatment of early-stage breast cancer: a systematic review and meta-analysis of randomized controlled trials. Oncologist. 2008 Jun;13(6):620–30.

50. Viani G a, Afonso SL, Stefano EJ, De Fendi LI, Soares F V. Adjuvant trastuzumab in the treatment of her-2-positive early breast cancer: a meta-analysis of published randomized trials. BMC Cancer. 2007 Jan;7:153.

51. Fabi A, Metro G, Ferretti G, Giannarelli D, Di Cosimo S, Papaldo P, et al. Do HER-2 positive metastatic breast cancer patients benefit from the use of trastuzumab beyond disease progression? A mono-institutional experience and systematic review of observational studies. Breast. 2008 Oct;17(5):499–505.

52. Smith I, Procter M, Gelber RD, Guillaume S, Feyereislova A, Dowsett M, et al. 2-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer: a randomised controlled trial. Lancet. 2007 Jan 6;369(9555):29–36.

53. Spielmann M, Roché H, Delozier T, Canon J-L, Romieu G, Bourgeois H, et al. Trastuzumab for patients with axillary-node-positive breast cancer: results of the FNCLCC-PACS 04 trial. J. Clin. Oncol. 2009 Dec 20;27(36):6129–34.

54. Mackey JR, Clemons M, Côté M a, Delgado D, Dent S, Paterson A, et al. Cardiac management during adjuvant trastuzumab therapy: recommendations of the Canadian Trastuzumab Working Group. Curr. Oncol. 2008 Jan;15(1):24–35.

55. Ng D, Ferrusi I, Khong H, Earle CC, Trudeau M, Marshall D, et al. Cardiac monitoring during adjuvant trastuzumab therapy for breast cancer (poster). San Antonio Breast Cancer Symp. San Antonio; 2012.

56. Romond EH, Perez E a, Bryant J, Suman VJ, Geyer CE, Davidson NE, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N. Engl. J. Med. 2005 Oct 20;353(16):1673–84.

57. Statistics Canada. Table 326-0021 (Health and personal care) - Consumer Price Index (CPI), 2009 basket, annual. Statistics Canada; 2012.

58. Hanna WM. HER2 test costs (personal communication). 2012.

59. Herceptin (trastuzumab) Product monograph. Mississauga: Hoffmann-La Roche Ltd.; 2008. p. 62.

60. Leighl NB. trastuzumab unit price & vial wastage (personal communication). 2012.

61. Ontario Ministry of Health and Long-Term Care. Schedule of Benefits for Physician Services. Toronto: Government of Ontario; 2012.

62. Mittmann N, Verma S, Koo M, Alloul K, Trudeau M. Cost effectiveness of TAC versus FAC in adjuvant treatment of node-positive breast cancer. Curr. Oncol. 2010 Feb;17(1):7–16.

63. Drucker A, Skedgel C, Virik K, Rayson D, Sellon M, Younis T. The cost burden of trastuzumab and bevacizumab therapy for solid tumours in Canada. Curr. Oncol. 2008 Jun;15(3):136–42.

64. Levy a R, Briggs a H, Demers C, O'Brien BJ. Cost-effectiveness of beta-blocker therapy with metoprolol or with carvedilol for treatment of heart failure in Canada. Am. Heart J. 2001 Sep;142(3):537–43.

65. Will BP, Berthelot JM, Le Petit C, Tomiak EM, Verma S, Evans WK. Estimates of the lifetime costs of breast cancer treatment in Canada. Eur. J. Cancer. 2000 Apr;36(6):724–35.

66. Skedgel C, Rayson D, Younis T. The cost-utility of sequential adjuvant trastuzumab in women with Her2/Neu-positive breast cancer: an analysis based on updated results from the HERA Trial. Value Heal. International

Society for Pharmacoeconomics and Outcomes Research (ISPOR); 2009;12(5):641–8.

67. Sorensen S, Brown R, Benedict A, Flood E, Revicki D. QL4 Patient-Rated Utilities in Postmenopausal Early Breast Cancer (EBC):A Cross-Country Comparison. Value Heal. Elsevier; 2004 Nov 1;7(6):641–2.

68. Miller G, Randolph S, Forkner E, Smith B, Galbreath AD. Long-term cost-effectiveness of disease management in systolic heart failure. Med. Decis. Mak. 2009;29(3):325–33.

69. Skedgel C, Rayson D, Dewar R, Younis T. Cost-utility of adjuvant hormone therapies with aromatase inhibitors in post-menopausal women with breast cancer: upfront anastrozole, sequential tamoxifen-exemestane and extended tamoxifen-letrozole. Breast. 2007 Jun;16(3):252–61.

70. Clarke LD, Plevritis SK, Boer R, Cronin K a, Feuer EJ. A comparative review of CISNET breast models used to analyze U.S. breast cancer incidence and mortality trends. J. Natl. Cancer Inst. Monogr. 2006 Jan;98264(36):96–105.

71. Vanni T, Karnon J, Madan J, White RG, Edmunds WJ, Foss AM, et al. Calibrating models in economic evaluation: a seven-step approach. Pharmacoeconomics. 2011 Jan;29(1):35–49.

72. Karnon J, Goyder E, Tappenden P, McPhie S, Towers I, Brazier J, et al. A review and critique of modelling in prioritising and designing screening programmes. Health Technol. Assess. (Rockv). 2007 Dec;11(52):iii–iv, ix–xi, 1–145.

73. Kleinbaum DG, Klein M. Kaplan-Meier Survival Curves and the Log-Rank Test. In: Gail M, Krickberg K, Samet J, Tsiatis A, Wong W, editors. Surviv. Anal. A Self-Learning Text. 2nd ed. New York: Springer; 2005. p. 45–82.

74. Kleinbaum DG, Klein M. Parametric Survival Models. In: Gail M, Krickberg K, Samet J, Tsiatis A, Wong W, editors. Surviv. Anal. A Self-Learning Text. 2nd ed. New York: Springer; 2005. p. 257–330.

75. Kurian AW, Thompson RN, Gaw AF, Arai S, Ortiz R, Garber AM. A cost-effectiveness analysis of adjuvant trastuzumab regimens in early

HER2/neu-positive breast cancer. J. Clin. Oncol. 2007 Feb 20;25(6):634–41.

76. Panjwani P, Epari S, Karpate A, Shirsat H, Rajsekharan P, Basak R, et al. Assessment of HER-2/neu status in breast cancer using fluorescence in situ hybridization & immunohistochemistry: Experience of a tertiary cancer referral centre in India. Indian J. Med. Res. 2010 Sep;132(September):287–94.

77. Riethdorf S, Hoegel B, John B, Ott G, Fritz P, Thon S, et al. Prospective multi-centre study to validate chromogenic in situ hybridisation for the assessment of HER2 gene amplification in specimens from adjuvant and metastatic breast cancer patients. J. Cancer Res. Clin. Oncol. 2011 Mar;137(2):261–9.

78. Ahmed SS, Iqbal J, Thike AA, Lim AST, Lim TH, Tien SL, et al. HER2/neu revisited: quality and interpretive issues. J. Clin. Pathol. 2011 Mar;64(2):120–4.

79. Goud KI, Dayakar S, Vijayalaxmi K, Babu SJ, Reddy PVA. Evaluation of HER-2/neu status in breast cancer specimens using immunohistochemistry (IHC) & fluorescence in-situ hybridization (FISH) assay. Indian J. Med. Res. 2012 Mar;135(March):312–7.

80. Vergara-Lluri ME, Moatamed N a, Hong E, Apple SK. High concordance between HercepTest immunohistochemistry and ERBB2 fluorescence in situ hybridization before and after implementation of American Society of Clinical Oncology/College of American Pathology 2007 guidelines. Mod. Pathol. Nature Publishing Group; 2012 Oct;25(10):1326–32.

81. Cancer Care Ontario. Surgical Oncology News. Ontario CC, editor. Surg. Oncol. Newsl. Toronto; 2009;4(1).

82. Mass RD, Press MF, Anderson S, Cobleigh MA, Vogel CL, Dybdal N, et al. Evaluation of clinical outcomes according to HER2 detection by fluorescence in situ hybridization in women with metastatic breast cancer treated with trastuzumab. Clin. Breast Cancer. 2005 Aug;6(3):240–6.

83. Wilking U, Karlsson E, Skoog L, Hatschek T, Lidbrink E, Elmberger G, et al. HER2 status in a population-derived breast cancer cohort: discordances

during tumor progression. Breast Cancer Res. Treat. 2011 Jan;125(2):553–61.

84. Nishimura R, Osako T, Okumura Y, Tashima R, Toyozumi Y, Arima N. Changes in the ER, PgR, HER2, p53 and Ki-67 biological markers between primary and recurrent breast cancer: discordance rates and prognosis. World J. Surg. Oncol. BioMed Central Ltd; 2011 Jan;9(1):131.

85. Al-Khattabi H, Kelany A, Buhmeida A, Al-Maghrabi J, Lari S, Chaudhary A, et al. Evaluation of HER-2/neu gene amplification by fluorescence in situ hybridization and immunohistochemistry in saudi female breast cancer. Anticancer Res. 2010 Oct;30(10):4081–8.

86. Kiyose S, Igarashi H, Nagura K, Kamo T, Kawane K, Mori H, et al. Chromogenic in situ hybridization (CISH) to detect HER2 gene amplification in breast and gastric cancer: comparison with immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH). Pathol. Int. 2012 Nov;62(11):728–34.

87. Lee Y, Ryu Y, Jeong H, Chang H, Kim Y, Kim A. Effectiveness of silver-enhanced in situ hybridization for evaluating HER2 gene status in invasive breast carcinoma: a comparative study. Arch. Med. Res. Elsevier Inc; 2012 Mar;43(2):139–44.

88. Mollerup J, Henriksen U, Müller S, Schønau A. Dual color chromogenic in situ hybridization for determination of HER2 status in breast cancer: a large comparative study to current state of the art fluorescence in situ hybridization. BMC Clin. Pathol. BioMed Central Ltd; 2012 Jan;12(1):3.

89. Panjwani P, Epari S, Karpate A, Shirsat H, Rajsekharan P, Basak R, et al. Assessment of HER-2/neu status in breast cancer using fluorescence in situ hybridization & immunohistochemistry: Experience of a tertiary cancer referral centre in India. Indian J. Med. Res. 2010 Sep;132(September):287–94.

90. Schiavon BN, Jasani B, de Brot L, Vassallo J, Damascena A, Cirullo-Neto J, et al. Evaluation of reliability of FISH versus brightfield dual-probe in situ hybridization (BDISH) for frontline assessment of HER2 status in breast cancer samples in a community setting: influence of poor tissue preservation. Am. J. Surg. Pathol. 2012 Oct;36(10):1489–96.

91. Sui W, Ou M, Chen J, Wan Y, Peng H, Qi M, et al. Comparison of immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH) assessment for Her-2 status in breast cancer. World J. Surg. Oncol. 2009 Jan;7(83):1–6.

92. Zhu X, Lu Y, Lu H, Yang W, Tu X, Cai X, et al. Genetic alterations and protein expression of HER2 and chromosome 17 polysomy in breast cancer. Hum. Pathol. Elsevier Inc.; 2011 Oct;42(10):1499–504.

4.6 TABLES

Table 4.1. Comparison of pre-specified calibration scenarios describing the rationale for relevant scenarios, which parameters were permitted to vary probabilistically vs. those that were deterministic at their uncalibrated or calibrated mean values, the best-fitting trial and the associated least-squares goodness of fit (GoF) estimate over all cycles, up to and including cycles 25 and 10. * indicates that secondary GoF estimates were not estimated owing to poor fit across all cycles. Abbreviations: DR: distant recurrence; EBCTCG: Early-Breast Cancer Trialists' Collaborative Group; HER2: human epidermal growth factor receptor-2; LRR: locoregional recurrence; +: positive; -:negative; SE: standard error

Parameter, parameter type & calibration assumptions						Least squares goodness of fit estimates across cycles:		
Scenario	p(recur|NED)	p(DR|recur)	p(DR|LRR)	Rationale	Best trial	all	≤ 25	≤ 10
1	deterministic; set HER2- values to ½ of HER2+ for years ≥3	deterministic	deterministic	HER2+ rates should be twice HER2- and steadily decline after year 2	391	0.3866	*	*
2	deterministic; set HER2- values to ½ of HER2+ for all years	deterministic	deterministic	HER2+ rates should be 2x HER2-;rates should steadily decrease after year 2	836	0.4983	*	*
3	deterministic; set HER2+ values to 2x HER2- for all years	deterministic	deterministic	HER2+ rates should be 2x HER2-;rates should steadily decrease after year 2	604	0.0760	0.0062	0.0039
4	probabilistic; fit narrow beta to the best fitting parameter set from scenarios 1-3 (scenario 3 chosen)	deterministic	deterministic	assumed SE from EBCTCG data for calibration since we assumed that HER2+ was 2x HER2- data from EBCTCG	863	0.0810	0.0047	0.0025
5	deterministic from scenario 4	probabilistic; fit narrow beta	deterministic		871	0.0803	0.0049	0.0027
6	deterministic from scenario 4	deterministic from Scenario 5	probabilistic; fit narrow beta		558	0.0801	0.0047	0.0028

Table 4.2. Characteristics of studies included in the review of HER2 test agreement. Age is represented as a median, unless the standard deviation is also noted. Age ranges are provided where reported. All study sample sizes shown are prior to sample attrition. Abbreviations: CISH: chromogenic *in situ* hybridisation; FFPE: formalin fixed paraffin embedded; FISH: fluorescence *in situ* hybridisation; IHC: immunohistochemistry; NR: not reported.

Study	Country	Period of Recruitment	Age	Female Proportion	Disease Stage	Study Sample Size	Laboratory Setting	Tissue Type	Tests Compared	Specific Tests or Antibodies
Ahmed 2011 [83]	Singapore	2007 - 2008	NR	NR	NR	106	Hospital	FFPE unknown	IHC CISH FISH	PathVysion clone SP3
Al-Khattabi 2010 [91]	Saudi Arabia	2002 -2008	52 (SD 13.14) range 24-81	NR	NR	75	NR	FFPE unknown	IHC FISH	HercepTest PathVysion
Goud 2012 [84]	India	Jul 2008 - Apr 2009	48 (SD13.7) range 30-50	100%	NR	90	Hospital	FFPE unknown	IHC FISH	PathVysion
Kiyose 2012 [92]	Japan	NR	NR	NR	NR	125	University medical center	FFPE unknown	IHC CISH FISH	HISTRA, HercepTest, PathVysion
Lee 2012 [93]	Korea	1992 - 2004	NR	NR	NR	543	University medical center	FFPE tumour excision	IHC SISH FISH	CB11, PathVysion, INFORM, Ultraview
Mollerup 2012 [94]	USA	NR	NR	NR	NR	365	Reference laboratory	FFPE unknown	IHC CISH FISH	HercepTest, pharmDx, PathVysion
Panjwani 2010 [95]	India	Apr 2006 - Sept 2008	range 29-78	99%	NR	200	NR	FFPE unknown	IHC FISH	3B5

Study	Country	Period of Recruitment	Age	Female Proportion	Disease Stage	Study Sample Size	Laboratory Setting	Tissue Type	Tests Compared	Specific Tests or Antibodies
Riethdorf 2011[82]	Germany	2003 - 2004	62 range 28-94	NR	9% IV	403	NR	FFPE tumour excision & core biopsies (<1%)	IHC CISH FISH	HercepTest PathVysion SPOT-Light
Schiavon 2012 [96]	Brazil	1990 - 2005	55 range 32-77	NR	I, II, III	320	Hospital	FFPE unknown	IHC CISH FISH	4B5, PathVysion, DuoCISH, ZytoVision
Sui 2009 [97]	China	July 2008 - Mar 2009	NR	NR	NR	50	Cancer cen-ter	FFPE unknown	IHC FISH	S-P Kit
Vergara-Lluri 2012 [85]	USA	2003-2010	NR	NR	NR	1473	University medical center	FFPE unknown	IHC FISH	AO485, HercepTest, PathVysion
Zhu 2011 [98]	China	Jun 2007 - Oct 2008	NR	NR	NR	481	Hospital	FFPE unknown	IHC FISH	HercepTest PathVysion

Table 4.3. Quality attributes of each study included in the systematic review and meta-analysis of HER2 test agreement and accuracy. Green indicators represent limited risk of bias, yellow indicators moderate risk of bias and red indicators represent high risk of bias. Abbreviations: NR: not reported. [careful - not sure that colour can be used for your thesis – I think everything is balck and white]

Study	Perspective	Sample Selection	Blinding	% of Sample Verified	Attrition	Overall Bias Indicator
Ahmed [83]	prospective	NR	blinded	100%	none	
Al-Khattabi [91]	NR	randomised	NR	100%	none	
Goud [84]	retrospective	consecutive	blinded	100%	none	
Kiyose [92]	NR	NR	blinded	100%	none	
Lee [93]	retrospective	consecutive	NR	100%	none	
Mollerup [94]	retrospective	consecutive	blinded	≥90%	explained	
Panjwani [95]	prospective	NR	NR	≥90%	explained	
Riethdorf [82]	prospective	NR	partly-blinded	100%	none	
Schiavon [96]	retrospective	NR	blinded	<90%	explained	
Sui [97]	prospective	NR	blinded	100%	none	
Vergara-Lluri [85]	retrospective	NR	NR	100%	none	
Zhu [98]	NR	NR	NR	≥90%	none	

Table 4.4. Study-level and pooled agreement between IHC and gold standard (FISH) test results to determine HER2 status. Agreement here is represented as the probability of being FISH+ or FISH-. All results were pooled by the inverse variance method of weighting individual study results. [might define this here if not in text? Please check] The arcsin transformation was used to accommodate studies that observed no events. Note that the probabilities across each category may not sum to 1 due to the effect of pooling multiple studies. The review was limited to studies that used the Canadian and ASCO/CAP cut-off for ISH positivity, a HER2:CEP17 gene ratio >2.2 or gene copy number >6. Abbreviations: eq: equivocal; FISH: fluorescence in situ hybridisation; IHC: immunohistochemistry; SE: standard error; +: positive; -: negative.

	IHC negative (0,1+)			IHC equivocal (2+)			IHC positive (3+)		
Study	**IHC- Sample Size**	**p(FISH+)**	**p(FISH-)**	**IHCeq Sample Size**	**p(FISH+)**	**p(FISH-)**	**IHC+ Sample Size**	**p(FISH+)**	**p(FISH-)**
Ahmed [83]	15	0.0000	1.0000	62	0.1774	0.7419	29	0.8276	0.0484
Al-Khattabi [91]	44	0.1591	0.8409	13	0.2308	0.7692	18	0.5000	0.6923
Goud [84]	30	0.0000	1.0000	28	0.7143	0.2857	32	0.7813	0.2188
Kiyose [92]	71	0.0282	0.9718	17	0.7647	0.2353	37	0.9730	0.0270
Lee [93]	372	0.0349	0.9651	59	0.4915	0.3559	112	0.8839	0.0179
Panjwani [95]	57	0.1228	0.8596	36	0.6667	0.3056	82	0.9390	0.0833
Riethdorf [82]	217	0.0138	0.9862	132	0.1970	0.8030	54	0.7593	0.0985
Sui [97]	26	0.1923	0.8077	10	0.7000	0.3000	14	0.9286	0.1000
Vergara-Lluri [85]	777	0.0270	0.9730	121	0.4298	0.5702	118	1.0000	0.0000
Zhu [98]	131	0.2595	0.7252	184	0.5652	0.4185	166	0.8313	0.1467
Total:	**1,740**			**662**			**662**		
Pooled mean:		**0.0424**	**0.9567**		**0.4448**	**0.5407**		**0.8777**	**0.0758**
SE:		**0.0159**	**0.0159**		**0.0159**	**0.0279**		**0.0279**	**0.0279**

Table 4.5. Study-level and pooled agreement between CISH and gold standard (FISH) test results to determine HER2 status. Agreement here is represented as the probability of being FISH+ or FISH-. All results were pooled by the inverse variance method of weighting individual study results. The arcsin transformation was used to accommodate studies that observed no events. Note that the probabilities across each category may not sum to 1 due to the effect of pooling multiple studies. The review was limited to studies that used the Canadian and ASCO/CAP cut-off for ISH positivity, a HER2:CEP17 gene ratio >2.2 or gene copy number >6. This resulted in the inclusion of only studies examining dual-colour CISH. Abbreviations: CISH: chromogenic *in situ* hybridisation; FISH: fluorescence *in situ* hybridisation; IHC: immunohistochemistry; SE: standard error; +: positive; -: negative.

	CISH Negative			**CISH Positive**		
Study	**CISH- Sample Size**	**p(FISH+)**	**p(FISH-)**	**CISH+ Sample Size**	**p(FISH+)**	**p(FISH-)**
Kiyose [92]	74	0.0000	1.0000	51	1.0000	0.0000
Mollerup [94]	306	0.0131	0.9869	44	0.9091	0.0909
Schiavon [96]	64	0.0469	0.9531	52	0.8462	0.1538
Total:	**74**			**51**		
Pooled mean:		**0.0144**	**0.9856**		**0.9346**	**0.0654**
SE:		**0.0004**	**0.0004**		**0.0013**	**0.0013**

Table 4.6. Study-level and pooled agreement between SISH and gold standard (FISH) test results to determine HER2 status. Agreement here is represented as the probability of being FISH+ or FISH-. All results were pooled by the inverse variance method of weighting individual study results. The arcsin transformation was used to accommodate studies that observed no events. Note that the probabilities across each category may not sum to 1 due to the effect of pooling multiple studies. The review was limited to studies that used the Canadian and ASCO/CAP cut-off for ISH positivity, a HER2:CEP17 gene ratio >2.2 or gene copy number >6. Abbreviations: FISH: fluorescence *in situ* hybridisation; IHC: immunohistochemistry; SE: standard error; SISH: silver *in situ* hybridisation;+: positive; -: negative.

	SISH Negative			**SISH Positive**		
Study	**SISH- Sample Size**	**p(FISH+)**	**p(FISH-)**	**SISH+ Sample Size**	**p(FISH+)**	**p(FISH-)**
Lee [93]	366	0.0000	1.0000	150	0.9000	0.1000
Schiavon [96]	135	0.0000	1.0000	36	1.0000	0.0000
Total:	**501**			**186**		
Pooled mean:		**0.0009**	**0.9991**		**0.9259**	**0.0741**
SE:		**0.0002**	**0.0002**		**0.0006**	**0.0006**

Table 4.7. Agreement between IHC and CISH among studies that verified test results against the gold standard (FISH) to determine HER2 status. Agreement here is represented as the probability of being CISH+ or CISH-. Agreement is shown separately for true positive (FISH+) and true negative (FISH-) samples. All results were pooled by the inverse variance method of weighting individual study results. The arcsin transformation was used to accommodate studies that observed no events. Note that the probabilities across each category may not sum to 1 due to the effect of pooling multiple studies. The review was limited to studies that used the Canadian and ASCO/CAP cut-off for ISH positivity, a HER2:CEP17 gene ratio >2.2 or gene copy number >6. This resulted in the inclusion of only studies examining dual-colour CISH. Abbreviations: CISH: chromogenic *in situ* hybridisation; FISH: fluorescence *in situ* hybridisation; IHC: immunohistochemistry; SE: standard error; +: positive; -: negative.

		IHC Negative (0,1+)			**IHC Equivocal (2+)**			**IHC Positive (3+)**		
Study	**FISH Sample Size**	**IHC- Sample Size**	**p(CISH+)**	**p(CISH-)**	**IHCeq Sample Size**	**p(CISH+)**	**p(CISH-)**	**IHC+ Sample Size**	**p(CISH+)**	**p(CISH-)**
True Positive Samples (FISH+)										
Kiyose [92]	51	2	1.0000	0.0000	13	1.0000	0.0000	36	1.0000	0.0000
Total:	**51**	**2**			**13**			**36**		
Transformed mean:			**0.8750**	**0.1250**		**0.9808**	**0.0192**		**0.9931**	**0.0069**
SE:			**0.0309**	**0.0309**		**0.0048**	**0.0048**		**0.0017**	**0.0017**
True Negative Samples (FISH-)										
Kiyose [92]	74	69	0.0000	1.0000	4	0.0000	1.0000	1	0.0000	1.0000
Total:	**74**	**69**			**4**			**1**		
Transformed mean:			**0.0036**	**0.9964**		**0.0625**	**0.9375**		**0.2500**	**0.7500**
SE:			**0.0009**	**0.0009**		**0.0155**	**0.0155**		**0.0612**	**0.0612**

Table 4.8. Agreement between IHC and SISH among studies that verified test results against the gold standard (FISH) to determine HER2 status. Agreement here is represented as the probability of being SISH+ or SISH-. Agreement is shown separately for true positive (FISH+) and true negative (FISH-) samples. All results were pooled by the inverse variance method of weighting individual study results. The arcsin transformation was used to accommodate studies that observed no events. Note that the probabilities across each category may not sum to 1 due to the effect of pooling multiple studies. The review was limited to studies that used the Canadian and ASCO/CAP cut-off for ISH positivity, a HER2:CEP17 gene ratio >2.2 or gene copy number >6. Abbreviations: FISH: fluorescence *in situ* hybridisation; IHC: immunohistochemistry; SE: standard error; SISH: silver *in situ* hybridisation;+: positive; -: negative.

		IHC Negative (0,1+)			IHC Equivocal (2+)			IHC Positive (3+)		
Study	**FISH Sample Size**	**IHC- Sample Size**	**p(SISH+)**	**p(SISH-)**	**IHCeq Sample Size**	**p(SISH+)**	**p(SISH-)**	**IHC+ Sample Size**	**p(SISH+)**	**p(SISH-)**
True Positive Samples (FISH+)										
Lee [93]	141	13	0.0000	1.0000	29	1.0000	0.0000	99	1.0000	0.0000
Total:	141	13			29			99		
Transformed mean:			**0.0192**	**0.9808**		**0.9914**	**0.0086**		**0.9975**	**0.0025**
SE:			**0.0048**	**0.0048**		**0.0022**	**0.0022**		**0.0006**	**0.0006**
True Negative Samples (FISH-)										
Lee [93]	382	359	0.0000	1.0000	21	0.6190	0.3810	2	0.0000	1.0000
Total:	**382**	**359**			**21**			**2**		
Transformed mean:			**0.0007**	**0.9993**		**0.6190**	**0.3810**		**0.8750**	**0.1250**
SE:			**0.0002**	**0.0002**		**0.0902**	**0.0030**		**0.0309**	**0.0309**

Table 4.9. Comparison of costs, life-years (LY), quality-adjusted life years (QALYs) and incremental outcomes associated with each strategy across different sequential and sequential+concurrent trastuzumab treatment benefit estimates. Deterministic and probabilistic mean results are shown. Abbreviations: FISH: fluorescence *in situ* hybridisation; IHC: immunohistochemistry; SISH: silver *in situ* hybridisation

Strategy	Costs	LYs	QALYs	Incremental Costs	Incremental LYs	Incremental QALYs	Incremental cost/LY	Incremental cost/QALY
Deterministic base case: sequential therapy								
A: IHC (FISH)	$44,840	11.58	10.55					
B: IHC (CISH)	$45,337	11.58	10.55	$497	-0.00076	-0.00111	dominated	dominated
C: IHC (SISH)	$50,556	11.58	10.55	$5,715	-0.00034	-0.00374	dominated	dominated
D: FISH	$43,939	11.59	10.56	-$902	0.00973	0.01117	-$92,693	-$80,728
E: CISH	$43,936	11.59	10.56	-$905	0.00702	0.00823	-$128,885	-$109,937
F: SISH	$45,364	11.59	10.56	$524	0.00939	0.00996	ext dominated	ext dominated
Deterministic sensitivity analysis: sequential+concurrent therapy								
A: IHC (FISH)	$43,951	11.63	10.60					
B: IHC (CISH)	$44,422	11.63	10.60	$472	-0.00107	-0.00146	dominated	dominated
C: IHC (SISH)	$49,369	11.63	10.60	$5,418	-0.00048	-0.00395	dominated	dominated
D: FISH	$43,092	11.64	10.62	-$859	0.01376	0.01555	-$62,400	-$55,211
E: CISH	$43,092	11.64	10.62	-$859	0.00993	0.01140	-$86,479	-$75,358
F: SISH	$44,443	11.64	10.62	$493	0.01328	0.01418	ext dominated	ext dominated
Probabilistic base case: sequential therapy								
A: IHC (FISH)	$44,308	11.61	10.60					
B: IHC (CISH)	$44,765	11.61	10.57	$458	0.00083	-0.02359	$549,537	dominated
C: IHC (SISH)	$49,600	11.61	10.58	$5,292	0.00234	-0.01610	$2,264,839	dominated
D: FISH	$43,493	11.62	10.60	-$815	0.01097	0.00217	-$74,272	-$374,766
E: CISH	$43,449	11.62	10.60	-$859	0.01022	-0.00036	ext dominated	ext dominated
F: SISH	$44,794	11.62	10.59	$486	0.01277	-0.00924	$38,038	dominated

Strategy	Costs	LYs	QALYs	Incremental Costs	Incremental LYs	Incremental QALYs	Incremental cost/LY	Incremental cost/QALY
Probabilistic sensitivity analysis: sequential+concurrent therapy								
A: IHC (FISH)	$43,510	11.66	10.65					
B: IHC (CISH)	$43,953	11.66	10.65	$443	0.00043	0.00164	$1,029,327	dominated
C: IHC (SISH)	$48,489	11.66	10.63	$4,978	0.00125	-0.01847	$3,973,959	dominated
D: FISH	$42,725	11.67	10.66	-$785	0.01679	0.01063	-$46,738	-$73,831
E: CISH	$42,729	11.67	10.66	-$781	0.01047	0.00481	-$74,560	ext dominated
F: SISH	$43,969	11.67	10.67	$459	0.01518	0.01829	$30,214	dominated

4.7 FIGURES

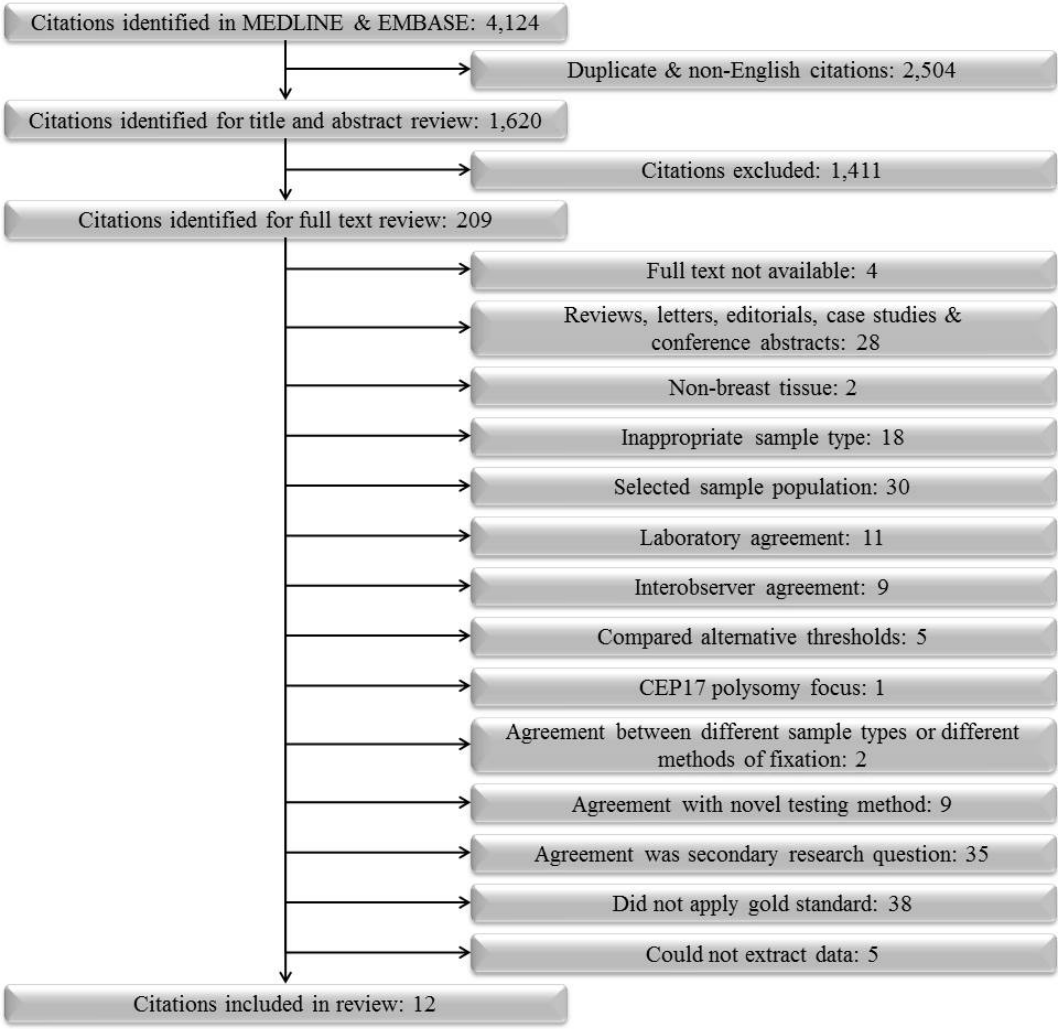


Figure 4.1. Quorum diagram representing the flow of citations through the review process. Reasons for exclusion and the number of excluded citations are shown for full text review only.

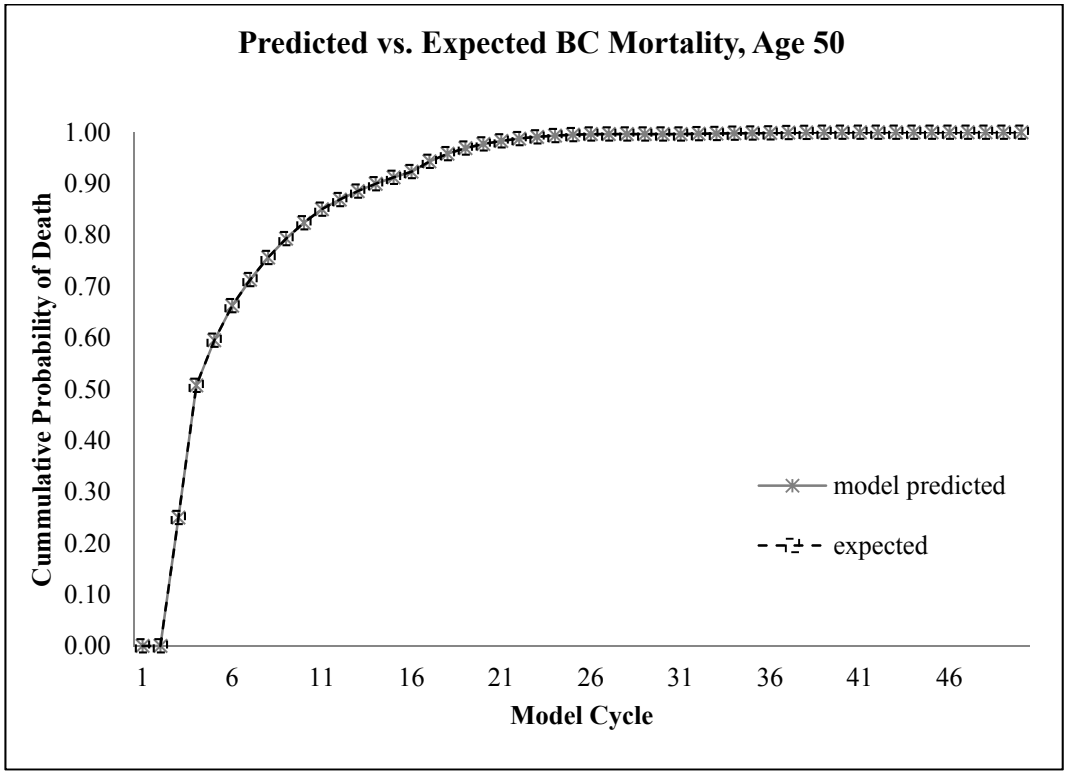


Figure 4.2. Plot of mortality due to breast cancer from the distant recurrence disease state, comparing cumulative expected mortality based on data provided by Cancer Care Ontario [23] and mortality predicted by the BC natural history model arm A. The figure demonstrates perfect agreement between predicted and expected mortality, indicating that the model functioned appropriately when predicting disease-specific mortality. This perfect agreement was consistent across all arms of the model (data not shown).

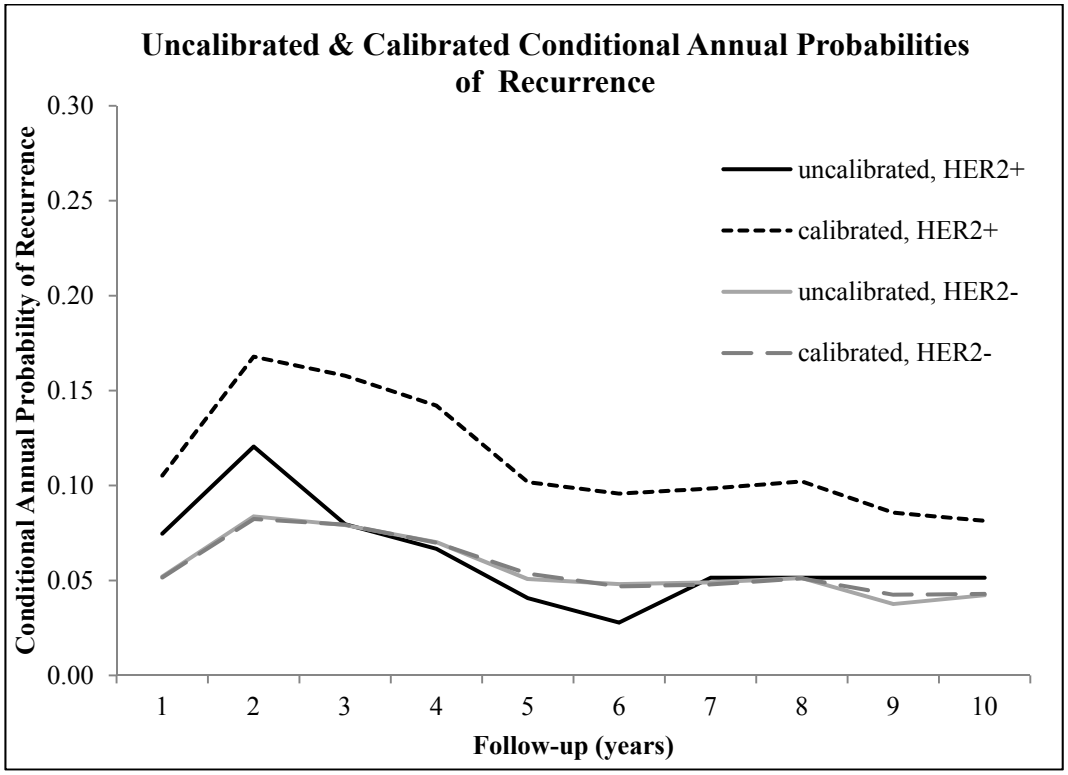
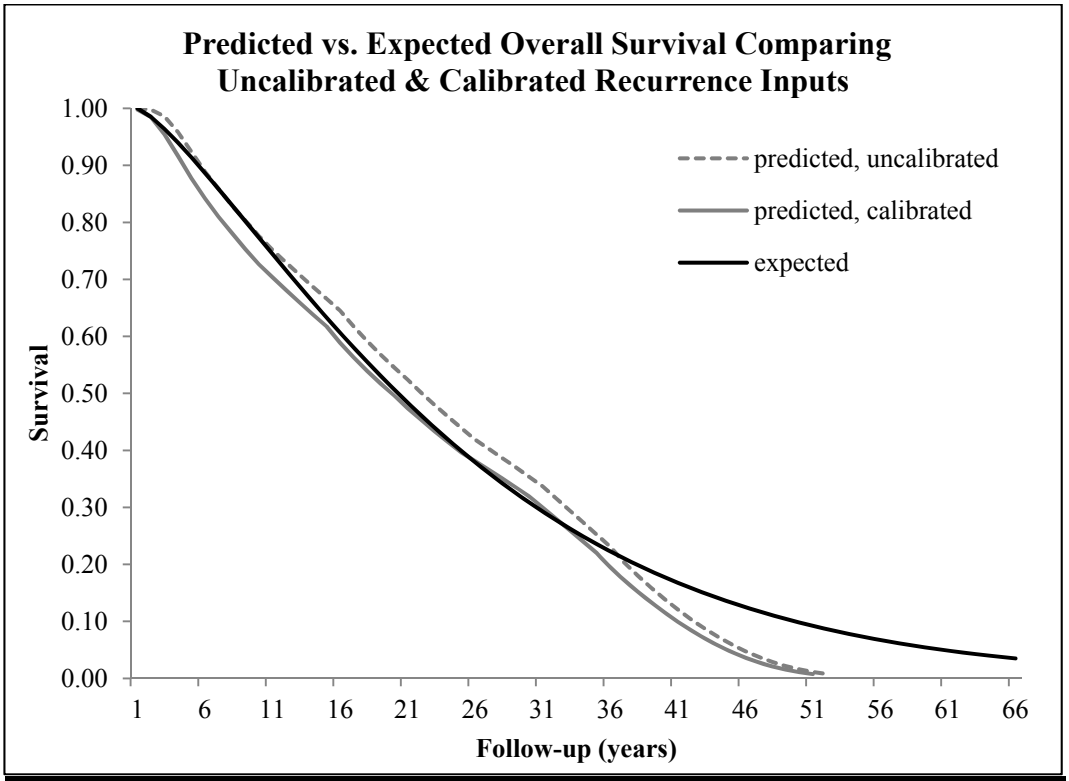


Figure 4.3. Conditional annual probabilities of recurrence (local or distant) from the NED state before and after calibration. Uncalibrated probabilities are shown as derived from the literature. Calibrated probabilities were identified through a random parameter search process wherein the probability was fit to a narrow beta distribution and random parameter sets were drawn from that distribution. The corresponding cumulative mortality associated with those parameters sets was compared to expected mortality by the least squares method of goodness of fit (GOF) estimation. The calibrated values shown above are those that produced the smallest GOF estimate.



Goodness of Fit Estimates	All Cycles	≤ Cycle 25	≤ Cycle 10
Predicted, uncalibrated	0.1442	0.0606	0.0112
Predicted, calibrated	0.0801	0.0047	0.0028

Figure 4.4. Expected survival vs. BC Markov model predicted survival in the lifetime perspective, before and after calibration. The uncalibrated predictions were based on disease recurrence parameters as derived from the literature. Optimal parameter values were identified through a random parameter search method and least squares method of goodness of fit estimation against expected survival.

4.8 APPENDICES

Appendix 4.1. Search strategy used to identify studies of HER2 test accuracy and concordance. The search was conducted in MEDLINE® In-Process and Other Non-Indexed Citations and MEDLINE 1946 to February 2, 2012.

#	Searches
1	("breast cancer" or "breast cancers").ti,ab.
2	("breast carcinoma" or "breast carcinomas").ti,ab.
3	1 or 2
4	*Breast Neoplasms/
5	3 or 4
6	*Receptor, erbB-2/
7	*Genes, erbB-2/
8	(HER2 or HER-2 or erbB2 or erbB-2 or C-erbB-2 or C-erbB2).ti.
9	6 or 7 or 8
10	5 and 9
11	Immunohistochemistry/
12	(immunohistochemistry or IHC or A0485 or CB11 or HercepTest).ti,ab.
13	11 or 12
14	in situ hybridization/ or in situ hybridization, fluorescence/
15	(in situ adj hybridi#ation?).ti,ab.
16	(FISH or CISH or SISH or PathVysion or pharmDx or UltraView or DuoCISH).ti,ab.
17	14 or 15 or 16
18	13 or 17
19	10 and 18
20	exp "Sensitivity and Specificity"/
21	False Positive Reactions/
22	False Negative Reactions/
23	"Reproducibility of Results"/
24	du.fs.
25	sensitivit*.tw.
26	(predictive adj4 value*).tw.
27	distinguish*.tw.
28	differentiat*.tw.
29	enhancement.tw.

#	Searches
30	identif*.tw.
31	detect*.tw.
32	diagnos*.tw.
33	accura*.tw.
34	comparison*.tw.
35	20 or 21 or 22 or 23
36	24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34
37	35 or 36
38	Comparative Study.pt.
39	(Validation Studies or Evaluation Studies).pt.
40	Randomized Controlled Trial.pt.
41	(Clinical Trial or Clinical Trial, Phase II or Clinical Trial, Phase III or Clinical Trial, Phase IV).pt.
42	Multicenter Study.pt.
43	38 or 39 or 40 or 41 or 42
44	(random* or sham or placebo*).ti,ab.
45	((singl* or doubl*) adj (blind* or dumm* or mask*)).ti,ab.
46	((tripl* or trebl*) adj (blind* or dumm* or mask*)).ti,ab.
47	(control* adj3 (study or studies or trial*)).ti,ab.
48	(non-random* or nonrandom* or quasi-random* or quasirandom*).ti,ab.
49	(allocated adj "to").ti,ab.
50	44 or 45 or 46 or 47 or 48 or 49
51	Cohort Studies/
52	Longitudinal Studies/
53	Prospective Studies/
54	Follow-Up Studies/
55	Retrospective Studies/
56	Case-Control Studies/
57	Cross-Sectional Study/
58	51 or 52 or 53 or 54 or 55 or 56 or 57
59	(observational adj3 (study or studies or design or analysis or analyses)).ti,ab.
60	cohort.ti,ab.
61	(prospective adj7 (study or studies or design or analysis or analyses or cohort)).ti,ab.

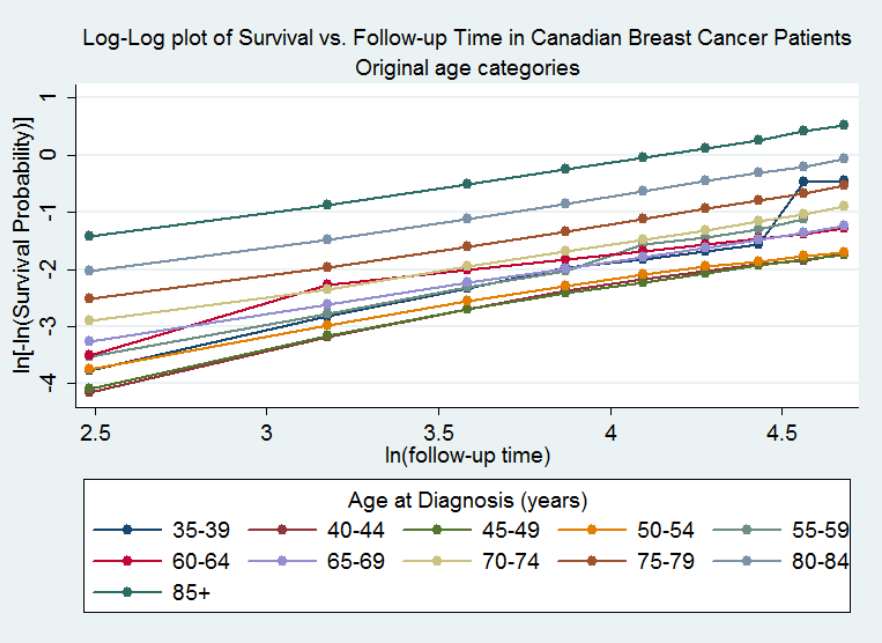
#	Searches
62	((follow up or followup) adj7 (study or studies or design or analysis or analyses)).ti,ab.
63	((longitudinal or longterm or (long adj term)) adj7 (study or studies or design or analysis or analyses or data or cohort)).ti,ab.
64	(retrospective adj7 (study or studies or design or analysis or analyses or cohort or data or review)).ti,ab.
65	((case adj control) or (case adj comparison) or (case adj controlled)).ti,ab.
66	(case-referent adj3 (study or studies or design or analysis or analyses)).ti,ab.
67	(population adj3 (study or studies or analysis or analyses)).ti,ab.
68	(cross adj sectional adj7 (study or studies or design or research or analysis or analyses or survey or findings)).ti,ab.
69	59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68
70	43 or 50 or 58 or 69
71	70 not Case Reports.pt.
72	37 or 71
73	19 and 72
74	limit 73 to animals
75	limit 74 to (animals and humans)
76	74 not 75
77	73 not 76

Appendix 4.2. Search strategy used to identify studies of HER2 test accuracy and concordance. The search was conducted in EMBASE® 1980 to week 5.

#	Searches
1	(HER2 or HER-2 or erbB2 or erbB-2 or C-erbB-2 or C-erbB2).ti.
2	*epidermal growth factor receptor 2/
3	*oncogene neu/
4	1 or 2 or 3
5	(breast adj3 (cancer$ or neoplasm$ or carcinoma$)).ti.
6	breast cancer/
7	breast carcinoma/
8	breast tumor/
9	5 or 6 or 7 or 8
10	4 and 9
11	(in situ adj hybridi#ation?).ti,ab.
12	(FISH or CISH or SISH or PathVysion or pharmDx or UltraView or DuoCISH).ti,ab.
13	fluorescence in situ hybridization/
14	in situ hybridization/
15	11 or 12 or 13 or 14
16	immunohistochemistry/
17	(immunohistochemistry or IHC or A0485 or CB11 or HercepTest or 4B5).ti,ab.
18	16 or 17
19	15 or 18
20	10 and 19
21	exp "Sensitivity and Specificity"/
22	false positive result/
23	false negative result/
24	reproducibility/
25	diagnostic accuracy/
26	du.fs.
27	sensitivit*.tw.
28	(predictive adj4 value*).tw.
29	distinguish*.tw.
30	differentiat*.tw.
31	enhancement.tw.

#	Searches
32	identif*.tw.
33	detect*.tw.
34	diagnos*.tw.
35	accura*.tw.
36	comparison*.tw.
37	specific*.tw.
38	di.fs.
39	21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38
40	cohort analysis/
41	longitudinal study/
42	prospective study/
43	follow up/
44	retrospective study/
45	case control study/
46	cross-sectional study/
47	40 or 41 or 42 or 43 or 44 or 45 or 46
48	(observational adj3 (study or studies or design or analysis or analyses)).ti,ab.
49	cohort.ti,ab.
50	(prospective adj7 (study or studies or design or analysis or analyses or cohort)).ti,ab.
51	((follow up or followup) adj7 (study or studies or design or analysis or analyses)).ti,ab.
52	((longitudinal or longterm or (long adj term)) adj7 (study or studies or design or analysis or analyses or data or cohort)).ti,ab.
53	(retrospective adj7 (study or studies or design or analysis or analyses or cohort or data or review)).ti,ab.
54	((case adj control) or (case adj comparison) or (case adj controlled)).ti,ab.
55	(case-referent adj3 (study or studies or design or analysis or analyses)).ti,ab.
56	(population adj3 (study or studies or analysis or analyses)).ti,ab.
57	(cross adj sectional adj7 (study or studies or design or research or analysis or analyses or survey or findings)).ti,ab.
58	48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57
59	(cross adj sectional adj7 (study or studies or design or research or analysis or analyses or survey or findings)).ti,ab.

#	Searches
60	validation study/
61	evaluation/
62	randomized controlled trial/
63	clinical trial/
64	phase 2 clinical trial/
65	phase 3 clinical trial/
66	phase 4 clinical trial/
67	multicenter study/
68	60 or 61 or 62 or 63 or 64 or 65 or 66 or 67
69	(random* or sham or placebo*).ti,ab.
70	((singl* or doubl*) adj (blind* or dumm* or mask*)).ti,ab.
71	((tripl* or trebl*) adj (blind* or dumm* or mask*)).ti,ab.
72	(control* adj3 (study or studies or trial*)).ti,ab.
73	(non-random* or nonrandom* or quasi-random* or quasirandom*).ti,ab.
74	(allocated adj "to").ti,ab.
75	69 or 70 or 71 or 72 or 73 or 74
76	47 or 58 or 68 or 75
77	case report/
78	76 not 77
79	39 or 78
80	20 and 79
81	limit 80 to human



Appendix 4.3. Scatterplot of the log (-log S(t)) Kaplan-Meier estimates against the log of time for observed mortality in Canadian breast cancer patients diagnosed between 1998 and 2006.

```
No. of subjects =       132388                     Number of obs   =    132388
No. of failures =        28717
Time at risk    =      7956480
                                                   LR chi2(1)      =   9220.84
Log likelihood  =   -84492.757                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        age3 |   .2176735   .0023332    93.29   0.000     .2131006    .2222465
       _cons |  -8.304679   .0343027  -242.10   0.000    -8.371911   -8.237446
-------------+----------------------------------------------------------------
       /ln_p |   .2608181   .0051039    51.10   0.000     .2508147    .2708215
-------------+----------------------------------------------------------------
           p |   1.297992   .0066248                      1.285072    1.311041
         1/p |   .7704211   .0039321                      .7627527    .7781665
------------------------------------------------------------------------------
```

The formula for survival in the Canadian breast cancer population diagnosed between 1998 and 2006 is:

$S(t) = e^{(-\lambda t^{p})}$

where,

λ = hazard parameter $= e^{(\beta_0 + \beta_1 age)}$

β_0 = intercept parameter = -8.304679

β_1 = *coefficient of age parameter = 0.2176735*

p = *shape parameter = 1.297992*

t = *time (in months)*

Appendix 4.4. Specifications of the final Weibull model fitted to the observed mortality in Canadian breast cancer patients diagnosed between 1998 and 2006.

Appendix 4.5. Conditional annual probabilities of recurrence (local or distant) from no evidence of disease (NED) state for HER2+ and HER2- patients age 50 at diagnosis. Uncalibrated estimates represent conditional annual probabilities as calculated from the literature. Only the calibrated estimates were used to estimate incremental cost-effectiveness ratios. Estimates shown for cycle 10 were used for all remaining cycles. All probabilities (μ) were fit to a beta distribution with the standard error (SE) shown.

	HER2+				HER2-			
	Uncalibrated [16–20]		Calibrated		Uncalibrated [21]		Calibrated	
cycle	μ	SE	μ	SE	μ	SE	μ	SE
1	0.0746	0.0113	0.1053	0.0113	0.0520	0.0011	0.0516	0.0011
2	0.1205	0.0149	0.1678	0.0149	0.0837	0.0014	0.0823	0.0014
3	0.0796	0.0144	0.1579	0.0144	0.0791	0.0014	0.0794	0.0014
4	0.0667	0.0150	0.1421	0.0150	0.0702	0.0013	0.0698	0.0013
5	0.0407	0.0227	0.1018	0.0227	0.0507	0.0016	0.0536	0.0016
6	0.0278	0.0070	0.0957	0.0070	0.0481	0.0015	0.0469	0.0015
7	0.0514	0.0155	0.0984	0.0155	0.0489	0.0016	0.0479	0.0016
8	0.0514	0.0155	0.1021	0.0155	0.0512	0.0016	0.0511	0.0016
9	0.0514	0.0155	0.0857	0.0155	0.0376	0.0014	0.0425	0.0014
10+	0.0514	0.0155	0.0814	0.0155	0.0422	0.0028	0.0429	0.0028

Appendix 4.6. Conditional annual probabilities of distant recurrence (DR) from the locoregional recurrence (LRR) state for HER2+ and HER2- patients. Uncalibrated estimates represent conditional annual probabilities as derived from the literature. Only the calibrated estimates were used to estimate incremental cost-effectiveness ratios. Estimates shown for cycle 10 were used for all remaining cycles. All probabilities (μ) were fit to a beta distribution with the standard error (SE) shown.

	All		HER2+		HER2-	
	Uncalibrated [22]		Calibrated		Calibrated	
cycle	μ	SE	μ	SE	μ	SE
1	0.2249	0.0250	0.2258	0.0250	0.2164	0.0250
2	0.2155	0.0282	0.2145	0.0282	0.2127	0.0282
3	0.1612	0.0284	0.1632	0.0284	0.1568	0.0284
4	0.1485	0.0298	0.1491	0.0298	0.1494	0.0298
5	0.0154	0.0113	0.0784	0.0113	0.0785	0.0113
6	0.0781	0.0304	0.0152	0.0304	0.0150	0.0304
7	0.0056	0.0117	0.0055	0.0117	0.0055	0.0117
8	0.0056	0.0117	0.0057	0.0117	0.0055	0.0117
9	0.0056	0.0117	0.0054	0.0117	0.0055	0.0117
10+	0.0056	0.0117	0.0058	0.0117	0.0058	0.0117

Appendix 4.7. Age-specific, conditional annual probabilities of death due to distant disease (stage IV, metastatic breast cancer) as derived from data provided by Cancer Care Ontario. [23] The same probabilities were used for HER2+ and HER2- patients, and were not calibrated.

	Age at Distant Recurrence							
cycle	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85+
1	0.2500	0.3352	0.2839	0.3309	0.4578	0.5068	0.5152	0.5581
2	0.3431	0.2105	0.2613	0.2258	0.3444	0.3425	0.3125	0.4167
3	0.1778	0.2778	0.1951	0.2113	0.1724	0.1667	0.2045	0.3333
4	0.1667	0.1111	0.1522	0.1163	0.1818	0.2143	0.1579	0.1250
5+	0.0833	0.1500	0.1250	0.2609	0.2667	0.0010	0.1667	0.5000

Appendix 4.8. Conditional annual probabilities of general mortality applied to patients in no evidence of disease (NED), early-discontinuation (ED), locoregional recurrence (LRR) and post-locoregional recurrence (PLRR) disease states. All estimates derived from Statistics Canada Life tables.[52]

age	μ	age	μ
50	0.0023	76	0.0274
51	0.0025	77	0.0307
52	0.0028	78	0.0342
53	0.0031	79	0.0381
54	0.0034	80	0.0424
55	0.0037	81	0.0475
56	0.0041	82	0.0535
57	0.0045	83	0.0607
58	0.0049	84	0.0687
59	0.0054	85	0.0776
60	0.0059	86	0.0870
61	0.0064	87	0.0970
62	0.0070	88	0.1077
63	0.0077	89	0.1190
64	0.0085	90	0.1309
65	0.0093	91	0.1432
66	0.0103	92	0.1559
67	0.0113	93	0.1709
68	0.0124	94	0.1868
69	0.0136	95	0.2038
70	0.0149	96	0.2218
71	0.0165	97	0.2408
72	0.0182	98	0.2609
73	0.0202	99	0.2821
74	0.0223	100	0.3043
75	0.0247		

Appendix 4.9. Search strategy for identifying systematic reviews of trastuzumab treatment effect in PubMed (up to August 2012).

Search	Query	Items found
1	Search (trastuzumab[Title/Abstract] AND "breast neoplasms"[MeSH Major Topic])	2402
2	Search ("review"[Publication Type] OR systematic[Title/Abstract] AND review[Title/Abstract] OR meta[Title/Abstract] AND analysis[Title/Abstract]))	85307
3	Search 1 AND 2	17
4	Keep:	7

Appendix 4.10. Description of studies excluded at data extraction stage of the targeted review of HER2 test agreement.

Study	Description	Reason for exclusion
Dressler 2005	US retrospective study of IHC and FISH concordance in tumour excision samples from node-positive patients registered to clinical trial CALGB 8541.	IHC reported as either negative or positive, unable to distinguish 0 from 1+, or 2+ or 3+.
Middleton 2009	US cohort study of IHC and FISH concordance in core biopsy samples.	Results reported graphically showing inconclusive cases over time.
Todorovic-Rakovic 2007	Serbian cohort study of IHC and FISH concordance.	Unable to distinguish IHC 0 from 1+, 2+ from 3+ in reported results.
Tse 2005	Multicentre cohort study of IHC and FISH agreement (also serum and PCR measurement) in tumour excision samples from metastatic patients enrolled in the HERMES trial.	Unable to distinguish IHC 0 from 1+, 2+ from 3+ in reported results.
Vogel 2010	German cohort study of IHC and FISH concordance in female core biopsy samples.	Data was reported as overall percentages of concordance without providing actual numbers; detailed concordance data reported for only 6 cases.

CHAPTER 5

CONCLUSION

Personalized medicine is viewed with great promise, but has not lived up to its full potential in current practice.[1] However, the identification of genetic disease markers and drug targets was curtailed by greater than expected complexity of biological and environmental interactions.[2] Targeted trastuzumab treatment of HER2+ BC stands out as an established example of personalised medicine used successfully in practice. We employed this example to explore: (1) the use of health technology assessment methods, including systematic review, meta-analysis, primary data collection and economic evaluation, to evaluate HER-targeted therapy as an example of personalised medicine [entire dissertation], (2) how companion diagnostics and targeted therapies are usde in practice (i.e. 'real world') compared to guidelines [Chapter 3], and, (3) how a decision-analytic model can be used to examine the 'comparative effectiveness' of current practice vs. guidelines in an economic evaluation of a personalised medicine [Chapter 4]. This concluding chapter discusses the implications of the main findings and their generalisability to other personalized medicines, the limitations of this work and areas for future research.

5.1 SUMMARY OF FINDINGS

We identified important gaps in published economic evaluations of trastuzumab targeted therapy through rigorous systematic review methods of evidence-based medicine research. The majority of published economic evaluations did not model the joint uncertainty of HER2 diagnosis and HER2-targeted thera-

py, and most analyses were conducted without taking into account the impacts of local disease epidemiology and clinical practice. A small selection of trastuzumab economic evaluations demonstrated that the ICER of treatment differs when HER2 testing is also modeled, and varies widely depending on the sequencing of alternative testing modalities.[3,4] These studies highlighted that initial FISH testing or FISH confirmation of IHC equivocal results were the strategies that resulted in the greatest health outcome gains at the lowest incremental costs by minimizing over and under treatment due to false positive and false negative diagnoses. However, these models considered test-treat alternatives against a strategy of no HER2 testing and no targeted therapy. We also noted an important gap regarding practice patterns in published CEAs of trastuzumab, all of which modeled scenarios of testing or treatment guideline adherence and strict clinical trial populations. We hypothesized that the aforementioned gaps in decision-analytic model framing would have an important impact on estimates of the incremental cost-effectiveness of targeted trastuzumab therapy. These hypotheses were based on small observational studies that demonstrated deviations from testing guidelines in practice, including a lack of access to HER2 testing, incomplete testing without confirmation of IHC equivocal results, or treatment in the absence of a HER2+ test result. [5] We hypothesized that these guideline deviations would have a negative impact on the ICER of trastuzumab by increasing rates of over and under treatment. The impact of guideline deviations, combined with the impact of test-

ing demonstrated in previous economic evaluations, have not been evaluated jointly.[3,4] These questions were the driving force behind our observational study designed to document HER2 testing and targeted treatment practice in Ontario. Our retrospective cohort study of patients newly diagnosed with early-stage BC in 2006-07 examined HER2 testing reported to a central pathology centre with linked data from several administrative health databases. This work documented several departures from guideline recommendations in routine practice documented in pathology reports. HER2 testing was not documented for all patients, and confirmatory FISH testing was not documented for all IHC equivocal results [6] (Chapter 3). We also documented cases of over- and under-treatment with trastuzumab that contradicted documented HER2 status. The clinical and economic impacts of these 'real-world' practice patterns were then estimated using a decision analytic model calibrated to represent the course of BC in Canada, including patterns of HER2+ disease prevalence documented for the Canadian population. We hypothesized that these deviations in practice would lead to fewer accurate diagnoses, and, when compounded by over-and under-treatment with trastuzumab, would affect the incremental cost-utility ratio of diagnosing and treating HER2+ BC.

Chapter 4 illustrates the potential impact of the documented deviations from HER2 targeted testing and treatment guidelines. This was the first economic evaluation to consider 'real-world' practice and disease prevalence factors beyond

recommended treatment practices and regional BC mortality. Thus, this work makes a unique contribution to the literature with its 'real-world' perspective by quantifying the consequences of population-level HER2 testing and targeted treatment practices. We used a decision analytic model to demonstrate that a policy of initial FISH testing would improve health outcomes and reduce costs in the long-term in a setting where FISH confirmation of IHC equivocal tumours is not performed consistently in practice. Moreover, we demonstrated the important influence of underlying HER2+ disease prevalence, and found that a more expensive but accurate diagnostic strategy (i.e. primary FISH) becomes more dominant as disease prevalence increases. This factor was particularly relevant given that documented HER2+ prevalence [6,7] may be lower than the generally accepted 20%.[8] These 'real-world' findings present novel additions to the body of economic literature around HER2-targeted therapy.

5.2 LIMITATIONS, GENERALIZABILITY AND IMPLICATIONS OF THIS WORK

Our pragmatic perspective, emphasizing real-world practices, was balanced against trade-offs in the generalizability of our findings. We documented testing and treatment practices specific to the Canadian province of Ontario. The generalizability of Ontario practices to other jurisdictions will depend on local testing and drug reimbursement policies, as well as healthcare system organization. Ontario's publicly-funded system reimburses for HER2 testing at a fixed rate, and

trastuzumab funding is provided for all patients with evidence of a HER2+ tumour. This practice is comparable to other Canadian provinces and some US health maintenance organisations (HMOs).[9] However, testing practices may be influenced by variations in reimbursement practice. For example, pathologists have reported a disincentive to FISH testing when laboratories are treated as cost-centres within a larger institution.[10] Indeed, pathologists in one study explained that this disincentive was driven by the disconnect between costs and revenues within medical institutions, where reimbursement for HER2 testing is provided to the institution and not passed onto the laboratory.[10] This was seen as a barrier to providing primary FISH testing for 67% of surveyed pathologists.

This reluctance to use primary FISH is similar to observed practice patterns, and suggests a wider generalizability of Ontario practice to some privately-funded healthcare systems. Moreover, this example highlights organizational complexities that were not considered in this analysis. These organizational complexities, including reimbursement for HER2 testing, laboratory budgeting processes, laboratory quality assurance programs, and the collection and storage of pathology information, all influence HER2 testing practice. These 'real world' elements are challenging to document at the population level, and thus represent another trade-off in this analysis. Indeed, the uptake of 'real world' or comparative effectiveness research is often inhibited due to its disconnect from system or institutional complexities.[10] To address this disconnect, researchers will need to engage in re-

search at a smaller level focusing on a single institution or health network. This trade-off in turn reduces the generalizability of findings, but may facilitate implementation of changes at the institution level.

Despite the challenges in conducting real world studies, we have documented important deviations from HER2 testing guidelines in practice. These deviations bear important consequences for patient care as they may lead to missed therapeutic opportunities or unnecessary exposure to ineffective treatment and harmful side effects. Clinical discussions of Canadian HER2+ BC management in the literature have focused on treatment without examining the continuity and accuracy of testing practice.[11] Similarly, discussions of HER2 testing in Canada have not addressed access to testing or the current status of quality assurance programs.[12] These discussion gaps exemplify the common disconnect between testing and treatment considerations when formulating targeted therapy policy.

These gaps may exist due to the challenge in studying the test-treat linkage in practice. Indeed, we faced several challenges in documenting HER2 testing practice. Better documentation of diagnostic testing practices, including the test type, result and sequencing of multiple results, would empower both researchers and institutions to study the test-treat linkage. Our study has demonstrated the tremendous effort and significant methodological challenges that are inherent in deriving diagnostic information from pathology reports. This was further complicated by the fact that centralized reporting of HER2 status was not mandatory at the

time of this study, despite Cancer Care Ontario's policy of restricting trastuzumab reimbursement in patients with evidence of a HER2+ test. This reflected a disconnect in HER2 reporting to the central Ontario Cancer Registry which did not permit us to comment on access to HER2 testing. We propose that improved documentation of diagnostic testing procedures, results and sequencing in databases will be essential as medical practice becomes more 'personalised' and directed by genomic markers of disease. Cancer registries are a natural starting point for documentation of genomic testing practice.[13] These centralized repositories would facilitate consistent data collection and reporting despite variations in pathology information management at the laboratory level. Without improved data collection, comparative effectiveness research on personalized medicines and quality studies will be limited in either scope or level of detail. This would in turn limit the ability of researchers to identify opportunities for and barriers to improving personalized medicine practice.

The authors of several economic evaluation guidelines have long encouraged a pragmatic perspective[14] before comparative effectiveness research became an international research priority. Those guidelines encouraged a pragmatic perspective through modelling of local disease epidemiology, local patterns of care, system-related influences and medication adherence patterns. However, we observed that contextual factors representing local practice were rarely included in published economic evaluations of targeted trastuzumab therapy. This study present-

ed a more pragmatic perspective by modelling the consequences of testing and treatment practices associated with alternative HER2 targeted test-treat strategies. Such deviations could have important consequences for personalized medicine policy, particularly in jurisdictions where policy-making is informed by economic evaluations. We have demonstrated that centrally documented HER2 testing practice does not adhere to guidelines despite a policy of universal test access and controlled drug reimbursement. We've also shown that current practice patterns produce a mean 11% increase in inaccurate diagnoses because FISH is not used to confirm all IHC equivocal results. Likewise, current treatment patterns resulted in undertreatment of 43% of HER2+ and overtreatment of 1% of HER2- patients, while 15% of IHC equivocal patients also received trastuzumab. Given the serious medical consequences of not treating aggressive HER2+ disease, and the high costs associated with inappropriately treating HER2- patients, our analysis provides quantitative evidence to support a primary FISH testing policy. This recommendation is novel, and could not be gleaned from previous analyses that considered guideline adherence against a no testing or treatment referent strategy. Our study demonstrates the additional importance of jointly modelling disease prevalence and its uncertainty, which is intertwined with the cost-effectiveness of testing. These considerations are important for future personalized medicines, particularly given recent moves towards using economic evaluation to inform 'value-based' drug price negotiations and policy development.[15,16] We recom-

mend that decision makers consider the following factors when forming personalized medicine policy using economic evaluations:

(1) The prevalence of the underlying genotype or condition and its uncertainty,

(2) The joint uncertainty of diagnostic test accuracy and treatment effectiveness in the target population,

(3) A range of diagnostic test sequences including single tests and confirmatory testing options, and

(4) The potential for varying degrees of adherence to testing guidelines and treatment uptake.

5.3 FUTURE RESEARCH RECOMMENDATIONS

While this research has made several contributions to the literature, it also raises important questions and avenues for research. Much of the focus of this chapter has been on the potential health outcome gains that could be achieved by using primary FISH testing to reduce inaccurate diagnoses. However, we also examined the potential to use similar CISH and SISH testing methods which use microscopes that are more widely available. We suggest that adequately powered, prospective diagnostic validation studies could evaluate equivalency with FISH. If proven equivalent, the easier conduct of either CISH or SISH could promote a shift towards primary ISH testing and therefore increase diagnostic accuracy in practice. This could have a relatively fast and measurable impact on care, as many

of the laboratory procedures are common among the ISH tests. Value of information analysis can shed light on the relative value of conducting research to reduce our uncertainty about testing technologies relative to other uncertain model parameters (e.g. HER2 prevalence, treatment effect) and identify the area offering greatest benefit relative to research costs.[17]

We've also alluded to some of the complex clinical and system-related factors that influence testing practice. A comparative cross-sectional study design could document a series of pathology laboratory funding schemes, pathology integration within a healthcare system, and related HER2 testing practices. This study would be an important first step to understanding what systems produce the highest quality of HER2 testing, and how to modify systems producing disincentives to accurate diagnosis. Decision-makers can then model existing and new pathology testing scenarios using discrete event simulation or system dynamic modelling methods. Although these models require more data and greater detail, they offer enhanced customization to the local setting, thereby increasing the pragmatic aspects of decision-analytic modelling.

From a clinical perspective, we documented a large gap in trastuzumab utilisation among eligible HER2+ patients. The reasons for this gap were not measured in this work, presenting an opportunity for future research to understand if patients were offered trastuzumab and reasons for declining offered treatment. This

research could inform modelling assumptions and may also reveal opportunities to improve the tolerability of treatment protocols.

5.4 REFERENCES

1. Issa AM. Personalized medicine and the practice of medicine in the 21st century. McGill J. Med. 2007;10(1):53–7.
2. Katsios C, Roukos DH. Individual genomes and personalized medicine: life diversity and complexity. Per. Med. 2010;7(4):347–50.
3. Lidgren M, Jönsson B, Rehnberg C, Willking N, Bergh J. Cost-effectiveness of HER2 testing and 1-year adjuvant trastuzumab therapy for early breast cancer. Ann. Oncol. 2008;19(3):487–95.
4. Elkin EB, Weinstein MC, Winer EP, Kuntz KM, Schnitt SJ, Weeks JC. HER-2 testing and trastuzumab therapy for metastatic breast cancer: a cost-effectiveness analysis. J. Clin. Oncol. 2004;22(5):854–63.
5. Wolff AC, Hammond MEH, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. J. Clin. Oncol. 2007;25(1):118–45.
6. Ferrusi IL, Earle CC, Trudeau M, Leighl NB, Pullenayegum E, Khong H, et al. Closing the personalized medicine information gap: HER2 test documentation practice. Am. J. Manag. Care. 2013;19(1):838–44.
7. Yang XR, Chang-Claude J, Goode EL, Couch FJ, Nevanlinna H, Milne RL, et al. Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. J. Natl. Cancer Inst. 2011;103(3):250–63.
8. Hicks DG, Kulkarni S. HER2+ breast cancer: review of biologic relevance and optimal use of diagnostic tools. Am. J. Clin. Pathol. 2008;129(2):263–73.
9. Haas JS, Phillips K a, Liang S-Y, Hassett MJ, Keohane C, Elkin EB, et al. Genomic testing and therapies for breast cancer in clinical practice. J. Oncol. Pract. 2011;7(3 Suppl):e1s–7s.
10. Trosman JR, Weldon CB, Schink JC, Gradishar WJ, Benson AB. What do providers, payers and patients need from comparative effectiveness research on diagnostics? The case of HER2/Neu testing in breast cancer. J. Comp. Eff. Res. 2013;2(4):461–77.
11. Verma S, Lavasani S, Mackey J, Pritchard K, Clemons M, Dent S, et al. Optimizing the management of HER2-positive early breast cancer: the clinical reality. Curr. Oncol. 2010;17(4):20–33.

12. Hanna W, Barnes P, Berendt R, Chang M, Magliocco A, Mulligan AM, et al. Testing for her2 in breast cancer: current pathology challenges faced in Canada. Curr. Oncol. 2012;19(6):315–23.
13. Phillips KA, Liang S-Y, Van Bebber S. Challenges to the translation of genomic information into clinical practice and health policy: Utilization, preferences and economic value. Curr. Opin. Mol. Ther. 2008;10(3):260–6.
14. Ferrusi IL, Leighl NB, Kulin N a, Marshall D a. Do economic evaluations of targeted therapy provide support for decision makers? J. Oncol. Pract. 2011;7(3 Suppl):36s–45s.
15. Henschke C, Sundmacher L, Busse R. Structural changes in the German pharmaceutical market: price setting mechanisms based on the early benefit evaluation. Heal. Policyolicy. 2013;109(3):263–9.
16. Sussex J, Towse A, Devlin N. Operationalizing value-based pricing of medicines : a taxonomy of approaches. Pharmacoeconomics. 2013;31(1):1–10.
17. Hoomans T, Fenwick EAL, Palmer S, Claxton K. Value of information and value of implementation: application of an analytic framework to inform resource allocation decisions in metastatic hormone-refractory prostate cancer. Value Heal. 12(2):315–24.