Abstract

Background: New genomic tests are being developed to predict an individual’s risk of cancer recurrence by analyzing the expression of multiple genes. However, it is unclear how to report the test results so that they would be most useful to clinicians. A mail-out questionnaire has the potential to help a) describe physicians’ attitudes towards the clinical use of new genomic tests, b) determine what information physicians prefer to have included in the test reports, and c) explore how physicians think the test results would impact their treatment recommendations.

Objectives: To design such a questionnaire that could be used in the eventual large-scale survey, and to ensure that the questionnaire a) is comprehensible, b) has face validity, c) appears interesting to, and d) does not place undue response burden on, the target population.

Methods: The first draft, based on a specific genomic test for breast cancer recurrence (Onco
type DX™) and on two case scenarios, was created. Cognitive interviews with practicing oncologists were conducted to identify problems in the questionnaire. The evaluation involved face-to-face interviews with Kingston oncologists who treat breast cancer, followed by telephone interviews with medical oncologists who treat breast cancer in other places in Ontario. Three-to-four oncologists were included in each round of interviewing after which the questionnaire was revised based on that round’s recommendations. Additional rounds of interviews were conducted until no new problems/issues were raised in one entire round.
Results: A medium-length questionnaire was drafted. Four rounds of interviews were conducted with no new problems/issues being raised in the fourth round. Most of the problems identified in the questionnaire related to comprehensibility, followed by logical issues which detected fundamental problems in the questionnaire design. There was no evidence of fatigue or disinterest in participants and they deemed the response burden reasonable.

Conclusion: The results suggest that the proposed questionnaire is comprehensible and has face validity. Additionally, it appears to be an interesting questionnaire to, and would not place undue burden on, the target population. Thus, the questionnaire is now ready for the field administration.
I am extremely grateful to my supervisors for their unfailing guidance and assistance in the completion of this thesis. Because of their support and talents, I am able to complete this project. I sincerely thank:

Dr. Julia Brettschneider, for her brilliant idea for this creative topic, and meticulous attention to detail. I am grateful that she brought me into the wonderful interdisciplinary world, which incorporates biostatistics, clinical epidemiology, and psychology. I am also thankful that she corrected mistakes that I didn’t notice, led me to think through issues around problems, and taught me how to have a rigorous attitude towards scientific research.

Dr. Michael Brundage, who provided the spark to ignite the project. His vision and energy shaped the direction and set the tone. His continuous encouragement and belief in my abilities were enormously instrumental in my progresses.

Dr. Deb Feldman-Stewart, who instructed me how to think critically, how to communicate effectively and efficiently, and how to write scientifically. Her thoughtful comments and suggestions helped me overcome challenges along with my studies, and effectively guided the evolution of my reluctantly tolerable text into a much improved and organized edition.

My supervisors have been the most supportive supervisors that a student could hope for. I can never thank them enough for the time and guidance that they have given me over the past years. Beyond the project, I have also learnt greatly because of these creative and dedicated mentors.
I would particularly like to thank Brenda Bass and Christine Tong for their support and help in “the pod” at the Cancer Care and Epidemiology at Queen’s Cancer Research Institute. I really appreciate their endless supply of knowledge and their kindness for revising my questionnaire and allowing me to practice interviews with them.

I would like to extend my deep gratitude to all the physicians who took the time to participate in this study. Their input was incredibly valuable to this study, and meant much more to me.

I would like to acknowledge funding from the CIHR Transdisciplinary Training Program in Cancer Research for this project. Without it this project could never have been initiated.

Last, but certainly not least, a depth of gratitude to all my friends and my family. My classmates, and my friends at Queen’s, thanks for being here—literally. You share my joy, you help me when I am in trouble, and you have seen each step that I have made. Thanks heartily to my family and my everlasting friends around the world. Thanks for being here—figuratively. You are the collective sunshine of my life, you help me on every level possible, and your constant support encourages me to continue doing what I love. I gratefully appreciate it.
Table of Contents

Abstract ............................................................................................................................................ ii
Acknowledgements ......................................................................................................................... iv
Table of Contents ............................................................................................................................ vi
List of Tables .................................................................................................................................. ix
Glossary .......................................................................................................................................... xi
Chapter 1 Introduction ..................................................................................................................... 1
  1.1 Research question .................................................................................................................. 1
  1.2 Motivation .............................................................................................................................. 1
Chapter 2 Literature Review ............................................................................................................ 5
  2.1 Introduction to new genomic tests ......................................................................................... 5
    2.1.1 The development of genetic research .............................................................................. 5
    2.1.2 New genomic tests for cancer prognosis ......................................................................... 6
  2.2 Implementation of new genomic tests ................................................................................. 10
    2.2.1 Implementation problems ............................................................................................. 10
    2.2.2 Health risk communication ............................................................................................ 11
    2.2.3 Test reports ............................................................................................................. ....... 12
    2.2.4 New genomic tests in clinical practice .......................................................................... 14
  2.3 Self-administered questionnaire ........................................................................................... 16
  2.4 Pretesting methods for questionnaire development ............................................................. 18
  2.5 Cognitive Interviewing ........................................................................................................ 21
    2.5.1 Thinking-aloud and Probing ......................................................................................... 21
    2.5.2 Application to self-administered paper questionnaires ................................................. 22
  2.6 Reliability and validity of the questionnaire ........................................................................ 24
  2.7 About the proposed questionnaire ....................................................................................... 25
    2.7.1 The selected test—Oncotype DX™ .............................................................................. 26
    2.7.2 Breast Cancer ................................................................................................................ 29
    2.7.3 The provision of scenarios ............................................................................................ 32
  2.8 Summary .............................................................................................................................. 33
Chapter 3 Study Design and Methods ............................................................................................ 35
  3.1 General Study Design .......................................................................................................... 35
  3.2 Study Goal ............................................................................................................................ 35
List of Tables

Table 1: New genomic tests that estimate cancer recurrence ......................................................... 8
Table 2: Definitions of the risk groups of Onco\textit{type} DX$^\text{TM}$ ............................................................... 39
Table 3: The issues covered and the corresponding scripted probes ............................................... 56
Table 4: The codes of problem types used in the study ..................................................................... 60
Table 5: A summary of survey questions .......................................................................................... 63
Table 6: Characteristics of the participants in this study ................................................................. 65
Table 7: Results of "recommendations for changes to wording"..................................................... 67
Table 8: Results of "need to clarify the intent" for Objective 1, Round 1 ........................................ 68
Table 9: Results of "problems related to ordering/formatting issues" .............................................. 68
Table 10: Results of "need to clarify the intent" for Objective 2, Round 1 ......................................... 69
Table 11: Results of "overall burden" for Objective 4, Round 1 .......................................................... 71
Table 12: Results of "limitations of the questionnaire" for Objective 4, Round 1 .......................... 71
Table 13: Results of "recommendations for changes to wording".................................................. 72
Table 14: Results of "need to clarify the intent" for Objective 1, Round 2 ......................................... 72
Table 15: Results of "problems related to ordering/formatting issues" .......................................... 73
Table 16: Results of "need to clarify the intent" for Objective 2, Round 2 ......................................... 74
Table 17: Results of "overall burden" for Objective 4, Round 2 ....................................................... 75
Table 18: Results of "limitations of the questionnaire" for Objective 4, Round 2 .......................... 75
Table 19: Results of "recommendations for changes to wording" .................................................. 76
Table 20: Results of "need to clarify the intent" for Objective 1, Round 3 ......................................... 76
Table 21: Results of "problems related to ordering/formatting issues" .......................................... 77
Table 22: Results of "need to clarify the intent" for Objective 2, Round 3 ......................................... 78
Table 23: Results of "overall burden" for Objective 4, Round 3 ....................................................... 78
Table 24: Results of "limitations of the questionnaire" for Objective 4, Round 3 .......................... 79
Table 25: Results of "need to clarify the intent" for Objective 1, Round 4 ......................................... 80
Table 26: Results of "need to clarify the intent" for Objective 2, Round 4 ......................................... 80
Table 27: Results of "overall burden" for Objective 4, Round 4 ....................................................... 81
Table 28: Results of "limitations of the questionnaire" for Objective 4, Round 4 .......................... 82
Table 29: The summary of the number of problems identified for each objective ......................... 82
Table 30: The summary of the problem codes assigned per interview round .............................. 83
Table 31: Potential outcomes of interest and the corresponding variables ....................................... 95
Table 32: Sample size determination

Table 33: Determination of widest possible confidence intervals
Glossary

Addition (carryover) effect

Respondents continue to think about the answers to the preceding question when they answer the next question. Especially when the second question is a general question relevant to the first one, the answers to the first one weigh much more than it should in answering the next question.

Anchoring or cognitive-based effect

It indicates the situations when a respondent’s answer to the preceding question guides the direction for answering the following question.

BRCA1/BRCA2

BRCA1 is a gene on chromosome 17 that normally helps to suppress cell growth. A person who inherits certain mutations (changes) in a BRCA1 gene has a higher risk of getting breast, ovarian, prostate, and other types of cancer. BRCA2 is a gene on chromosome 13 that normally helps to suppress cell growth. A person who inherits certain mutations (changes) in a BRCA2 gene has a higher risk of getting breast, ovarian, prostate, and other types of cancer.

Central tendency effect

Central tendency effect is the tendency to select the response option listed in the middle. It is common in questions asking for numbers.
DNA

It is a very large molecule composed of a string of smaller chemicals called nucleotides held together by a chain of 5-carbon sugars called ribose.

**Extreme tendency**

Extreme tendency is the tendency to pick the first or the last option. This effect is most often seen in questions asking for a name, an idea or a statement, especially when the list is long.

**Gene**

The functional and physical unit of heredity passed from parent to offspring. Genes are pieces of DNA, and most genes contain the information for making a specific protein.

**Gene expression pattern**

This is an equivalent term currently in use to refer to “gene expression profile.”

**Gene expression profile**

It is the expression of a set of genes in a biologic sample (eg, blood, tissue) using microarray, RT-PCR, or other technology capable of measuring gene expression.

**Gene expression profiling**

This term refers to any genomic techniques that quantify genes that are expressed in a specific sample (by assigning an expression value for each gene). This definition refers to techniques that allow the assessment of more than one gene at a time, especially microarray and real time RT-PCR.
Gene expression signature

This is an equivalent term currently in use to refer to a specific “gene expression profile,” usually associated with a specific phenotype (33).

Genetic predisposition

It is an inherited increase in the risk of developing a disease. Also called genetic susceptibility.

Increased positiveness of summary items when asked after specific items on the same subject

The tendency for respondents to think about the preceding concrete aspects when they answer the summary question on a subject.

Microarray

In the thesis, it refers to a DNA microarray, specifically. A DNA microarray, also commonly called “gene chip” or “DNA chip”, is a collection of microscopic DNA spots (defined “features”), commonly representing single genes or transcripts, arrayed on a solid surface by covalent attachment to chemically suitable matrices, or directly synthesized on them. DNA microarrays use DNA as part of their detection system. Qualitative or quantitative measurements with DNA microarrays use the selective nature of DNA-DNA or DNA-RNA hybridization under high-stringency conditions and fluorophore-based detection. DNA arrays are commonly used for gene expression profiling, i.e., monitoring expression levels of thousands of genes simultaneously, or for comparative genomic hybridization.
**mRNA**

Messenger RNA, the key intermediary in gene expression, and the template for protein synthesis. It translates the DNA's genetic code into the amino acids that make up proteins. In detail, each set of three bases, called codons, specifies a certain protein in the sequence of amino acids that comprise the protein. The sequence of a strand of mRNA is based on the sequence of a complementary strand of DNA.

**Norm of evenhandedness**

It happens on value-based questions. Respondents tend to be fair or evenhanded when they finish a previous value-based question, and respond to a subsequent similar question.

**Primacy effect**

Primacy effect is the tendency that the first response option has a higher likelihood of being selected over other options in the list. Primacy effect becomes common when the response option list is long.

**Protein**

A molecule made up of amino acids that are needed for the body to function properly. Proteins are the basis of body structures such as skin and hair and of substances such as enzymes, cytokines, and antibodies. Proteins are both the building blocks of living tissue and the builders that do the work.
Real Time Reverse Transcriptase Polymerase Chain Reaction (RT-PCR)

Real-time RT-PCR is a molecular biology technique that allows the amplification and the quantification in real time of defined RNA molecules from specific specimens. This technology has been used for decades in research and clinical settings to measure RNA molecules. In the first step DNA, copies of the investigated RNA molecules present in the template are obtained by a reaction named reverse transcription. Then DNA amplification is obtained using PCR, while the quantification of the accumulating DNA product is accomplished by the use of specific fluorescent reagents. The quantification of the target RNA molecule is based on the analysis of the accumulation curve of the complementary DNA, as measured by the fluorescence detected at each cycle of the reaction.

Recency effect

Recency effect arises when an option at the end of the list is more likely to be chosen.

RNA

RNA is similar to DNA in that it comprises a string of nucleotides held together by a chain of ribose molecules. Each nucleotide in RNA corresponds to a similar nucleotide in DNA. RNA communicates the genetic information from DNA to the proteins that are produced in the cell.

Subtraction effect

A seemingly opposite effect of the “addition effect”. When the part of the next question has already been answered in the previous question, the respondent tends to “subtract” out reasons
used for the first question, and therefore base his/her answer to the second question.
Chapter 1

Introduction

1.1 Research question

The goal of this study was to design a questionnaire that could be used in a later study to systematically a) describe physicians’ attitudes towards the clinical use of new genomic tests, b) determine what information physicians prefer to have included in the reports of the new tests, and c) explore how physicians think the results of the new genomic tests would impact their treatment recommendations.

1.2 Motivation

In recent years, a new type of genomic test has emerged. These new tests are designed to estimate the individual risk of cancer recurrence by analyzing the expression of multiple genes expressed in tumor samples (1-5). Although the tests have the potential to refine cancer prognosis, the future role of new genomic tests can be substantially influenced by the conditions in which the tests are implemented. The content and the layout of the test report, as an important part of the implementation process, may affect whether the tests are likely to be used correctly, widely and efficiently in the future, and whether they can assist physician-patient communication regarding medical decision-making (6). Currently, there are no guidelines for reporting new genomic tests.
Oncologists, the potential users of these tests, may shed some light on how to optimize test reporting. However, communication of new genomic test results has only been described from patients’ perspectives (7) and it is unclear what information physicians desire to have from these new tests. In physicians’ clinical practice, new genomic tests, as a novel type of prognostic factor, have already demonstrated an impact on physicians’ treatment recommendations to some extent (8-11). Still, little is known about physicians’ attitudes towards the tests or how to make their reports most useful to the clinicians.

Driven by these research needs, a self-administered questionnaire was thought to be an effective way of determining physicians’ attitudes and needs regarding these new genomic tests. However, such research required the creation of a questionnaire, as none had been used previously. The motivation for this thesis, therefore, was the creation of such a questionnaire. The main goal of this thesis was determined to design a questionnaire with face validity as judged by content experts.

The creation of a new questionnaire designed to explore new contents required a large amount of effort. The research process is described in detail in Chapters 2 and 3 to follow. Briefly, it required selecting a typical test, introducing the test to oncologists, designing new questions, and evaluating the questionnaire. However, easier said than done. First, a representative genomic test had to be selected in order to place the survey questions in an appropriate context, and then the corresponding disease site and the target population could be determined. Second, the selected test had to be introduced to physicians because it is possible that
the test would not be known to many oncologists. It was necessary to determine what information should be included as the background for oncologists to understand the test, and to decide how to present the information. Beyond the provision of background information, the potential role of the test in clinical decision making also needed to be described and explored. To do so, clinical case scenarios were designed and included in the questionnaire. Concrete examples were used to facilitate physicians’ understanding of the test. The scenarios also provided conditions under which physicians’ attitudes could be judged, and the clinical application of the test could be illustrated. In order to design reasonable case scenarios that could guide physicians’ thinking about making clinical decisions, literatures on all prognostic factors involved in physicians’ decision-making process, and treatment recommendation guidelines were reviewed. Third, it was necessary to create a large number of novel questions, since few questions could be extracted from the literature, and since no previous hypotheses were available to guide what variables should be sampled by the survey or what types of relationships might exist among the variables. Therefore, the exploration of physicians’ information needs in the test report, including for example, the content, the expression, and the presentation of the test results, used principles from the literature on health risk communication. Finally, the logistics of seeking the opinions of local content experts, and the role of cognitive interviewing techniques in meeting the thesis objectives required the development and the implementation of cognitive interviews at the local cancer clinic level.
In conclusion, this thesis intended to develop a self-administered questionnaire, in an attempt to inform optimal clinical implementation of new genomic tests which have great potential to refine cancer prognosis.
Chapter 2

Literature Review

2.1 Introduction to new genomic tests

2.1.1 The development of genetic research

Many achievements in human genetic research can be traced back to the Human Genome Project (1990-2003). The project aimed at identifying all genes (approximately 20,000-25,000) in human DNA, and determining the sequences of all chemical base pairs (about 3 billion) that make up human DNA (12). This program provided a means for searching for a molecular explanation of health and disease (12).

With the help of genes identified through this project, it is possible to research structures and functions of genes, gene-gene interactions, and gene-environment interactions. This newly acquired knowledge, combined with modern molecular technology, is changing the world of medicine. For example, many complex diseases, such as certain types of cancers (12), brain disorders (13), have been studied at the molecular level. Studies of molecularly targeted therapies have also been launched. Commercial tests for cancer-screening and prognosis are also becoming available.
2.1.2 New genomic tests for cancer prognosis

A genomic test is a test which looks at many or all disease-related genes of an individual, and tries to understand how active these genes are, as well as how they interact and function. These tests may help prevent, diagnose, and treat the disease (14). It is worthwhile to mention that a genomic test is not a genetic test. The difference between genomic tests and genetic tests comes from the difference between genomics and genetics. Usually, genomics involves studying genes, their activities and functions, interactions and related techniques, while genetics focuses on heredity (15), such as single genes and their effects (16). For example, in breast cancer, a genomic test might provide a snapshot of which activities of a group of 70 genes are present. A genetic test, however, would test whether there is a mutation in the breast cancer predisposition gene BRCA1/BRCA2 (see glossary). Sometimes, genetic tests are deemed to be a subset of genomic tests (17).

In cancer research, newly developed genomic tests are based on an emerging research method, called “gene expression profiling” (see glossary), which measures how active particular genes are. In these genomic tests, a list of genes that are correlated with different cancer diagnoses or prognoses is identified; the expression of the genes is analyzed: how many times the genes are transcribed into cellular messenger ribonucleic acid (mRNA) transcripts, or how many times the genes are made into functional products (e.g., proteins) in the end. Mostly, new genomic tests were developed to target mRNA transcripts rather than genes themselves (DNAs) or the genes’ products (e.g., proteins). One reason is that mRNA is considered more accurately in the analysis of tumor behaviour compared to DNA because genes in cancer cells can be altered in
terms of their expression level (RNA level) without mutations in their structures (DNA level) (18). These non-mutated genes with constantly altered expression are actually a significant component of carcinogenesis. Another reason is that the technology which deals with RNAs is more advanced than that dealing with proteins. While genetic changes can be identified by screening at the protein level, recovering and utilizing cellular proteins in experiments is still an obstacle to be overcome (18).

There are two common molecular technologies currently used to portray a gene expression profile, Real-Time Reverse Transcriptase Polymerase Chain Reaction (Real-Time RT-PCR) and Microarray (see glossary). Microarrays offer the advantage of profiling expression levels of tens of thousands of genes simultaneously; although RT-PCR is only capable of analyzing hundreds of genes at the same time, it provides highly precise quantification of gene expression levels (19). RT-PCR serves as the “gold standard” against which other methods are compared (19, 20), especially in the validation of microarray results.

The US Food and Drug Administration (FDA) has drafted a guideline for defining and regulating new genomic tests under the class of In Vitro Diagnostic Multivariate Index Assays (IVDMIAs) (21). The American Society of Clinical Oncology (ASCO) and the National Comprehensive Cancer Network (NCCN) have recommended the use of particular new genomic tests in clinical practice (22, 23). Thus, new genomic tests are quickly moving into the realm of routine clinical care.
The general uses of new genomic tests can be divided into three categories:
a) distinguishing molecular subtypes of cancer (2, 24-28), b) providing cancer prognosis (1-5),
and c) aiding the selection of treatment for individual patients (29, 30). This thesis is focused, in
particular, on genomic tests that estimate cancer recurrence (cancer prognosis).

Genomic tests estimating cancer recurrence supplement traditional prognostic factors
with information about the expression of cancer-related genes. Some researchers have deemed
that new genomic tests are potentially more powerful prognostic and predictive tools than the
ones currently used (31). By refining cancer prognosis, these tests may have the potential to
individualize treatment recommendations. That is, an individual would receive the treatment from
which he/she can gain a substantial benefit and avoid being over-treated or under-treated. In other
words, the right patient can receive the right treatment. Below is a list of new genomic tests that
are commercially available (CA) or are under development (UD):

Table 1: New genomic tests that estimate cancer recurrence

<table>
<thead>
<tr>
<th>Test name</th>
<th>Disease Site</th>
<th>Measurement</th>
<th>What to measure</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oncoty DX (32-34)</td>
<td>Breast Cancer</td>
<td>RT-PCR</td>
<td>21 genes</td>
<td>Commercially Available (CA) since 2004</td>
</tr>
<tr>
<td>MammaPrint (32-34)</td>
<td>Breast Cancer</td>
<td>Microarray</td>
<td>70 genes</td>
<td>CA, approved by FDA in 2007</td>
</tr>
<tr>
<td>H/I ratio (BCP assay) (32, 33)</td>
<td>Breast Cancer</td>
<td>RT-PCR</td>
<td>Breast Cancer Gene Expression Ratio of HOXB13 and IL17BR transcripts</td>
<td>CA</td>
</tr>
<tr>
<td>Mammostrat (35)</td>
<td>Breast Cancer</td>
<td>Immunohistochemistry (IHC)</td>
<td>5 biomarkers: p53, HTF9C, CEAMC5, NDRG1, SLC7A5</td>
<td>CA</td>
</tr>
<tr>
<td>Src homology collagen-like</td>
<td>Breast Cancer</td>
<td>Shc protein-based tests</td>
<td>Level of activated tyrosine</td>
<td>CA</td>
</tr>
<tr>
<td>Test Name</td>
<td>Disease</td>
<td>Test Type/Technique</td>
<td>Biomarkers/Details</td>
<td>Approval Status</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Protein test</td>
<td>Prostate Cancer</td>
<td>Immunohistochemistry (IHC) on tissue microarray</td>
<td>12 biomarkers: 6 markers for target antigens based on their cell distribution, and 6 for cancer associated antigens</td>
<td>CA, approved by FDA in 2007</td>
</tr>
<tr>
<td>Guanylyl cyclase C (GCC) test</td>
<td>Colorectal Cancer</td>
<td>Ultrasensitive messenger RNA-based amplification technology</td>
<td>Shc protein and p66 Shc protein in tissue specimens</td>
<td>CA</td>
</tr>
<tr>
<td>Pulmostrat</td>
<td>Lung Cancer</td>
<td>Immunohistochemistry (IHC)</td>
<td>Four antibodies</td>
<td>CA</td>
</tr>
<tr>
<td>eXagenBC</td>
<td>Breast Cancer</td>
<td>Computational approach and FISH platform</td>
<td>specific DNA sequences on chromosomes</td>
<td>Under Development (UD)</td>
</tr>
<tr>
<td>TBA (14-gene prognostic score)</td>
<td>Breast Cancer</td>
<td>TBA</td>
<td>N/A</td>
<td>UD</td>
</tr>
<tr>
<td>TBA (ER/PR receptor status)</td>
<td>Breast Cancer</td>
<td>TBA</td>
<td>N/A</td>
<td>UD</td>
</tr>
<tr>
<td>Metagene prognostic signature</td>
<td>Breast Cancer</td>
<td>Mathematical algorithm</td>
<td>Discriminatory information in a given cluster of genes</td>
<td>UD</td>
</tr>
<tr>
<td>Rotterdam signature (76-gene signature)</td>
<td>Breast Cancer</td>
<td>Microarray</td>
<td>18400 RNA transcripts (14500 well characterized human genes)</td>
<td>UD</td>
</tr>
<tr>
<td>Wound-response gene</td>
<td>Breast Cancer</td>
<td>Microarray</td>
<td>512 genes were included in the signature</td>
<td>UD</td>
</tr>
<tr>
<td>Genomic grade</td>
<td>Breast Cancer</td>
<td>Microarray</td>
<td>97 genes were included in the signature</td>
<td>UD</td>
</tr>
<tr>
<td>p53 signature</td>
<td>Breast Cancer</td>
<td>Microarray</td>
<td>Functional status of p53 (32 genes were included in the signature)</td>
<td>UD</td>
</tr>
<tr>
<td>Death-from-cancer signature</td>
<td>Breast Cancer</td>
<td>Microarray</td>
<td>BMI1 oncogenic pathway self renewal (11 genes were included in the signature)</td>
<td>UD</td>
</tr>
<tr>
<td>Invasiveness gene signature</td>
<td>Breast Cancer</td>
<td>Microarray</td>
<td>Tumorigenic cancer cells CD44+</td>
<td>UD</td>
</tr>
</tbody>
</table>
### CD24<sup>−/low</sup>
(186 genes were included in the signature)

<table>
<thead>
<tr>
<th>Colorectal cancer</th>
<th>Microarray</th>
<th>N/A</th>
<th>UD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ColoPrint (2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lung cancer (30, 38)</td>
<td>Microarray</td>
<td>N/A</td>
<td>UD</td>
</tr>
<tr>
<td>Duke’s B Colon Cancer (39)</td>
<td>Microarray</td>
<td>22000 RNA transcript</td>
<td>UD</td>
</tr>
<tr>
<td>Predictive system (12 genes) (40)</td>
<td>Hepatocellular carcinoma</td>
<td>Microarray</td>
<td>N/A</td>
</tr>
</tbody>
</table>

### 2.2 Implementation of new genomic tests

#### 2.2.1 Implementation problems

The implementation of any new laboratory-based test frequently experience the following problems:

- **a)** errors and difficulties in quality control in tissue sampling, laboratory processing and statistical analysis (41),
- **b)** logistic challenges,
- **c)** financial constraints, and
- **d)** insufficient evidence to support its routine use.

Implementation of new genomic tests faces those problems:

- **a)** Error or bias may be introduced by the noisiness of the data obtained from still immature large-scale measurement technologies such as some of the microarray platforms (41).
- **b)** The logistic challenges may be caused by the fact that there are few laboratories certificated to analyze the tumor samples.
c) Financial constraints are imposed by the high costs of new genomic tests which ironically result from complicated technical steps rather than the clinical value of these tests.

d) There are few outcome studies to address whether the introduction of these tests in clinical practice can actually lead to better health care (42).

Beyond the properties and requirements of the tests themselves, how the test reports are provided to physicians and patients can affect whether or not these tests will be used in the future (6). This is of particular concern for genomic tests estimating cancer recurrence risk, where physicians only receive the risk category results from some tests but do not know what the results represent or how to explain them.

2.2.2 Health risk communication

For genomic tests to be of any use, their results will need to be communicated. Communicating information about prognosis can be informed by the field of health risk communication, the study of exchanging information about types and levels of risk and about methods to manage risks among individuals, groups, and institutions (43).

In cancer prognosis, the core of health risk information is the probability of cancer recurrence for a patient. This information can be expressed as a percentage, proportion or frequency, can be categorized into a risk group, or can be transferred into odds; however, the accurate and effective communication of this important information may vary based on:

a) methods of risk communication (calculated risk estimates, role model description/vignettes,
etc), and b) ways of framing information (expression of probabilities, how the information is presented, etc) (44).

Over a number of years, prominent in risk communication research is “knowledge, attitudes, and behaviour” (44). The health risk communication research on new genomic tests, however, has only been described from patients’ perspectives to date (7). How to communicate the health risk information derived from these new genomic tests should also be studied from the perspectives of physicians, in that physicians are the ones who have to understand these tests, make clinical decisions, introduce the tests and interpret the test results to their patients. It has been noted that the information in traditional genetic tests is not always helpful for physicians’ decision making (45) or for physician-patient communication (45, 46).

2.2.3 Test reports

Test laboratories usually communicate with physicians through test reports. Currently, no guidelines for genomic test reporting are available but there are guidelines or recommendations for genetic test reporting. Clinical Laboratory Improvement Amendments (CLIA) guidelines, for example, suggest that reports should include the test result, an interpretation of the result, comments on the result, recommendations for further testing or clinical consultation, and a summary of the test method and its limitations, if applicable (47). The FDA has also drafted a template for general genetic test reporting (48). Test reports of current new genomic tests, however, are disparate. One reason could be that little is known about what information about
new genomic test results physicians would prefer to receive. Presumably, the test report should be written in language that physicians could understand accurately and efficiently during routine clinical practice; otherwise, the test results might be misinterpreted (49). In addition, the complexity of the test report should match the information need of a typical physician. Previous evidence suggested that geneticists may underestimate the amount of information that a typical physician would like to receive (50).

An important tool in risk communication through test reports is the use of graphical displays of quantitative information. It has been hypothesized that presenting data in a graphical display can reduce the cognitive effort required to understand quantitative information (51). In addition, a graph allows the reader to process quantitative information in a format that is easier to retain than a textual presentation (52). Evidence also suggests that the best visual format varies according to the type and the complexity of the task for which the information is used (53, 54).

It is suggested that an excellent visual display of quantitative information “consists of complex ideas communicated with clarity, precision, and efficiency” (55). That is, a good design of visual formats should have two key elements: simple in design, and complex in data presentation (55). Visually attractive and friendly graphics can provoke curiosity, and also gather power from content and interpretations beyond the immediate display of some numbers (55).

Examples of visual formats include line graphs, bar graphs, pie charts, systematic ovals, and more complex formats. To display a trend over time, a line graph is superior to a bar chart in terms of distinguishing differences between points in time; while a bar chart is better in
illustrating the differences in discrete quantities (53). For summarizing data, line graphs appear to help speed decision-making process, while bar charts are more understandable and readable (53).

There are some recommendations available on how best to display quantitative information graphically although these suggestions tend not to be empirically based. For example, it is claimed that tables can be the best choice to show exact numerical values, and tables are preferable to graphics for many small data sets (55). Words should run horizontally, particularly along the Y-axis. Words run in several different directions should be avoided (55).

The graph presentation of cancer recurrence information in new genomic tests has not been investigated yet. Studies to determine which visual formats yield best performance (e.g., most accurately and easily understood, etc.) are needed.

2.2.4 New genomic tests in clinical practice

Clinical decision-making is a matter of fact uncertainty (56). When making a decision about using a new test or not, a physician typically estimates how much the potential new information would change the uncertainty about a patient’s true state (57). Therefore, the selection of a new test depends on the level of uncertainty that a physician has about a patient’s true state beforehand and at which level a physician is willing to start a certain treatment (57). Theoretically, if a physician was virtually certain of his/her estimate, providing treatment recommendation without further testing would be preferred. When a patient’s risk of cancer recurrence is intermediate, new tests are more likely to be performed.
To use a new test or not also depends on physicians’ attitudes towards risk and uncertainty about using new technologies in clinical practice. Physicians who are comfortable dealing with uncertainty will be more willing to use new tests, or embrace new technology. However, physicians who try to avoid such a risk will feel more comfortable with established procedures or tests (58).

In addition, many other factors may also affect whether a test will be utilized or not (57). These factors include the cost of the test, the patient’s need for reassurance, unpleasant effects of the test, etc. Currently, little is known about clinical conditions in which physicians are likely to use new genomic tests, and about what concerns would be if using a new genomic test is not preferred. Thus, studies on physicians’ concerns and expectations with regard to these new genomic tests are needed.

Another aspect of the clinical application that needs investigation is to determine how a result from a new genomic test would affect a treatment recommendation. Some preliminary studies do suggest that new genomic tests could impact physicians’ treatment recommendations (8, 9, 11), increase physicians’ confidence in the treatment recommendations (8), and change actual treatments administered (8, 9, 11). Regarding the direction of treatment changes between hormonal therapy and chemotherapy, the results were controversial. Two studies found that the most frequent change in treatment decision were made from chemotherapy to hormonal therapy (8, 11), while the third study claimed that the change made from hormonal therapy to chemotherapy occurred mostly (9). More evidence is needed to address situations under which
treatment recommendations are likely to be changed, to address directions of changes, and to address how confident physicians are about changes.

### 2.3 Self-administered questionnaire

Often, a survey is an effective instrument to collect data where insufficient information exists (59). To determine physicians’ opinions towards the clinical implementation of new genomic tests, a survey would be a direct and efficient instrument.

Among types of questionnaires, a self-administered questionnaire offers a number of advantages:

a) allowing respondents to answer at their leisure,

b) presenting various visual formats,

c) minimizing social desirability bias,

d) eliminating interviewer distortion,

e) producing the same response rates as personal or telephone interviews among physicians(60, 61), and

f) requiring relatively limited resources.

However, compared with other survey methods (phone interviews, etc), a self-administered questionnaire also has some disadvantages. It:

a) tends to have more concerns about the unknown bias from non-respondents,

b) is less successful in obtaining answers to open-ended questions,

c) has less control over the sequence in which questions are completed,

d) may have higher item non-response rate,
e) lacks an opportunity for immediate problem-solving by an interviewer when respondents complete the questionnaire, and

f) requires a potentially longer time for the survey implementation (61).

For research of physicians’ attitudes and preferences for genomic test reports, a self-administered questionnaire should be a good strategy, because:

a) Members of our target population—oncology physicians—have demanding work schedules (62); their accessibility is a great concern. In order to obtain a relatively high response rate, providing a flexible time for oncologists to answer the questionnaire, and no restriction in locations where the questionnaire can be completed are very important.

b) As part of learning what information physicians would like included in the test reports, a self-administered questionnaire can present various aspects of information that could be included without requiring additional equipments or resources.

c) For questions regarding physicians’ attitudes and concerns, a self-administered questionnaire can eliminate bias caused by using different interviewers.

d) A self-administered questionnaire allows for the collection of data from a large number of physicians relatively inexpensively.

Although a self-administered questionnaire would be a preferred strategy to learn about physicians’ attitudes and information needs in this context, literature reviews related to new genomic tests, and to physicians’ attitudes, concerns or information needs towards other diagnostic or prognostic tests, revealed few helpful questionnaires, except one that addresses...
patients’ perceptions (7). As a result, selection of content to be covered was driven by the characteristics of new genomic tests, and layout design had to rely on general theories underlying survey development.

2.4 Pretesting methods for questionnaire development

Such novel questionnaires require thorough pretesting as part of their development. Potential pretesting methods include cognitive interviewing, formal expert review, focus groups, behavior coding, and re-interview surveys. The features of these interviewing methods are described below:

a) Cognitive interviewing: it aims to investigate specific survey questions/wording on topics or major themes of interest. It is characterized by testing potential problems iteratively. Rounds of a fixed number of interviews are conducted until no new problems are identified. Cognitive interviewing is effective for studying the individual in isolation, for identifying covert problems (e.g., silent misinterpretation), and for identifying reasons for problems. However, the biggest concern with cognitive interviewing is that interviewing conditions are different from field environments. As a result, questions arise in a cognitive interview may not happen in the field. However, preliminary evidence suggests that the differences are not likely to be an issue. There are some studies that support the contention that types of problems found in cognitive interviews appear to be consistent with findings from a field situation (63).
b) Formal expert review: it “involves the systematic collection and processing of reviews from sources outside the immediate design team” (63). Expert review may include questionnaire design experts, subject matter experts, questionnaire administration experts (e.g., experienced interviewer), and computer-based expert systems (e.g., QUEST, a computer model which can identify difficulty in question clarity based on a linguistic model of text comprehension.) (63).

Cognitive interviewing and expert review are not mutually exclusive. A form of expert review—expert appraisal is involved in the first step of cognitive interviewing (preparation stage). The formal expert review is not as practical as other pretest methods because well-documented procedures for formal expert review are not available (63).

c) Focus groups: it is a qualitative approach, and often used to investigate topics/major themes by asking an interactive group of topic-relevant people in a room. Specifically, focus groups are suited to investigate what people should be asked and what they think about key concepts (63). Questions can be asked in focus groups to obtain a full, in-depth discussion of a topic. However, it may not be effective in identifying problems by reviewing questions one by one. Participants of focus groups can affect each other; as a result, focus groups cannot study the individual in isolation. The requirement of a fixed time for focus groups can be a challenge for individual participants, especially when their schedules are busy.
d) Behavior coding: it is a systematic means for cording and quantifying errors or
difficulties during interviews. Behavior coding is mainly concerned with counting
problems rather than identifying reasons for those problems. Compared with other
pretesting methods, this one needs a relatively large sample size (at least 50 interviews)
to produce a valid code distribution. In addition, behavior coding is more oriented
towards focusing on overt problems, detecting problems of field interviewers, and it
focuses strictly at the question level. Consequently, it is weak in detecting covert
problems, and it neglects the interaction between questions. Behavior coding is usually
conducted at a later stage of questionnaire development, when the questionnaire is in a
form appropriate for field administration rather than early drafts.

e) Reinterview: it is a pretest method that involves asking respondents the same question
twice, in order to determine the consistency of respondents’ answers. It is an occasionally
used quantitative means for measuring the reliability of responses. By asking questions at
multiple time points, this method can shed some light on the consistency of responses to
individual items (63). However, it can neither be used to identify problems in the
questionnaire nor to clarify what underlies the problem so it can be corrected in a
meaningful manner.

Based on objectives and application environment of different pretesting methods,
cognitive interviewing is the most appropriate choice for testing the drafted questionnaire. Other
pretesting methods may also be applied at a different stage of the questionnaire development.
2.5 Cognitive Interviewing

Cognitive interviewing, also called “intensive interviewing” or “expansive interviewing”, is defined as “the administration of draft survey questions while collecting additional verbal information about the survey responses, which is used to evaluate the quality of the response or to help determine whether the question is generating the information that its author intends.” (64). Cognitive interviewing is derived from Cognitive Aspects of Survey Methodology (CASM), which is the interdisciplinary combination of survey methodology and cognitive psychology (63). Cognitive interviewing, with the purpose of evaluating survey questions, is applied CASM research. In the practice of questionnaire development, cognitive interviewing serves the function of “product testing”.

2.5.1 Thinking-aloud and Probing

Two main techniques have been initiated during the development of CASM, think-aloud interviewing and verbal probing techniques. Think-aloud interviewing is used to explore how the subject arrived at an answer to a specific question. Subjects are asked to utter or think aloud explicitly as they answer the survey questions. The interviewer interjects little, except to say “tell me what you're thinking” when the subject pauses (63).
Advantages of the think-aloud technique include: a) limited interviewer-imposed bias because of little interjection, b) minimal interviewer training requirements, and c) an open-ended format that allows subject to talk about the topic freely.

Drawbacks of think-aloud include: a) requiring subject training before formal interviews, b) producing simple answers without elaboration when subjects are not proficient at thinking aloud, c) placing more burden on subjects, d) subjects may stray from the task and delve into irrelevant areas, and e) introducing bias due to the difference of mental effort in think-aloud speech compared to simply answering the question (65).

As an alternative to think-aloud methods, intensive verbal probing targets the basis of responses by asking probes relevant to a particular question, or to the specific answer given. Verbal probing has two strengths: a) the interviewer can have good control of the interview, and b) it is easy to induce subjects to answer probe questions. However, the limitations of this method include: a) artificial survey environment, and b) potential bias from leading probe questions. Artificiality is a problem in every pretest method. Encouragingly, careful selection of “non-leading” probes can minimize bias (65).

2.5.2 Application to self-administered paper questionnaires

Interviewer-administered cognitive interviews can be effective in developing self-administered questionnaires, although the administration mode of cognitive interviews is different
from that of self-administered questionnaires. There are several reasons why cognitive interviewing is useful developing this type of questionnaires:

a) Problems, such as terminology and question vagueness, can still persist, no matter which administration mode is used. Face-to-face cognitive interviews can be useful for identifying these problems.

b) The subject can interact directly with the to-be-completed questionnaire without interviewer interjection. Retrospective probing is a form of verbal probing that leaves the subject alone as he/she completes the questionnaire and asks probe questions at the end. The environment of retrospective probing is very close to that of a self-administered questionnaire. Retrospective probing can be used close in time to when a situation that evoked questions occurred.

c) Concurrent probing, a form of verbal probing which asks the subject probe questions when he/she completes the questionnaire, can overcome the limitations of think-aloud and retrospective probing. Think-aloud is potentially more disruptive, whereas retrospective probing is prone to identify fewer problems and it may introduce bias because it asks the respondent to reconstruct thoughts to individual questions. Concurrent probing, can help minimize such problems (65).
2.6 Reliability and validity of the questionnaire

The goal of drafting then revising a questionnaire through cognitive interviews to resolve problems regarding comprehension, logic and selection of information is to produce a questionnaire that is both reliable and valid. The reliability of a questionnaire refers to “… the stability and equivalence (or reproducibility) of measures of the same concept over time or across methods of gathering the data” (66). There are three types of reliability: test-retest reliability, inter-rater reliability, and internal consistency reliability.

Test-retest reliability refers to the degree of consistency between answers to the same questions asked of the same respondents at different points in time. It evaluates the stability that a questionnaire has in measuring a concept. Inter-rater reliability examines the equivalence of the information obtained by different data gatherers on the same or comparable groups of respondents. Internal-consistency reliability analyzes the consistency or equivalence of different questions intended to measure the same concept. Internal consistency reliability is used primarily in constructing and evaluating summary scales. Reliability of this questionnaire can be considered in future studies.

The validity of survey questions refers to “…the degree to which there are systematic differences between the information obtained in response to the questions relative to (1) the full meaning of the concept they were intended to express, (2) related questions about the same concept, and (3) theories or hypotheses about their relationships to other concepts.” (66)
There are three types of validity: face validity, criterion validity, and construct validity. Face validity, also called content validity, refers to the extent to which questions can reflect concepts they intended to. The analysis of face validity relies on expert judgments, and it can be conducted in cognitive interviews. Criterion validity refers to the extent to which the survey instrument can predict or agree with the “gold standard”—the criterion of a “true” value for the instrument. Construct validity can examine whether and how many of the relationships predicted by theories or hypotheses do come true when data are analyzed. Without the “gold standard” and definite relationships between survey variables, criterion validity and construct validity cannot be evaluated at the current stage.

2.7 About the proposed questionnaire

Today’s new genomic tests are very complex and diverse. They target different disease sites; they differ in their underlying biological technologies; and they are at different stages in the process of development and validation. To facilitate interpretation of responses, it was decided to focus the questionnaire on one specific genomic test and Oncotype DX™ was selected as it is one of the most developed (29). The test estimates the likelihood of distant recurrence in women who have been diagnostic with early-stage breast cancer. Because of the selected Oncotype DX™, breast cancer was the disease site chosen for the proposed questionnaire. Then scenarios with breast cancer patients were created for the questionnaire.
2.7.1 The selected test—Onco\textsuperscript{type} DX\textsuperscript{TM}

2.7.1.1 Introduction to Onco\textsuperscript{type} DX\textsuperscript{TM}

Onco\textsuperscript{type} DX\textsuperscript{TM} is a 21-gene panel RT-PCR assay that can be used to predict the risk of distant recurrence in node-negative, Estrogen Receptor (ER)-positive breast cancer patients who will be treated with tamoxifen. A Recurrence Score (RS), the output of the test, is used to quantify and classify the predicted risk. A patient with a high-RS (≥31) tumor is believed to have a high risk of recurrence. Besides hormonal therapy, she may obtain more therapeutic benefit from adjuvant chemotherapy (specifically Cyclophosphamide, Methotrexate, and Fluorouracil ([C]MF) ). A patient with a low-RS (≤18) tumor may need hormonal therapy only because of a low risk of cancer recurrence. The clinical benefit of intermediate-RS category (18<RS<30) is still undetermined.

Onco\textsuperscript{type} DX\textsuperscript{TM} can be performed on formalin-fixed paraffin embedded tissue (FPET). This assay is capable of quantifying up to 400 genes simultaneously using high-throughput RT-PCR technology (29). The development of this test consisted of three studies involving 449 patients: 224 patients with node-negative, ER-positive breast cancer treated with tamoxifen, 79 patients with 10 or more positive axillary nodes, and 146 additional patients with operable breast cancer. In developing the test, 250 candidate genes were selected from the published literature, genomic databases, and previous experiments based on DNA microarrays (67). Candidate genes were analyzed using univariate and multivariate Cox models, and genes associated most significantly with distant-recurrence-free survival were selected. Then 16 cancer-related genes
and 5 reference genes made up of the final 21-gene panel. The final panel could provide a variety of information. Some genes could provide information regarding cancer cell proliferation and invasion. Some genes could provide information currently available in clinical practice, e.g., Estrogen Receptor (ER) and Her2/neu status. After the process of gene selection, 16 cancer-related genes were grouped based on function, correlated expression or both. Subsequently, a score for each group was calculated using a weighted average based on the expressions of individual genes in this group. For example, HER2 group score = 0.9 x GRB7 + 0.1 x HER2. Later, a total score, denoted as the Recurrence Score, was calculated as a linear combination of the group scores using coefficients that were predefined in the development studies for cancer-related gene groups and some individual genes that had emerged as being associated with breast cancer recurrence risk. Finally, the Recurrence Score was rescaled from 0 to 100 for convenience. More detail about the RS calculation has been attached in Appendix A.

The Onco\textsuperscript{type} DX\textsuperscript{TM} test was subsequently evaluated in 668 patients with ER-positive, node-negative breast cancer enrolled in the National Surgical Adjuvant Breast and Bowel Project (NSABP) trial B-14. A Cox proportional hazard model indicated that the RS was the only factor that emerged as a significant prognostic factor for 10-year recurrence (hazard ratio=3.21 [2.23-4.61]; p<0.00001). The other two factors used in the analysis were age (hazard ratio=0.71 [0.48, 1.05]) and tumor size (hazard ratio=1.26 [0.86, 1.85]) (67).

The limitations of Onco\textsuperscript{type} DX\textsuperscript{TM} include: a) this test has only been validated retrospectively, b) the clinical benefit of the intermediate risk group is uncertain, c) risk estimates
for the intermediate and high risk groups have wide confidence intervals, d) risk category cutoff values selected by the test developer may be arbitrary and may not be optimal, and e) the test’s prediction of therapeutic benefit from adjuvant chemotherapy may only be restricted to [C]MF (33).

2.7.1.2 Rationale for the selection of this test

Each new genomic test is only valid for one specific cancer situation for a predefined population. One or more test prototypes can illustrate the concept “new genomic test”. However, if two or more tests were included, the questionnaire would place a heavy burden on respondents. Therefore, only one test was chosen to represent the whole series of new genomic tests.

The choice of Oncotype DX™ was based upon the following considerations:

a) Oncotype DX™ is a standardized, multi-gene RT-PCR-based molecular assay performed in a single laboratory, so there is no between-laboratory variation.

b) It uses tissue specimens routinely processed in clinical pathology laboratories.

c) It has received Clinical Laboratory Improvement Amendments (CLIA) approval in the United States.

d) In patients with ER-positive, node-negative early-stage breast cancer, Oncotype DX™ makes prognosis more accurate than standard clinical criteria, including age, tumor size, and histological grade.

e) Its validity has been demonstrated in an independent large population.
f) Preliminary data indicates that the Recurrence Score (RS), the output of this test, may also predict the magnitude of benefit from adjuvant chemotherapy.

g) The target population of this test accounts for nearly 50% of all breast cancer patients diagnosed in the United States, and this population is the group in which patients receive unnecessary adjuvant chemotherapy more often than those with other breast cancer conditions (29).

2.7.2 Breast Cancer

2.7.2.1 Introduction to breast cancer prognosis

Among Canadian women, breast cancer continues to lead in incidence in 2008. Between the ages of 20 and 59, breast cancer remains the most commonly diagnosed type of cancer, and is the leading cause of death, accounting for 36% of all new cancer cases and 24% of cancer deaths in Canada (68).

Types of breast cancers include: ductal carcinoma in situ (DCIS), lobular carcinoma in situ (LCIS), invasive (or infiltrating) ductal carcinoma (IDC), invasive (or infiltrating) lobular carcinoma, and other less common types of breast cancer. In some cases, a single breast tumor can have a combination of ductal and lobular carcinoma, or can have a mixture of invasive and in situ cancer (69).
After an original breast cancer is removed, the cancer can recur locally, loco-regionally, or it can metastasize to other parts of the body (70). Cancer patients who have undergone surgery, radiation therapy, and/or chemotherapy are often given adjuvant (additional) systemic therapy to prevent cancer recurrence. Systemic therapy includes one or more types of the following therapies: chemotherapy, hormone therapy, and targeted therapy (which may contain radiotherapy) (71). During the past 10 years, the increasing use of adjuvant hormonal therapy and chemotherapy has made a great contribution to falling breast cancer mortality (29).

The selection of an appropriate adjuvant therapy for a breast cancer patient depends on the prognosis estimate based on a collection of prognostic factors. The essential prognostic factors well-established to date in primary invasive breast cancer include:

- The patient’s menopausal status (premenopausal/post menopausal, sometimes approximated by age, < 50/ >=50)
- Lymph node status (negative nodes/number of positive nodes)
- Tumor size ( \( \leq 2.0 \text{ cm} \), \( 2.0 \text{ cm} < \text{T} \leq 5.0 \text{ cm} \), >5.0cm )
- Histologic grade (well/moderate/poor-differentiated tumors)
- Histologic type (infiltrating ductal carcinomas, tubular, mucinous, etc.)
- Mitotic figure count
- Hormone receptor status (positive/negative estrogen and progesterone receptor status)
- Tumor Recurrence or Metastasis Following Primary Therapy (e.g., the site of first relapse following primary therapy)
- Extent of response to Local and Systemic Treatment (e.g., patients who have a complete response to doxorubicin-containing chemotherapeutic regimens have a
significant survival advantage compared with those who show partial or no response to treatment) (72).

Among these factors, the lymph node status, menopausal status, tumor size, tumor grade and hormone receptor status are routinely used to predict breast cancer recurrence and are incorporated into current treatment guidelines (67). Other novel prognostic factors are promising as well. For example, Urokinase plasminogen activator (uPA) and plasminogen activator inhibitor (PAI-1) are recommended for breast cancer prognosis by the American Society of Clinical Oncology (ASCO) for node-negative patients (23). Based on these essential and promising factors, a patient’s cancer recurrence risk can be classified into either low or high risk, and sometimes, intermediate risk group if the criterion/guideline that a physician follows includes this risk category. A systemic adjuvant treatment can be recommended then accordingly.

The guidelines that are most commonly used currently to classify breast cancer recurrence risk and guide the provision of adjuvant treatment are: the National Comprehensive Cancer Network (NCCN) guideline (73), the National Institutes of Health (NIH) criterion (74, 75), St. Gallen criterion (76, 77), Nottingham Prognostic Index (78), and Cancer Care Ontario (CCO) guideline (79). In addition to these guidelines, Adjuvant!Online (80), a web-based decision making tool for breast cancer physicians, is a popular prognostic tool that incorporates clinical prognostic factors to predict the 10-year disease-free and overall survival rates of patients.

Current guidelines routinely recommend adjuvant chemotherapy to nearly all patients with larger tumors or positive axillary lymph nodes. However, not all these people are destined to
progress to metastatic disease because patients with the same set of prognostic factor status can actually have markedly different prognoses (81). Preliminary studies suggest new genomic tests in breast cancer may potentially be more accurate than conventional guidelines in classifying patients into low/intermediate/high risk groups (82, 83). These new genomic tests could be very promising in minimizing patients over-treated or under-treated with chemotherapy by refining current guidelines.

2.7.2.2 The development of prognosis in breast cancer

Prognosis research has amongst the longest histories and is among the most developed in breast cancer, of all cancers. For example, the research on estrogen receptor (ER) and axillary node status, has been continuing for more than 30 years (84). Even the prognostic factors at the molecular level, for example her2/neu oncogene and DNA ploidy, have been researched for more than 10 years (85). In the research of new genomic tests, those estimating breast cancer recurrence risk stay at the forefront of this field, with 3 commercial products available and at least 4 tests being developed. The prognosis research in other cancer types, such as lung cancer, prostate cancer, and colon cancer, has not been studied as deeply as in breast cancer, though studies of these cancer types are also of significant importance.

2.7.3 The provision of scenarios

To eliminate variance in responses due to physicians thinking of different clinical situations, it was decided to include case scenarios in the proposed questionnaire. In addition,
knowing the specific situation that respondents would be thinking about provides a means to measure respondents’ opinions that is more valid and more reliable than "simpler" abstract questions (86). Scenarios, also called vignettes, “...are short stories or descriptions of a hypothetical person that are used to investigate a subject’s cognitive processing with respect to survey-relevant decisions.” (63) Survey participants are typically asked to respond to these stories with what they would do in a particular situation or how they think a third person would respond (87). A vignette presented to respondents as concrete and detailed as possible would more closely approximate a real-life decision-making or judgment-making situation. A vignette can be used in self-administered questionnaires and in face-to-face interviews; it can also be presented on audiotape, videotape, and computers (87).

Regarding the novel test Oncotype DX™, providing case scenarios can also help participants better understand the clinical use of the novel test Oncotype DX™, should it be unfamiliar to them. This case-based learning strategy has been widely used in medical education. Therefore, physicians would be comfortable learning about Oncotype DX™ and its usage through scenarios.

2.8 Summary

In summary, new genomic tests have potential prognostic value and clinical implications for individualizing treatment decisions. However, in the implementations of new genomic tests, it is unclear about physicians’ attitudes and concerns about these tests, what information physicians
prefer to have included in the test reports, and how physicians think the use of these tests would impact their treatment recommendations. This study sought to develop a mail-out questionnaire to explore this area of research. Cognitive interviewing was used to evaluate various drafts of the questionnaire.
Chapter 3
Study Design and Methods

3.1 General Study Design

A questionnaire intended for medical oncologists who treat breast cancer patients was
drafted to explore the physicians’ attitudes towards the clinical implementation of new genomic
tests. The draft questionnaire was then evaluated by Kingston oncologists through face-to-face
interviews, followed by medical oncologists in other places in Ontario through telephone
interviews.

3.2 Study Goal

The goal of this thesis study was to design a valid questionnaire that could be used in the
eventual large scale survey to systematically a) describe physicians’ attitudes towards the clinical
use of new genomic tests, b) determine what information physicians prefer to have included in the
reports of the new tests, and c) explore how physicians think the results of the new genomic tests
would impact their treatment recommendations. Thus, the goal of this study was to create the
questionnaire and to establish its face validity.
3.3 Study Objectives

Stage 1: Draft the questionnaire

Objective: To produce a first draft of the self-administered questionnaire.

Stage 2: Obtain feedback from oncologists on the first draft

Objective 1: To ensure that the questionnaire is comprehensible.

Objective 2: To assess the questionnaire’s face validity.

Objective 3: To determine if the target population finds the questionnaire interesting.

Objective 4: To ensure that the questionnaire will not place undue burden on respondents.

3.4 Target population

The target population for our proposed questionnaire was medical oncologists who treat breast cancer patients. This particular sub-group of oncologists was targeted because of the selected test Oncotype DX™. Oncotype DX™ was specifically designed to help medical oncologists with the decision to use chemotherapy to control breast cancer recurrence after patients have had their initial surgery treatment. Therefore, basic familiarity with breast cancer and its prognostic factors was a prerequisite for specifying targeted oncologists.

There are 5 medical oncologists who treat breast cancer in Kingston. Considering that this is a tiny sample and that it is recommended that people other than the targeted group be consulted during questionnaire development (61, 66), we expanded our study sample to all breast
cancer oncology physicians in Kingston. These physicians, including medical oncologists, radiation oncologists, and surgeons, can all shed some light on the development of the questionnaire from different viewpoints because all the physicians are familiar with medical scenarios involving breast cancer patients.

There were no further inclusion or exclusion criteria.

3.5 Stage 1: Drafting the questionnaire

There have been no standard epidemiological methodologies developed to date to study new genomic tests from physicians’ perspectives. Therefore the choice of content to be included in the questionnaire was guided by literatures on development, clinical validation and risk communication of new genomic tests; the layout design was guided mainly by Aday (66) and Dillman’s (88, 89) principles for survey design.

3.5.1 Content considerations

3.5.1.1 Introduction to physicians about Oncotype DX™

Oncotype DX™ is a novel test, and therefore, may not be familiar to medical oncologists in Canada. Thus, the questionnaire needed to include a section to introduce the test. To minimize burden placed on respondents, the introduction section only covered fundamental topics about Oncotype DX™: the appropriate population for testing, the purpose of the test, the nature of the
test, the validation of the test, and the results of the test with explanations (risk group classification and breast cancer Recurrence Score (RS)).

To balance the burden on respondents against having access to the amount of detail they might want about the test, further detailed information was provided in an appendix at the end of the questionnaire. The appendix included: the process of development and clinical validation, its algorithm for computing the Recurrence Score (RS), its benefits and limitations, its certification and regulation, and additionally how the prognosis generated from this test compares to conventional risk classifiers.

Although we chose Oncotype DX™ as the test prototype, we redefined its risk categories because of an ongoing Phase III clinical trial, Trial Assigning Individualized Options for Treatment (TAILORx), which involves Queen’s NCIC Clinical Trials Group. The TAILORx trial had reset the cut-off Recurrence Score for the low/intermediate/high risk group relied on previous clinical studies, thus many local practitioners in the trial would be using the new defined thresholds. Because we were interested in the whole class of new genomic tests, not Oncotype DX™ specifically, it was more important to keep the presentation consistent with current standards of clinical care as indicated by the TAILORx protocol (29) rather than with the standard of Oncotype DX™ developers. The original and the TAILORx trial’s definitions of the low, intermediate, and high risk groups are presented below:
Table 2: Definitions of the risk groups of Oncotype DX™

<table>
<thead>
<tr>
<th>Risk Category</th>
<th>Oncotype DX™</th>
<th>TAILORx</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recurrence</td>
<td>Mean 10-yr Distant Recurrence Rate</td>
</tr>
<tr>
<td>Low</td>
<td>&lt; 18</td>
<td>6.8% (95% CI: 4-9.6%)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>18-30</td>
<td>14.3% (95% CI: 8.3-20.3%)</td>
</tr>
<tr>
<td>High</td>
<td>≥ 31</td>
<td>30.5% (95% CI: 23.6-37.4%)</td>
</tr>
</tbody>
</table>

3.5.1.2 The selected case scenarios

The provision of scenarios not only reduces heterogeneity caused by survey participants thinking about different clinical situation but also helps participants better understand the clinical use of Oncotype DX™. Because the usefulness of Oncotype DX™ could vary according to clinical scenarios, we decided to provide more than one case scenario. The results of Oncotype DX™ can either be concordant with conventional prognostic tests, or be discordant with them. A preliminary study found that the correlation between Oncotype DX™ and Adjuvant! Online is only modest ($R^2=0.43$, $P<0.0001$) (9), which suggests approximately 60% of cases would have discordant predicted results. That is, one case would be predicted by a model as having a high risk of recurrence but she could be classified into low risk group by the other model. We are
particularly interested in how physicians weigh Oncotype DX™ in these types of situations. Considering the response burden of the questionnaire, we limited the study of the implementation situations of Oncotype DX™ to two scenarios.

Scenario A was created so that the risk result of Oncotype DX™ for the patient would be concordant with that of conventional prognostic factors combined together. In theory, one could do so by having a scenario where both the clinical risk estimate and the Oncotype DX™ result classify the patient into the same risk group, either "low (clinically) vs. low (Oncotype DX™)" or "high (clinically) vs. high (Oncotype DX™)". The case scenario “low (clinically) vs. low (Oncotype DX™)” was selected for the questionnaire. We hypothesized that, in such circumstances, the availability of Oncotype DX™ might increase the certainty of physicians’ risk estimates; alternatively, the result of Oncotype DX™ might not be considered useful at all in this setting. The treatment recommendation should stay the same before and after receiving the Oncotype DX™ results.

Scenario B was created so that the result of Oncotype DX™ would be discordant with the prognosis derived from the conventional prognostic factors. The situation “low (clinically) vs. high (Oncotype DX™)” was chosen. To limit other potentially complicating differences between the two scenarios, we elected to provide a second case whose recurrence risk fell in the same risk group with, but was slightly higher than the case in scenario A. Thus, the second case was designed to have a greater uncertainty in the assessment of low risk of recurrence. We hypothesised that the greater the uncertainty in risk estimate was, the more useful the Oncotype
DX\textsuperscript{TM} would be because if a physician was very certain about his/her risk estimate, he/she might not require additional testing. In scenario B, we expected that physicians would have less certainty about their risk estimate after receiving the Oncotype DX\textsuperscript{TM} result; they could even change their treatment decisions.

The proposed questionnaire does not include scenarios associated with an intermediate risk category, either by Oncotype DX\textsuperscript{TM} or by clinical prognostic factors. The clinical significance of Oncotype DX\textsuperscript{TM}’s intermediate risk category is undefined currently. In addition, an intermediate risk group was considered a more difficult scenario in which to detect a physician’s treatment change than a low or high risk group. People in a low risk group only receive none or hormonal therapy, and people in a high risk group receive hormonal therapy plus chemotherapy. However, people in an intermediate risk category can receive hormonal therapy, or hormonal therapy plus chemotherapy. This varied treatment practices would make it more difficult to discern the impact of Oncotype DX\textsuperscript{TM}, specifically, on treatment changes that happened before and after receiving Oncotype DX\textsuperscript{TM} results.

3.5.1.3 The selection of study questions

Questions were selected to achieve three objectives of the questionnaire: to learn about a) physicians’ attitudes, b) their information needs, and c) the test’s impact on their treatment decisions. Description of each objective’s questions and the rationale for their inclusion are summarized below:
a) Attitudes:

Four baseline questions were asked first, relating to each scenario: “risk classification”, “quantitative risk estimate”, “treatment recommendation”, and “confidence about treatment”. These questions were asked to learn about the physicians’ clinical judgments based on conventional prognostic factors towards the case in each scenario.

After the Oncotype DX result was provided, the attitude questions asked addressed: “change in certainty of the risk estimate”, “the test’s usefulness in making a treatment recommendation”, “general willingness to recommend the test to this type of patient”, and “If no, what are the concerns”. We believed the implementation of a new genomic test would depend on physicians’ positive attitudes towards it. If physicians deemed that the information derived from the test is helpful, if the use of the test could provide some benefit (e.g., changes of certainty level or/treatment recommendation), or if physicians would be willing to recommend the test to a particular patient, we would conclude that the physicians would have positive attitudes toward the test.

Two other questions were asked independent of scenarios: “information useful in deciding on treatment” and “information useful in communicating with the patient”. We believed that the two questions could help determine how to make the test more useful in clinical practice from physicians’ perspectives.
b) Information needs:

“Types of risk classification” (a question asked). This is a central issue because the classification of cancer recurrence risk is the result from every prognostic test but different genomic tests present distinct types of risk classification. For example, some classify into low/high risk groups, and others classify into low/intermediate/high risk groups.

“Average rate for patients in a risk category”, and “individual risk for a patient”. We were interested in identifying whether the information about the average rate in each risk category is enough for physicians, or if they would prefer to know the specific rates for individual patients, or if they would want to know both. Currently, there are no guidelines for providing recurrence risk information, and different new genomic tests provide different pieces of the information.

“Cancer recurrence risk over 5 or 10 years”. Low risk in distant recurrence over 5 years is different from that over 10 years. The validation study of Oncotype DX™ found that about two-thirds of relapses in the low risk group occurred between 5 and 10 years (29). Currently, some new genomic tests in breast cancer provide recurrence risk at 5 years (90); others provide risk at 10 years (e.g., Oncotype DX™).

“Test results”. We intended to find out whether physicians would like to know the results of some specific genes of the 21 gene panel, especially the well-known ones, for instance, Estrogen Receptor (ER) gene, Progesterone Receptor (PgR) gene, and HER2/neu gene. The main reasons are that Oncotype DX™ could provide the information about the expression of ER, PgR and HER2/neu genes and that the status of ER, PgR, and HER2/neu are pieces of information currently provided in a pathological report. However, it is unclear if physicians want such
information when it is included in the overall Oncotype DX™ test result. It is interesting to note that Oncotype DX™ may have the potential to replace the conventional immunohistochemistry (IHC) in future for detecting the positivity of ER and PgR proteins, and the amplification of HER2/neu gene. Thus far, the RT-PCR platform that Oncotype DX™ relies on could define positive ER status as well as immunohistochemistry (IHC) technology (k statistics ranging from 0.80 to 1) (91-95). With the development of RT-PCR technology, Oncotype DX™ might also work well on PgR and HER2/neu status. If so, Oncotype DX™ could provide detailed information on specific genes in the test report.

“Test itself”. Regarding the information needs about the test per se, we have mainly covered the following topics: the accuracy of the test, what is measured and which technology is used, its development and validation trials involved, and its comparison with existing risk classifiers (Ajuvant!Online, exclusively). This information, we believe, is important for physicians to understand the test, assess its accuracy, and explain the test results to patients.

“Prognostic model”. Researchers are devoted to integrated models which combine genomic information with clinical data for personalized prediction in cancer recurrence (96). However, an integrated model is not always preferred; sometimes physicians may prefer to see various prognostic factors separately, and to give their own judgments. In addition, some prognostic factors may not be significant in statistical analysis of prognostic tools; however, they remain important in physicians’ decision-making, for example, a patient’s age.

“Words vs. numbers in risk expression”, and “recurrence rates vs. survival rates”. Risk communication is also the communication of probability. Although it makes little mathematical
difference whether statistical results are expressed in words or numbers, recurrence rates or survival rates, it does make a psychological difference (97-100). Different ways in which recurrence risks are expressed can have distinct impacts on risk perception and the following treatment decisions-making (97). We hoped that risk communication questions in this questionnaire could provide some insight into how physicians would like recurrence risk to be presented.

“95% confidence interval”. The provision of a 95% confidence interval can present the uncertainty of risk estimates directly. For example, a low risk estimate may have a 95% confidence interval whose upper bound falls into the intermediate or high risk category. That is, the true recurrence risk may be low, intermediate or even high.

“Visual formats”. There is evidence that people understand different formats with differing ease and accuracy (54, 101). By presenting diverse graphs of the test results, we hope to find one or more graphs that could help physicians understand the test results most accurately and effectively. The design of visual formats in the questionnaire consulted current new genomic tests for cancer diagnosis and prognosis, coupled with graph design theories for quantitative information.

c) Oncotype DX™’s impact on treatment decisions:

Questions in this part are scenario-based questions. Questions, for example, “how the test influences his/her certainty of recurrence risk estimate”, “whether the test is helpful in treatment decision-making”, and “whether the test result has changed his/her treatment decision”, could shed some light on the potential effect of Oncotype DX™ in clinical practice.
3.5.2 Layout issues

The questionnaire design described below was guided by Aday’s (66) and Dillman’s guidelines (88, 89). These two guidelines complement each other in identifying and addressing formatting issues.

3.5.2.1 The cover pages

The front cover. The front cover can attract and motivate respondents to participate in the survey by establishing trust with respondents and creating feelings of connectedness and importance among respondents (89). To be simple and official looking, our front cover had: a) a short, simple title, b) a simple, distinctive and neutral graph (for a positive first impression), c) the sponsor’s name and address (for fostering trust that the survey is legitimate and useful), and d) initial directions on the opposite page (88, 89).

The back cover. The back cover is the place to invite respondents to make additional comments, and to thank them for their participation. The back cover of the questionnaire was much simpler than the front cover, so that it would not detract from the front cover.

3.5.2.2 Questions

Words. “Clarity” is the criterion used to evaluate the chosen words (66). In the questionnaire, we tried to choose the words which could capture or convey the concept of interest adequately and clearly, and could be understood by respondents completely and clearly.
**Phrases.** The criterion for assessing appropriate phrases is the relative balance among the words which constitute phrases (66). In the questionnaire, three dimensions of “balance” were considered: a) addressing both sides of a question or an issue adequately, b) avoiding loading response options in one direction or another, and c) avoiding double-barreled questions.

**Sentences.** The length of sentences that compose questions is the criterion for evaluating survey questions (66). For our mail-out questionnaire, I tried to use short questions. If a medium-length question (16 to 24 words) or a long question (25 or more words) had to be used, a moderate introductory text (16 to 64 words) was considered in the first place in conjunction with the question.

**Question groupings.** Question grouping is to organize questions logically. In this study, questions were grouped by content and were kept in a logical order. Different colors were applied to convey groupings, as well as spacing.

**Question ordering.** Question ordering effects have to be eliminated or kept to a minimum. When constructing our questions, we tried to avoid the following ordering effects identified by Dillman (66, 88): the tendency for “norm of evenhandedness”, “anchoring or cognitive-based effect”, “addition (carryover) effect”, “subtraction effect”, and “increased positiveness of summary items when asked after specific items on the same subject” (see glossary for explanations of these terms).

**Question layout.** Questions have to be presented clearly, logically and consistently in appearance. During the design process of the questionnaire, great attention was paid to: a) font
type, font size, and indentation, b) the length of the questionnaire, c) blank space between
questions and between line spacing within questions, d) print for questions and for response
options, e) answer spaces especially on the same page, f) the display of response options, g) a box
to allow putting “X” in front of each response option, h) visual elements to improve skip pattern
appearance, and i) abbreviations.

3.5.2.3 Response options

*Open-ended questions.* An open-ended format generally encourages respondents to talk
about what is at the top of their heads or what comes to mind first without limitation (66). A few
open-ended questions were placed in the questionnaire because there were no standard or obvious
response categories that could be provided. One additional open-ended question was provided at
the end of the entire questionnaire as a debriefing section for respondents to write down the
problems they had encountered, as well as any suggestions or important issues that we missed.

*Closed-ended questions.* We tried to provide comprehensive response categories to
closed-ended questions.

*Partially closed-ended questions.* A partially closed-ended question is a question with an
open-ended response option among other closed-ended options. One feature of our questionnaire
is that most questions are partially closed-ended questions. These questions can help elicit
respondents’ feedback if a question is ambiguous, or if response options are not adequately
appropriate or comprehensive.
Response-order effects. A response-order effect (66) often arises when the content of a question is general, and/or a response that a question requires is vague (102). In the questionnaire design, the following response-order effects were minimized: “primacy effects” and “recency effects”. Two response tendencies, central tendency and extreme tendency, were minimized as well. (See glossary for explanations of these terms).

Rating scales. A 5-point rating scale is expected to have the highest reliability and validity (66). We followed the suggested 5-point rating scale in designing our questionnaire. A typical set of response options provided in the questionnaire is: “Definitely would not / Probably would not / Unsure or neutral / Probably would / Definitely would”.

3.5.2.4 The overall questionnaire

Instructions. Instructions aim to ensure that there is no misunderstanding between response options interpreted by researchers and by responders (66). In our draft questionnaire, general directions were placed at the beginning of the questionnaire. Relevant specific instructions were placed immediately before where they were to be applied. For example, the subheading of a particular set of questions, served as a brief instruction for that set of questions.

Navigational path. The navigational path describes how the questionnaire branches, and when particular paths should be taken. In the questionnaire, we used visual guides to make the navigational path clear, and to inform respondents which path they should take (89).

Survey length. Physicians’ interest and patience were considered when determining the most appropriate length for the survey (89).
Page design. The criteria for page design are “professional looking” and “easy to follow”. A booklet format is usually recommended (89). For the design of our booklet, we chose white letter size paper, printed it double-sided, assembled it into a booklet, and stapled on the spine. No questions on the front or back cover. In addition, an envelope with a return address printed on it will be provided.

Survey color. Attention was paid on the contrast effect of colors. Colors believe to tire eyes (66) or carry with them strong emotional reactions (103) were avoided.

Other issues. Other details of guidelines that we followed are summarized in a table (major DO’s and DON’T’s of questionnaire design) provided in Appendix B.

3.5.2.5 The cover letter in the interview stage

A cover letter which introduced the research project and the interviewing process was provided with the questionnaire to participants in interviews. The design of the cover letter followed Dillman’s methods (61, 66, 88, 104) for mail-out survey development. The following items were included: a) the current date, the participant’s name, b) names of the researchers who are conducting this study and organizations involved, c) one medium-length (16 to 64 words) introduction to the study, and an introduction to the interviewing process, d) the reason why the recipients and their participation are important, e) a conservative estimate of how long the survey will take, f) ethical issues, and g) whom to contact (66). At the end, it also included a handwritten signature in a color that was in contrast to the black type displayed on the letter to make this letter more personalized (66).
3.6 Stage 2: Interviewing oncologists

3.6.1 Sampling method and recruiting strategy

3.6.1.1 Sampling method

Purposeful sampling was used in the study. Purposeful sampling is usually adopted when recruiting people is difficult, and the chosen individuals are believed to have ideas or experiences that will help achieve the goal of the research (105).

3.6.1.2 Recruiting strategy—local oncologists

First, we contacted the chair of Breast Cancer Disease Site Group at Cancer Centre of Southeastern Ontario, Kingston General Hospital, and asked for an opportunity to make a presentation about our study at their Site Group regular meeting at Center. During my presentation, I invited them to take part in my study. In addition, I asked for advice on how they recommended that I recruit participants to the project. It was suggested that I send an email to each breast cancer oncologist for his/her feedback on my draft questionnaire. Later I sent a follow-up email to the chair for the email list of all breast cancer oncologists in Kingston. An invitation email was sent to everyone on the list (a copy of the invitation email is provided in Appendix C).

In addition, an internet search about the Kingston General Hospital Physician referrals and the members of the Breast Cancer Disease Site Group on Cancer Care Ontario website identified two Kingston physicians who were not on the list provided by on-site chair. An invitation email was then sent to each of them.
3.6.1.3 Sample increment

Because of the small number of local oncologists, we allowed for expanding the sample (should the stopping rule, described in the “interview process” section, not be met) by approaching medical oncologists in Ontario outside Kingston. Since the stopping rule was, indeed, not met in the Kingston sample, expanding the sample was required. The Cancer Care Ontario website provides a list of members of the Ontario Breast Cancer Disease Site Group. Medical oncologists on the list (N=9) were identified subsequently. Then, 6 of these medical physicians were picked out randomly, and an invitation email was sent to each of them requesting for a telephone interview (approximately 20 minutes each). For non-respondents, a second email notice, guided by Dillman’s tailored design, was sent a week after the initial email with a goal of recruiting 3 participants per additional round required.

3.6.2 Interview process

3.6.2.1 Interview type

For Kingston oncologists, a semi-structured interview guided by the questionnaire itself was the basis for in-person individual interviews. For medical oncologists in other places, a structured interview was conducted with each of them over the telephone.

3.6.2.2 Interview strategy—Cognitive interviews

“Verbal probing” was used in the interviewing process, to ensure that the questionnaire
was both rational and meaningful to respondents. This strategy consists of two approaches: a) asking specific probe questions at the time a question is asked (concurrent probing), and b) asking probe questions at the end of the interview (retrospective probing). Considering that oncologists might have different debriefing habits, both of these “verbal probing” approaches were used in the interview. Concern that too many concurrent probing questions might interrupt the flow of physicians’ thinking, meant that the concurrent probing was applied only when respondents were stuck on specific questions, or they paused for clarification or to raise any issues.

There are two types of probe questions involved: “scripted probes” and “spontaneous probes”. “Scripted probes”, applied in retrospective probing, are probes developed prior to the interview; “spontaneous probes”, applied in concurrent probing, are probes that are usually “thought up” during the interview as a result of how the respondent reacts to the questionnaire (65). Scripted probing questions in our interviews covered issues of face validity, comprehensibility, etc. to initially assess the quality of the questionnaire. Spontaneous probes were generated, according to the interviewing conditions, from the pool of probes based on Willis’ guidelines for developing standard cognitive probes (63).

3.6.2.3 Cognitive interview process

The sequence of the interviewing process followed the cognitive interviewing guide from U.S. National Cancer Institute (65). It was made up of two stages: preparation stage for interviewing and following cognitive interviewing rounds. In the preparation stage, the tasks
included reviewing the questionnaire, making suggestions for modifications prior to testing, and
developing basic probes to use in the first round of interviewing. In the stage of interviewing
rounds, three-to-four in-depth individual interviews were conducted in each interviewing round.
Findings from each interview were summarized on a question-by-question basis, by entering
comments directly under each question (65). Then suggestions for revising questions were made.
After each interview round, suggestions were discussed, and the feasible modifications were
decided with my supervisors. The revised questionnaire was then assessed in the next round of
interviews. The interviewing rounds could stop when no new issues were identified by any
respondents in the round. The flowchart of cognitive interviewing activities has been provided in
Appendix D.

During the in-depth interviews, the behavior of “hesitating before answering”, “answer
erasures or modifications”, or “skipping a question by error” and facial expression of “fatigue”,
“time-consuming” or “confusing” of an interviewee were observed and marked under particular
questions (nonverbal communication) for probing problems later. Additionally, a tape recorder
was used, with approval of the interviewee, to record each interview dialogue. This process
facilitated the data collection process, and was a great help in identifying and revising
problematical questions (66).

3.6.2.4 Phone interviews

The invitation email in phone interviews was modified from the one for face-to-face
interviews. To increase the credibility of the study, the modified invitation email emphasized that,
the Breast Cancer Site Group in Kingston had already evaluated the draft questionnaire, and we were soliciting further feedback on the revised questionnaire. Once potential participants replied to our invitation email, the draft questionnaire and the cover letter were mailed to them immediately. The electronic versions were also attached in emails in case the mailed materials were delayed. In order to keep the interview as short as possible, and to use the time most efficiently, we asked both in emails and in mailed materials that the respondent fill out the questionnaire ahead of time.

Phone interviews were conducted following Dillman’s guidelines for phone interviews (61, 104). Scripted probes in face-to-face cognitive interviews were used as the probe questions in phone interviews. For those who filled out or had looked at the draft questionnaire before interviews, only probe questions were asked. One participant who had not looked at the questionnaire ahead of time was informed that the phone interview would be longer than anticipated (20 minutes anticipated). Then the participant was instructed to finish the questionnaire. Interviewer interventions were kept as few as possible. Tape recording was not used in phone interviews. Great attention was paid to non-verbal information, e.g., changes of the tone of voice, long periods of silence etc.

3.6.3 Scripted probe questions

Scripted probes were developed prior to the interview stage to initially assess the quality of the questionnaire. The scripted probes and the issues they covered are identified below.
<table>
<thead>
<tr>
<th>Issues</th>
<th>Targeted section</th>
<th>Interview Round</th>
<th>Probes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face validity</td>
<td>Q2</td>
<td>1</td>
<td>Were the answer options for question 2 designed properly?</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>Q10</td>
<td>1</td>
<td>How was it for you to go through the answer list in question 10? Did that cause any difficulties?</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>Q26</td>
<td>1</td>
<td>I used terms “ER gene”, “ER status”, and “Her2/neu gene amplification” in this question. Did they cause any problems?</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>Q27</td>
<td>1</td>
<td>I know this question might be a little burdensome. Did any terms cause any problems?</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>Appendix section</td>
<td>1</td>
<td>In general, how do you feel about this appendix?</td>
</tr>
<tr>
<td>Face validity</td>
<td>Scenarios</td>
<td>1</td>
<td>1. Did these two scenarios make sense to you?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Are there any case-scenarios that you think may be more appropriate/representative?</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>Title</td>
<td>1</td>
<td>Are you satisfied with the name of the questionnaire?</td>
</tr>
<tr>
<td>Face validity</td>
<td>Introduction to Oncotype DX™</td>
<td>1</td>
<td>How about the format of the test introduction section? (e.g., the bullet format to present prognostic information?)</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>Questionnaire</td>
<td>2, 3, 4</td>
<td>Were there any terms that were difficult to understand?</td>
</tr>
<tr>
<td>Face validity</td>
<td>Questionnaire</td>
<td>1, 4</td>
<td>What additional information would you like to have included in the questionnaire?</td>
</tr>
<tr>
<td>Comprehensiveness</td>
<td>Questionnaire</td>
<td>2, 3</td>
<td>Is there any important information that we missed to include in the questionnaire?</td>
</tr>
<tr>
<td>Face validity</td>
<td>Questionnaire</td>
<td>1, 2, 3, 4</td>
<td>What information/questions do you think should be deleted from the questionnaire?</td>
</tr>
<tr>
<td>Necessity</td>
<td>Questionnaire</td>
<td>1, 2, 3, 4</td>
<td>In general, are you satisfied with the structure, format/style of our questionnaire?</td>
</tr>
<tr>
<td>Format</td>
<td>Questionnaire</td>
<td>1, 2, 3, 4</td>
<td>Were the skip patterns easy to understand? Were you confused? Because the skip patterns in the questionnaire were not exactly the same.</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>Skip pattern</td>
<td>4</td>
<td>We used arrows in some questions, so respondents could skip one or two questions and moved forward to other questions down the survey. Were you confused by these arrows?</td>
</tr>
<tr>
<td>Response burden</td>
<td>Questionnaire</td>
<td>1, 2, 3, 4</td>
<td>1. Is it difficult for you and your colleagues to answer this questionnaire?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Is it burdensome?</td>
</tr>
</tbody>
</table>
In the first round of interviews, our probes targeted both particular question wordings and major concepts that the questionnaire was attempting to cover. In the later rounds of interviews, we mainly assessed the quality of the questionnaire, such as comprehensiveness, face validity, etc. Questions targeting particular questions/response options in the questionnaire were excluded. However, respondents could still raise issues regarding wording or understanding of specific questions/response options during interviews.

**3.6.4 Practice interviews**

Before conducting the face-to-face interview with local oncologists, two practice interviews were conducted with non-oncologists in the Division of Cancer Care and Epidemiology, Cancer Research Institute. Suggestions about the interviewing process and English wording of particular questions guided further revisions. The draft of the self-administered questionnaire after this stage was used in the first round of interviews with oncologists.
3.6.5 Qualitative data analyses for interviewing

Each step in these analyses involved a significant degree of judgment and interpretation according to our research goals. The method of analyzing cognitive interview results described below was based on Willis’ guidelines (63).

3.6.5.1 Characterizing cognitive interview outcomes

The interview outcomes, in our case, are the types of problems in questionnaire design, have been classified into the following categories:

a) Recommendations for changes to wording
b) Need to clarify the intent
c) Problems related to ordering/formatting issues
d) Respondents’ interest in the questionnaire
e) Overall burden
f) Limitations of the questionnaire

The above six categories covered the main questions/issues usually found in the questionnaire design. The interpretation of these outcome categories follows below:

a) Recommendations for changes to wording: It captures the inappropriate/confusing words identified. Recommendations for changes of these words can be made accordingly.

b) Need to clarify the intent: This category is to evaluate whether concepts are conveyed precisely in the questionnaire. In other words, we aim to ask the correct questions (measuring the constructs they were supposed to measure) and warrant that the objectives of the questions are clear. The “Need to clarify the intent” category encompasses two
dimensions. One dimension concentrates on understanding/effective communication. When the intent of an item (question/response option/other part) is unclear, a need to clarify/specify the intent of the questionnaire component is identified. The other dimension refers to the face validity of the questionnaire. It detects two failings in the design of the questionnaire: bias in the design and incompleteness of the questionnaire items. Bias regarding a question exists when a respondent thought he/she should answer the question in a certain way, when a respondent was unable to answer the question, or when a question which did not reflect the concept accurately was asked. Biased response options arise when response options were classified inappropriately, or when one response option was more socially desirable than the others. Bias can also be introduced in other parts of the questionnaire when improper information was provided. With respect to the completeness issue, the purpose is to ensure that all but only necessary questions have been covered, the widest possible range of answers has been captured, and no important issue/information has been missed.

c) Problems related to ordering/formatting issues: This outcome category collects issues of question ordering effects or other interactive effects, in order to address problems regarding the interaction between questions or between questionnaire sections.

d) Respondents’ interest in the questionnaire: This outcome category intends to investigate this issue. We were aware during the interviews, whether people were interested in the questionnaire and whether there were offensive/off-putting questions. These aspects were
judged by interviewees’ oral comments and their non-verbal expressions in the interviewing process.

e) Overall burden: This category assesses the burden of the questionnaire, because a questionnaire with a reasonable burden on the target population is a key factor to obtain a high response rate. This category focuses both on individual questions and on the questionnaire as a whole. In the two aspects, respondents’ comments and facial expressions of impatience/irritation/frustration/fatigue were used in the assessment.

f) Limitations of the questionnaire: As a medium-length self-administered questionnaire regarding novel tests, there are limitations on what can be asked of survey respondents. Therefore, some issues were raised by respondents that were deemed to be beyond the scope of the questionnaire, and these issues would be categorized in this section.

3.6.5.2 Analysis of interview results: coding schemes

A coding system can be applied to outcome categories for the purpose of method evaluation, rather than as a means to guide the review and modification of questions (63). The codes of problem types in our study have been presented in the following table:

<table>
<thead>
<tr>
<th>Problem code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehension/communication of intent</td>
<td>Reflecting that the subject has difficulty understanding or has comments on: a) semantic issues b) the intent of a question/response option/other parts of the questionnaire</td>
</tr>
</tbody>
</table>
Bias/sensitivity (of the questions/response options)

Capturing:
- whether there is any unbalanced or leading question
- whether there is any incorrect question asked
- whether there is any question unanswerable
- whether one response option seems to be the correct answer over the others

Response category

Detecting whether the given categories:
- match the answers that people normally use
- have covered all practical answers

Logical issues

Detecting whether there are fundamental problems in the design of the questionnaire

The above codes of problem types were assigned to cognitive interview outcomes, in order to identify the major problems in the design of the questionnaire.

3.6.5.3 Outcome categories and problem codes assigned to each objective

Objective 1: To ensure that the questionnaire is comprehensible

Outcome categories:
- Recommendations for changes to wording
- Need to clarify the intent
- Problems related to ordering/formatting issues

Problem Codes assigned:

“Comprehension/communication of intent”, and “logical issues”.

Objective 2: To assess the questionnaire’s face validity

Outcome categories:
- Need to clarify the intent

Problem Codes assigned:

“Bias/sensitivity”, “response category”, and “logical issues”.

61
Objective 3: To determine if the target population finds the questionnaire interesting

**Outcome categories:**

a) Respondents’ interest in the questionnaire

**Problem Codes assigned:**

“Bias/sensitivity” and “logical issues” for detecting offensive/off-putting questions.

Objective 4: To ensure that the questionnaire will not place undue burden on respondents

**Outcome categories:**

a) Overall burden

b) Limitations of the questionnaire

**Problem Codes assigned:**

“Comprehension/communication of intent”, and “logical issues”.

### 3.7 Ethical considerations

This study was approved by the Queen’s University and Affiliated Teaching Hospitals Health Sciences Human Research Ethics Board on June 4th, 2008 (please refer to Appendix E for a copy of the ethics approval form). A consent form was not required from participants. An introduction letter (cover letter) explaining the study, the interview process and the ethical issues was provided instead (please refer to Appendix F).

All information obtained during the course of this study was strictly confidential and the anonymity of physicians would be protected at all times. Data were stored in locked files and was only available to principal investigators of this study. Physicians would not be identified in any publication or reports.
Chapter 4

Results

4.1 Stage 1: Drafting the questionnaire

Wordings as well as formatting problems were revised iteratively, which resulted in seven drafts before the questionnaire was deemed to be adequately clear to solicit feedback from oncologists in interviews.

In the following table, the generated survey questions are linked to the three objectives: a) to describe physicians’ attitudes towards the clinical use of new genomic tests, b) to determine what information physicians prefer to have included in the reports of the new tests, and c) to explore how physicians think the results of the new genomic tests would impact their treatment recommendations. In addition, the type of measurement, and the questionnaire section of survey questions are presented.

<table>
<thead>
<tr>
<th>Question # (version 1) *</th>
<th>Study question</th>
<th>Type of measurement</th>
<th>Objective</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Part I</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Scenario A</strong></td>
<td><strong>A patient scenario with concordant risk results (clinical prognostic factors vs. OncoType DX™)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>Baseline recurrence-risk classification</td>
<td>Nominal</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Baseline quantitative risk estimate</td>
<td>Ordinal</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Baseline treatment recommendation</td>
<td>Nominal</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Baseline confidence about treatment</td>
<td>Ordinal</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Introduction to OncoType DX™</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>OncoType DX™ test result for this patient is presented</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Change in certainty of the risk estimate</td>
<td>Nominal</td>
<td>a, c</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>Scale</td>
<td>Option</td>
</tr>
<tr>
<td>---</td>
<td>------------------------------------------------------------------------------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td>6.</td>
<td>The test’s usefulness in making a treatment recommendation</td>
<td>Ordinal</td>
<td>a, c</td>
</tr>
<tr>
<td>7.</td>
<td>Whether the test can change a treatment recommendation</td>
<td>Ordinal</td>
<td>c</td>
</tr>
<tr>
<td>8.</td>
<td>If yes, what is the new treatment recommendation</td>
<td>Nominal</td>
<td>c</td>
</tr>
<tr>
<td>9.</td>
<td>General willingness to recommend the use of the test</td>
<td>Ordinal</td>
<td>a</td>
</tr>
<tr>
<td>10.</td>
<td>if no, what are the concerns</td>
<td>Nominal</td>
<td>a</td>
</tr>
</tbody>
</table>

**Part II**

**Scenario B**

*Clinical prognostic factors vs. Oncotype DX™*

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Scale</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.</td>
<td>Baseline recurrence-risk classification</td>
<td>Nominal</td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>Baseline quantitative risk estimate</td>
<td>Ordinal</td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>Baseline treatment recommendation</td>
<td>Nominal</td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>Baseline confidence about treatment</td>
<td>Ordinal</td>
<td></td>
</tr>
</tbody>
</table>

**Oncotype DX™ test result for this patient is presented**

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Scale</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.</td>
<td>Change in certainty of the risk estimate</td>
<td>Nominal</td>
<td>a, c</td>
</tr>
<tr>
<td>16.</td>
<td>The test’s usefulness in making a treatment recommendation</td>
<td>Ordinal</td>
<td>a, c</td>
</tr>
<tr>
<td>17.</td>
<td>Whether the test can change a treatment recommendation</td>
<td>Ordinal</td>
<td>c</td>
</tr>
<tr>
<td>18.</td>
<td>If yes, what is the new treatment recommendation</td>
<td>Nominal</td>
<td>c</td>
</tr>
<tr>
<td>19.</td>
<td>General willingness to recommend the use of the test</td>
<td>Ordinal</td>
<td>a</td>
</tr>
<tr>
<td>20.</td>
<td>If no, what are the concerns</td>
<td>Nominal</td>
<td>a</td>
</tr>
</tbody>
</table>

**Part III**

**Information preferred to be included in the test report**

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Scale</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.</td>
<td>Type of risk classification</td>
<td>Nominal</td>
<td>b</td>
</tr>
<tr>
<td>22.</td>
<td>Average recurrence rate of each risk category</td>
<td>Ordinal</td>
<td>b</td>
</tr>
<tr>
<td>23.</td>
<td>If yes, the preference of average recurrence rate at 5/10 years</td>
<td>Nominal</td>
<td>b</td>
</tr>
<tr>
<td>24.</td>
<td>Individual recurrence rate</td>
<td>Ordinal</td>
<td>b</td>
</tr>
<tr>
<td>25.</td>
<td>If yes, the preference of individual recurrence rate at 5/10 years</td>
<td>Nominal</td>
<td>b</td>
</tr>
<tr>
<td>26.</td>
<td>Other test results that Oncotype DX™ report can provide</td>
<td>Ordinal</td>
<td>b</td>
</tr>
<tr>
<td>27.</td>
<td>Test itself that Oncotype DX™ report can provide</td>
<td>Ordinal</td>
<td>b</td>
</tr>
<tr>
<td>28.</td>
<td>Anything else to be included in the test report</td>
<td>Open-ended</td>
<td></td>
</tr>
</tbody>
</table>

**Formats**

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Scale</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>29.</td>
<td>Presentation of the risk (integrated estimate vs. separated estimate for each indicator)</td>
<td>Nominal</td>
<td>b</td>
</tr>
<tr>
<td>30.</td>
<td>Verbal expression or risk information</td>
<td>Nominal</td>
<td>b</td>
</tr>
<tr>
<td>31.</td>
<td>Verbal expression or risk information</td>
<td>Nominal</td>
<td>b</td>
</tr>
<tr>
<td>32.</td>
<td>95% confidence interval</td>
<td>Ordinal</td>
<td>b</td>
</tr>
<tr>
<td>33.</td>
<td>Recurrence rate or recurrence-free survival rate</td>
<td>Nominal</td>
<td>b</td>
</tr>
<tr>
<td>34.</td>
<td>Visual formats</td>
<td>Nominal</td>
<td>b</td>
</tr>
</tbody>
</table>

**Usefulness of the test**
4.2 Stage 2: Interviewing oncologists

4.2.1 Participants

Eight local oncology physicians, 1 local pathologist, and 4 oncologists outside of Kingston participated in 4 rounds of interviews in the study. The characteristics of the participants are summarized below:

<table>
<thead>
<tr>
<th>Round</th>
<th>Subject ID</th>
<th>Gender</th>
<th>Specialty</th>
<th>Interview type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>M</td>
<td>Surgery</td>
<td>Face-to-face</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>M</td>
<td>Radiation Oncology</td>
<td>Face-to-face</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>M</td>
<td>Radiation Oncology</td>
<td>Face-to-face</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>F</td>
<td>Surgery</td>
<td>Email</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>M</td>
<td>Pathologist</td>
<td>Face-to-face</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>F</td>
<td>Medical Oncology</td>
<td>Face-to-face</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>F</td>
<td>Medical Oncology</td>
<td>Face-to-face</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>M</td>
<td>Medical Oncology</td>
<td>Face-to-face</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>F</td>
<td>Radiation Oncology</td>
<td>Face-to-face</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>M</td>
<td>Medical Oncology</td>
<td>Face-to-face</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>F</td>
<td>Medical Oncology</td>
<td>Phone</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>M</td>
<td>Medical Oncology</td>
<td>Phone</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>M</td>
<td>Radiation Oncology</td>
<td>Face-to-face</td>
</tr>
</tbody>
</table>
4.2.2 The organization of the Interview Results

The Interview Results are presented below for each round of interviews, successively.

Within each interview round, the results are grouped by objectives. Within each objective, suggestions and comments are categorized into classes of interview outcomes that clarify the types of the problems. The results of each outcome are presented in a table. The table has four columns: a) “Questionnaire section” indicates the questionnaire component in which the problems/issues were identified; b) “Question # (version 1)” includes the question number in the first version of the questionnaire to which the comment/suggestion is directed; c) “Change suggested” is the change or comment made by participants; and d) the “Change made” column reports the change actually made after consulting with supervisors, indicated as either “Yes” or “No”. For cases when problems were identified but solutions were not suggested, the change finally made is recorded in the “change made” column directly. In the table, “--” denotes that no issues were raised, and no changes were made. The first version and the final version of the questionnaire are provided in the appendix for ease of reference (first version: Appendix G; final version: Appendix H).
4.2.3 Round 1

4.2.3.1 Objective 1: To ensure that the questionnaire is comprehensible

Outcome category: recommendations for changes to wording

Suggestions related to this category of outcome are listed separately for the questions, for the response options, and for other parts of the questionnaire (scenarios, introduction to Oncotype Dx™, headings, directions, and skip patterns, etc.).

Table 7: Results of "recommendations for changes to wording"
for Objective 1, Round 1

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions</td>
<td>29, 31</td>
<td>Change “what would you prefer” to “how would you prefer”.</td>
<td>Yes</td>
</tr>
<tr>
<td>Responses</td>
<td>6</td>
<td>Change “Probably would not” to “Possibly would not”</td>
<td>No (We followed the 5-category scale recommended)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>- Change “this test may have not been used widely” to “this test has not been used widely”</td>
<td>Yes (Similar change was made to the question 20 following scenario B)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Similar grammar corrections should be made to other response options</td>
<td></td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>Correct mistake in the first graph: changing “low risk” to “high risk”</td>
<td>Yes</td>
</tr>
<tr>
<td>Other parts</td>
<td>Skip patterns</td>
<td>Change “go to next page” to “proceed to question #”</td>
<td>“go to question #”</td>
</tr>
</tbody>
</table>

Outcome category: need to clarify the intent

As above, the suggestions are identified for specific questions, for response options, and for other parts of the questionnaire.
Table 8: Results of "need to clarify the intent" for Objective 1, Round 1

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions</td>
<td>2</td>
<td>Add “without adjuvant treatment” to the question</td>
<td>Yes (Similar change was made to the question 12 following scenario B)</td>
</tr>
<tr>
<td>Questions</td>
<td>7</td>
<td>Add the phrase “you made before” at the end of question</td>
<td>Yes</td>
</tr>
<tr>
<td>Questions</td>
<td>26</td>
<td>ER status should be addressed clearer</td>
<td>Add “(positive/negative)” after “The ER/PR status determined by the test”</td>
</tr>
<tr>
<td>Questions</td>
<td>26</td>
<td>The second items seemed to have included the first item. They should be separated from each other.</td>
<td>Add the word “only” in the first item, and bold this word</td>
</tr>
<tr>
<td>Response options</td>
<td>21</td>
<td>A respondent suggested no risk group classification, only percentage of risk of distant recurrence. Apparently, the option “None” was ignored or was understood incorrectly.</td>
<td>Add “(percentage of recurrence risk only)” after the word “None”</td>
</tr>
<tr>
<td>Other parts</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Outcome category: problems related to ordering/formatting issues

Problems related to ordering or formatting issues have been identified during the interviews and are listed below, again organized for the questions, for response options, and for other parts of the questionnaire.

Table 9: Results of "problems related to ordering/formatting issues" for Objective 1, Round 1

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Response options</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>
4.2.3.2 Objective 2: To assess the questionnaire’s face validity

Outcome category: need to clarify the intent

The confusing or problematic aspects of the questionnaire sections regarding face validity are listed below relating to a) bias in the design and b) incompleteness (of the component).

Table 10: Results of "need to clarify the intent" for Objective 2, Round 1

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions (bias in design)</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Questions (incompleteness)</td>
<td>New question</td>
<td>One participant suggested there was another clinical condition that Oncotype DX&lt;sup&gt;TM&lt;/sup&gt; would be more useful. In that situation, oncologists were struggling to make treatment recommendations</td>
<td>Add an open-ended question to ask for physicians’ opinions about the case scenario for which Oncotype DX&lt;sup&gt;TM&lt;/sup&gt; would be most useful</td>
</tr>
</tbody>
</table>

*: this suggestion was implemented after two rounds of modifications.
<table>
<thead>
<tr>
<th>Response options (bias in the design)</th>
<th>2</th>
<th>Two participants believed that the categories should be designed more rationally</th>
<th>The question should be changed to an open-ended question (Similar change was made to the question 12 following scenario B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response options (incompleteness)</td>
<td>10</td>
<td>Add a response option “Not available as a routine test”</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>One physician suggested that we could introduce a ‘very high’ risk group (e.g., &gt;50% at 10 years) in the questionnaire</td>
<td>Add an answer option “Low / intermediate / high / very high risk group” in question 22</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>Add “The average rate of recurrence risk in each category at 20 years” as an answer option.</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>Add “patient ID” in each diagram</td>
<td>Yes</td>
</tr>
<tr>
<td>Scenarios</td>
<td>Patients in both scenarios were premenopausal women. If no particular reasons, these scenarios could have varied patient conditions</td>
<td>Change the patient in this scenario to a postmenopausal woman</td>
<td></td>
</tr>
<tr>
<td>Other parts (bias in the design)</td>
<td>Scenarios</td>
<td>Use “lumpectomy” instead of “mastectomy” in scenarios</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>From the completed questionnaire, we found that people might have the same answers towards the first and the second scenarios. The patient in Scenario A was not deemed to be a completely low risk case. Oncotype test result for the patient in the second scenario did not conflict with participants’ clinical judgments.</td>
<td>- The patient in scenario A should be modified as a completely low risk case</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Redesign the scenario to assure that the Oncotype DX™ test result of the patient would be more likely to conflict with clinical judgments</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.2.3.3 Objective 3: To determine if the target population finds the questionnaire interesting

Outcome category: respondents’ interest in the questionnaire

During the interviews, participants provided comments on the questionnaire:

“professional”, “easy”, “straightforward”, and “not annoying when going through”. The
participants were active in the interviewing process by kindly sharing their expertise as much as possible. None of the questions was perceived as offensive/off-putting during the interviews.

4.2.3.4 Objective 4: To ensure that the questionnaire will not place undue burden on respondents

Outcome category: overall burden

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions</td>
<td>22</td>
<td>Make this question shorter *</td>
<td>Yes</td>
</tr>
<tr>
<td>Overall questionnaire</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Facial expression</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

*: this suggestion was implemented after two rounds of modifications.

Outcome category: Limitations of the questionnaire

<table>
<thead>
<tr>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>New question suggested</td>
<td>“Does the result change in patients with familial breast cancer? (BRCA1, or BRCA2, e.g.) if so, by how much (+ contralateral risks too)?”</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>“Do you want the test to be available? / Would you ask for the test? / Would you routinely order the test for patients?”</td>
<td>No</td>
</tr>
</tbody>
</table>
4.2.4 Round 2

4.2.4.1 Objective 1: To ensure that the questionnaire is comprehensible

Outcome category: recommendations for changes to wording

**Table 13: Results of "recommendations for changes to wording" for Objective 1, Round 2**

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>To simplify this question</td>
<td>- Change “What range in percentage would you estimate” to “How would you quantify (e.g., percentage, range etc.)”.</td>
</tr>
<tr>
<td>Questions</td>
<td></td>
<td></td>
<td>- Similar change was made to the question 12 following scenario B</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>This question was neglected because the words “no change” in the sentence “If your answer is no change (Definitely would not/Probably would not/Unsure) …” was confusing</td>
<td>- Delete the words “no change”, and change the font of sentences that described the formats/skip patterns, in order to separate them from the content of a question</td>
</tr>
<tr>
<td>Response options</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Other parts</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Outcome category: need to clarify the intent

**Table 14: Results of "need to clarify the intent" for Objective 1, Round 2**

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions</td>
<td>26</td>
<td>A written comment: “correlation with local test”</td>
<td>Add a footnote on the margin saying that ER/PR/Her-2/neu status determined by this test “may differ from the status arrived at using conventional methods, e.g., immunohistochemistry (IHC) or fluorescent in situ hybridization (FISH).”</td>
</tr>
</tbody>
</table>
To state explicitly that this question refers to the situation that one ordered the test, and got the results back, then what he/she would want them to be included in the test report.

(It is not necessary to add/emphasize “if you ordered the test” in the question)

Response options: -- -- --

Other parts: -- -- --

Outcome category: problems related to ordering/formatting issues

Table 15: Results of "problems related to ordering/formatting issues"

Table for Objective 1, Round 2

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Response options</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Other parts

| 23 | The sentence “you can mark as many options as you want” in this question was not noticed by a participant | Change the font of the sentence “You can mark as many options as you want”, and separate them from the main body of the question |
|-----------------------|------------------------|------------------|-------------|

Skip pattern

To clarify how to branch among questions 24-26

Yes *

*: this objective was achieved after two rounds of modifications.

4.2.4.2 Objective 2: To assess the questionnaire’s face validity

Outcome category: need to clarify the intent
Table 16: Results of "need to clarify the intent" for Objective 2, Round 2

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions (bias in design)</td>
<td>34</td>
<td>To clarify that whether this question was asking about what visual format physicians would like themselves first to look at</td>
<td>No (Oncotype DX™ does not have a separated result report for patients, so the visual format a physician saw would be the same one as a patient saw. Besides, we study the visual format issue preliminarily, so we only expect to find a variation in physicians' preferences of visual formats.)</td>
</tr>
<tr>
<td>Questions (incompleteness)</td>
<td>28</td>
<td>A written comment: “risk categories by score” (to refer to the issue that we missed)</td>
<td>No (we already covered this issue in question 21)</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>Add a question: “how familiar are you with this test?” Or “are you familiar that this test does exist?”</td>
<td>Add a question in the “background” section: “How much did you know about Oncotype DX™ before you saw this questionnaire?”</td>
</tr>
<tr>
<td>Response options (bias in the design)</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Response options (incompleteness)</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Other parts (bias in design)</td>
<td>Scenario B</td>
<td>Two participants commented that the Recurrence Score provided did not indicate a high risk of cancer recurrence</td>
<td>Change the Recurrence Score to a score indicating an absolutely high risk of cancer recurrence *</td>
</tr>
<tr>
<td></td>
<td>Introduction to Oncotype DX™</td>
<td>Two participants suggested changing the cut-points of Oncotype DX™ back to its original ones</td>
<td>Yes, and add “Note: the schema of Oncotype DX™ in this questionnaire retains as originally published” as a footnote</td>
</tr>
<tr>
<td>Other parts (incompleteness)</td>
<td>Scenarios</td>
<td>Add surgical margins and DCIS situations in scenarios:</td>
<td>Add “No DCIS, and surgical margins are clear” in the list of pathological information in each scenario.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Surgical margin are clear, or negative margins</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Minimum or no DCIS (suggestions for a patient with low risk of recurrence)</td>
<td></td>
</tr>
</tbody>
</table>

*: this objective was achieved after two rounds of modifications.
4.2.4.3 Objective 3: To determine if the target population finds the questionnaire interesting

Outcome category: respondents’ interest in the questionnaire

No problems were uncovered in this category. Participants were active during the interviews, and they claimed that the questionnaire was easy and straightforward to complete. None of the questions was perceived as offensive/off-putting during the interviews.

4.2.4.4 Objective 4: To ensure that the questionnaire will not place undue burden on respondents

Outcome category: overall burden

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions</td>
<td>34</td>
<td>Put the five graphs on the same page (to facilitate the comparison). Then respondents would not need to turn over the page to see the last three graphs</td>
<td>Place questions 33 and 34 at the beginning of the format section, in order to facilitate the comparison of graphs in this question</td>
</tr>
<tr>
<td>Overall questionnaire</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Facial expression</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Outcome category: Limitations of the questionnaire

<table>
<thead>
<tr>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>New question suggested</td>
<td>A written comment: “consider examining how we think this information is best presented to patients.” “We may ask ‘will you be more confident to use it, if it has been validated in clinical trials?’ ‘Would you wait for the results before you even think to apply this stuff or not?’ ”</td>
<td>No (Beyond the scope of this study)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No (we had covered a similar issue in the questionnaire regarding physicians’ concerns)</td>
</tr>
</tbody>
</table>
“We may ask ‘if there is no concern about the cost, how likely would you be to use it on patients?’ ”

No (Beyond the current stage of the development of new genomic tests)

Three participants found the questionnaire was repetitive, but then they realized that we actually asked different questions. One claimed that “It (the questionnaire) maybe a little repetitive, but not too much. I think that’s ok.”

No (the repetition was unavoidable)

4.2.5 Round 3

4.2.5.1 Objective 1: To ensure that the questionnaire is comprehensible

Outcome category: recommendations for changes to wording

Table 19: Results of "recommendations for changes to wording"
for Objective 1, Round 3

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Response options</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Other parts Scenarios</td>
<td>The original style “15 of 15” may be misleading. There are three ways to describe nodes appropriately: - Node negative (0/15 nodes) - Node negative (15 nodes identified) - Node – ve (0/15 nodes)</td>
<td>Change “node-negative (15 of 15)” to “Node negative (0/15 nodes)”</td>
<td></td>
</tr>
</tbody>
</table>

Outcome category: need to clarify the intent

Table 20: Results of "need to clarify the intent" for Objective 1, Round 3

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions</td>
<td>2</td>
<td>Change “breast cancer recurrence” to “distant recurrence”</td>
<td>Yes * (Similar change was made to the question 12 following scenario B)</td>
</tr>
</tbody>
</table>
3-4  Add “systemic” before “treatment recommendation”  Yes  (Similar changes were made to the questions 13-14 following scenario B).

27  To clarify which model the question referred to  Add “statistical” in front of “model”.

| Response options | -- | -- | -- |
| Other parts | The title of the questionnaire | Add “in breast cancer” at the end | Yes |
| | The title of part III | What “test report” should be emphasized | Change the title to “Oncotype DX™ test report: what do you prefer” |
| | Scenarios | “It’s better to say: ‘right BREAST lumpectomy’ because you can have the lumpectomy all over the place.” | Add “breast” between the words “right” and “lumpectomy” |

*: this objective was achieved after two rounds of modifications.

Outcome category: problems related to ordering/formatting issues

**Table 21: Results of "problems related to ordering/formatting issues" for Objective 1, Round 3**

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions</td>
<td>The order of 1 and 2</td>
<td>Switch the order of question 1 and question 2, and add “Given your above estimate of risk” at the beginning of question 2</td>
<td>Yes</td>
</tr>
<tr>
<td>Response options</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Other parts</td>
<td>Scenarios</td>
<td>Use “stage I breast cancer” as a heading, and put “node negative” in the list of prognostic factors”</td>
<td>Separate “Stage IIA breast cancer, node-negative (15 of 15), M0” as three indicators</td>
</tr>
</tbody>
</table>
4.2.5.2 Objective 2: To assess the questionnaire’s face validity

Outcome category: need to clarify the intent

Table 22: Results of "need to clarify the intent" for Objective 2, Round 3

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Response options</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Other parts (bias in design)</td>
<td>Scenarios</td>
<td>Change “liver function test” to “liver imaging”</td>
<td>Yes</td>
</tr>
<tr>
<td>Other parts (incompleteness)</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

4.2.5.3 Objective 3: To determine if the target population finds the questionnaire interesting

Outcome category: respondents’ interest in the questionnaire

No problems were discovered in this category. None of the questions was perceived as offensive/off-putting during the interviews.

4.2.5.4 Objective 4: To ensure that the questionnaire will not place undue burden on respondents

Outcome category: overall burden

Table 23: Results of "overall burden" for Objective 4, Round 3

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions</td>
<td>30, 31</td>
<td>Delete both of these questions</td>
<td>Delete question 31</td>
</tr>
<tr>
<td>Overall questionnaire</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Facial expression</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>
Outcome category: Limitations of the questionnaire

Table 24: Results of "limitations of the questionnaire" for Objective 4, Round 3

<table>
<thead>
<tr>
<th>Question #</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>(version 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenarios</td>
<td>“A more advanced case scenario should be provided”</td>
<td>No (Beyond the current stage of the development of new genomic tests)</td>
</tr>
<tr>
<td>New question suggested</td>
<td>“There are concerns that people may think patients may not need aggressive chemotherapy, but they do need chemotherapy. It’s also a change in treatment management.” (comment on whether we need to address the issue about aggressive/ less aggressive chemotherapy)</td>
<td>No (Beyond the scope of this study)</td>
</tr>
<tr>
<td></td>
<td>“If the test has been fully validated, in what situation would it be helpful?”</td>
<td>No (Beyond the current stage of the development of new genomic tests)</td>
</tr>
</tbody>
</table>

4.2.6 Round 4

4.2.6.1 Objective 1: To ensure that the questionnaire is comprehensible

Outcome category: recommendations for changes to wording

There were no recommendations for changes to inappropriate/confusing wording in this round of interviews.
Outcome category: need to clarify the intent

Table 25: Results of "need to clarify the intent" for Objective 1, Round 4

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions</td>
<td>2</td>
<td>To clarify whether the adjuvant treatment includes radiation therapy or not</td>
<td>No (Radiation therapy does not influence the risk of distant recurrence)</td>
</tr>
<tr>
<td>Questions</td>
<td>26, 27</td>
<td>To explain what RNA expression is, and what RT-PCR is, in case some oncologists are not aware of these terms</td>
<td>(Breast cancer oncologists should know well about basic biological terms, e.g., RNA expression, through pathology and the information about other existing molecular prognostic factors. Additionally, a brief introduction to bio-tech terms has been provided in the Appendix.)</td>
</tr>
<tr>
<td>Response options</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Other parts</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Outcome category: problems related to ordering/formatting issues

No issues were identified during this round of interviews.

4.2.6.2 Objective 2: To assess the questionnaire’s face validity

Outcome category: need to clarify the intent

Table 26: Results of "need to clarify the intent" for Objective 2, Round 4

<table>
<thead>
<tr>
<th>Questionnaire section</th>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions (bias in design)</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Questions (incompleteness)</td>
<td>New question suggested</td>
<td>A comment suggesting providing an additional question: “Have you ever used this test before?”</td>
<td>No (Not a necessary question)</td>
</tr>
</tbody>
</table>
4.2.6.3 Objective 3: To determine if the target population finds the questionnaire interesting

**Outcome category: respondents’ interest in the questionnaire**

Objective 3 was accomplished because there were no new/additional problems identified during this round of interviews.

4.2.6.4 Objective 4: To ensure that the questionnaire will not place undue burden on respondents

**Outcome category: overall burden**

<table>
<thead>
<tr>
<th>Table 27: Results of &quot;overall burden&quot; for Objective 4, Round 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Questionnaire section</strong></td>
</tr>
<tr>
<td>Questions</td>
</tr>
<tr>
<td>Questions</td>
</tr>
<tr>
<td>Overall questionnaire</td>
</tr>
</tbody>
</table>
Outcome category: Limitations of the questionnaire

Table 28: Results of "limitations of the questionnaire" for Objective 4, Round 4

<table>
<thead>
<tr>
<th>Question # (version 1)</th>
<th>Change suggested</th>
<th>Change made</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A (target population)</td>
<td>We could include both experimental oncologists and (medical oncology) residents</td>
<td>No</td>
</tr>
</tbody>
</table>
|                        |                                                                                  | (Experimental oncologists are not the members of the target population, because they do not treat patients. We have already included residents) |}

4.2.7 Analysis of interview results

Below is a summary of the number of problems identified per round for each objective:

Table 29: The summary of the number of problems identified for each objective

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Round 1</th>
<th>Round 2</th>
<th>Round 3</th>
<th>Round 4</th>
<th>Total number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. To ensure that the questionnaire is comprehensible</td>
<td>12</td>
<td>5</td>
<td>9</td>
<td>✓</td>
<td>26</td>
</tr>
<tr>
<td>2. To assess if the questionnaire has face validity</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td>✓</td>
<td>14</td>
</tr>
<tr>
<td>3. To determine if the target population finds the questionnaire interesting</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>0</td>
</tr>
<tr>
<td>4. To ensure that the questionnaire will not place undue burden on respondents</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>✓</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td>10</td>
<td>11</td>
<td>0</td>
<td>43</td>
</tr>
</tbody>
</table>

“✓”: indicates that no new problems were assigned to the specific objective in the entire round of interviews.

From this table, it can be seen that problems relating to objective 1 were the most frequent, indicating that the most effort that we had to make was to ensure that the questionnaire would be comprehensible.
Below is a summary of types of problems in the design of our questionnaire. The frequency of types of problems is presented by the problem codes assigned per each round of interviews.

**Table 30: The summary of the problem codes assigned per interview round**

<table>
<thead>
<tr>
<th>Problem code category</th>
<th>Round 1</th>
<th>Round 2</th>
<th>Round 3</th>
<th>Round 4</th>
<th>Total number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehension / communication of intent</td>
<td>13</td>
<td>5</td>
<td>9</td>
<td>✓</td>
<td>27</td>
</tr>
<tr>
<td>Bias/sensitivity</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>0</td>
</tr>
<tr>
<td>Response category</td>
<td>5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>5</td>
</tr>
<tr>
<td>Logical issues</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>✓</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td>10</td>
<td>11</td>
<td>0</td>
<td>43</td>
</tr>
</tbody>
</table>

“✓”: indicates that no new problems were identified in the specific category in the entire round of interviews.

From the table above, it can be seen across the four rounds of interviews that half of the problems were identified through the first round of interviews, and the most frequent problems in the whole process of the questionnaire design were in the comprehension/communication of intent category. Logical issue problems were the next most frequent, followed by the design of response categories.

The above tables also demonstrate that there were no new issues/problems raised in the fourth round. The result from this round met our stopping criterion, so the interviewing process ended.
Chapter 5

Discussion

The discussion of the thesis consists of five sections. Section 5.1 summarizes the results and associated insights of the study. Section 5.2 covers the strengths and limitations of the study. Section 5.3 looks at the next steps that may follow this study, section 5.4 concludes with the contribution of this study, and section 5.5 draws the conclusions of the study.

5.1 Summary of Results and associated insights

All of the objectives of the study have been achieved. The stopping criterion for the cognitive interviewing has been met which suggests that the proposed questionnaire is comprehensible and has face validity. In addition, we have gained some insight into the target population’s concerns about a particular new genomic test.

The results of the study suggest that, through four rounds of interviews, the final version of the questionnaire is much more precise and comprehensible than the first version. The title was made more specific, confusing questions and response options have been clarified, important issues that were missed have now been integrated into the questionnaire, all of which has increased the face validity of the questionnaire.

Beyond that, some problems relating to underlying questionnaire logic, which could potentially affect the validity of responses, were corrected. For example, it became evident that
the original response options provided for physicians’ estimates of cancer recurrence risk did not align with how the physicians naturally estimate the recurrence risk; thus, how respondents adapted their natural responses to the provided options varied as a result of mapping issues. Another example of a logic issue was related to the skip pattern. It became apparent that some skip patterns were not optimally designed because some physicians skipped some questions by mistake. The most significant logic concern was revealed when we found that some physicians’ risk estimates for the two scenarios were both concordant with the Onco
type DX™ results in the original questionnaire. That is, for some physicians, the second scenario failed to present a discordant situation. After revisions, the two scenarios produced the cases that we intended: one concordant with the Onco
type DX™ result and one discordant but realistic. In the fourth round of the interviews, a participant even commented on the scenario B that he once had a patient who had the same controversial situation as described in scenario B. His comment provided evidence confirming the realistic design of the scenarios.

The semi-structured interviews not only allowed me to probe the questions in the questionnaire, but also provided me with an opportunity to discover some specific concerns of the target population. I was interested to find, for example, that currently, the biggest concern of our oncology participants is the validity of the new genomic tests, including Onco
type DX™. Until a new genomic test is validated completely, physicians are likely to be conservative in changing their treatment recommendations only based on the result of the test, even though these physicians are willing to offer patients the test.
Although we did not intend to analyze the data from the participants of the interviews beyond ensuring that they understood the intent of the questions and that their responses appeared meaningful, the responses to particular questions provided some insight regarding the degree of physicians’ desire for the test report. The data suggest that, in general, physicians will require a large amount of information in the test report, more than the information covered in the current report. For example, the RNA expression results of some well-known genes (ER gene, PR gene, Her-2/neu, Ki67, etc.) were strongly desired by many participants in the study; however, these results are not explicitly described in the current report of Oncotype DX™. In addition, it appears that physicians vary in their preferred visual formats for quantitative information included in the test report: some physicians preferred the original graph in Oncotype DX™ test report, while others preferred simpler or still others more complicated graphs.

5.2 Strengths and limitations of the study

Clinical implementation of new genomic tests is a novel research area. There are some particular strengths of this study. For example, including participants from different backgrounds were very helpful in identifying fundamental issues in the questionnaire design. The selection of Oncotype DX™ also brought great benefits. For example, all medical oncologists and many radiation oncologists in interviews were aware of Oncotype DX™ before this study. Some of them were involved with the TAILORx trial that utilized Oncotype DX™ results as an indication for chemotherapy. Because they were familiar with Oncotype DX™, some of the physicians
suggested that the schema of Oncotype DX™ results classification in the questionnaire should retain as originally published. The schema used in the TAILORx trial was not appropriate because it has not been widely adopted clinically. So the change of the schema of Oncotype DX™ was made correspondingly. Finally, the results of the cognitive interviews suggest that the final self-administered questionnaire will be able to achieve its objectives in the future large scale survey.

This study also has some limitations. For example, cognitive interviewing has to count on subjective judgments, interpretation and corrections. The modifications of the questionnaire were based both on dominant trends across interviews (problems that seem to emerge repeatedly) and on "discoveries" (problems that occur in only a single interview) (65). It is possible that a problem identified during the interviews might occur more frequently in real life, or might not actually affect the resulting data adversely. To minimize such concern, all problems were vetted with my supervisors before being addressed; thus, only those considered substantive were addressed and the corrections were subsequently tested.

Another limitation is the small number of interviews conducted within a testing round. We could only recruit three interviewees in each round due to the limited number of oncologists in the area (the literature typically recommends 5-10 interviews in one round (63)). If more oncologists had been available to be recruited, we would have had stronger confidence that we have collected comprehensive insight into the performance of the questionnaire. Despite these
concerns, our sample size proved sufficient for meeting an a priori stopping rule in cognitive interviewing.

Several aspects of the questionnaire design should also be considered when implementing the results of the study because these aspects may restrict the generalizability of the eventual large scale survey results to some extent. The following paragraphs discuss these potential limitations focused on three aspects of our study design: basing the questionnaire on scenarios and Oncotype DX™, the participants of the study and their behavior, and the utilization of phone interviews in the cognitive interviews.

Related to questionnaire construction, two potential limitations may arise from the particular scenarios included. First, the two scenarios may not be fully representative of concordant and discordant situations. In concordant examples, a “low (clinically) vs. low (Oncotype DX™)” situation and a “high (clinically) vs. high (Oncotype DX™)” situation may not be weighed equally to a physician. In discordant situations, a “low (clinically) vs. high (Oncotype DX™)” scenario and a “high (clinically) vs. low (Oncotype DX™)” scenario may also have different impact on a physician’s decision. So it is possible that, in the large scale survey, physicians’ attitudes towards concordant and discordant situations would only be addressed to some extent. Second, only two case scenarios were provided. Additional scenarios were excluded out of concern for heavy response burden that would be placed on respondents and due to the inadequate research evidence of Oncotype DX™. For example, during the interviews, some participants deemed that Oncotype DX™ would be most useful when coping with a cancer
patient at an intermediate risk of distant recurrence. This issue could not be addressed both because physicians vary in their definition of what is intermediate recurrence risk, and because the clinical value of the intermediate risk category of Oncotype DX™ is still unclear. As a tradeoff, an open-ended question that asked for a case scenario in which Oncotype DX™ would be most useful was added.

Potential limitations in the questionnaire may also relate to our choice of Oncotype DX™ as the test prototype. First, Oncotype DX™ is only one of a group of diverse tests; thus, there are many features of other tests that cannot be represented by Oncotype DX™. Second, how the test results of Oncotype DX™ are presented will have an impact on physicians’ attitudes and judgments towards this test. Possibly, these attitudes and judgments may not generalize to other new genomic tests. Third, the selection of Oncotype DX™ has limited what can be asked in the questionnaire: for example, questions regarding comparable tests are not applicable. During the interviews, a participant suggested that we could ask for physicians’ preferences of genomic tests if there was another test comparable in quality but less expensive. However, all other tests are different from Oncotype DX™ either in quality or in the stage of development. Without tests of equal or better quality, we cannot investigate the complexity of physicians’ attitudes when they consider ordering both tests, one test, or neither test.

Potential limitations of the large scale survey also derive from the following two dimensions: the particular participants of the study, and their behaviour in the interviews. Our particular participants were Kingston oncologists who work together and that may present a
limitation. We cannot guarantee that their suggestions would be endorsed by the majority of the
target population. For example, Kingston oncologists deemed that the critical values to
distinguish low, intermediate and high risk group from one another should be retained as
originally published. However, the schema of Oncotype DX\textsuperscript{TM} results classification in the
TAILORx trial is clearly preferred by the oncologists designing the trial and might have been
preferred more generally if more interviews could have been conducted. The modification of
Scenario B is another example. As a result of the interviews, the Oncotype DX\textsuperscript{TM} result for the
patient in scenario B was shifted to a much higher but reasonable Recurrence Score because the
original score was deemed by Kingston oncologists as presenting an intermediate risk. However,
we could not assume that the majority of Canadian medical oncologists would have the same
judgment. It might still be possible that the higher score would not be considered high risk by
some oncologists. It is worthwhile to be aware of these potential limitations; they are not likely to
cause serious problems because the relevant modifications survived through several rounds of
testing afterwards. In addition, participants outside of Kingston, although few, had no problems
regarding these modified items. As long as the stopping criterion for the cognitive interviews was
met, it is not likely that there will be serious problems.

Another limitation was that during the interviews, some respondents provided comments
without “formally” filling in the questionnaire, as we had instructed participants that the purpose
of the interview was to assess the questionnaire rather than to collect the data. It is possible that
the comments might have been different if these respondents had systematically filled in the
questionnaire. Specifically, some problems identified when respondents just took a glance at the questionnaire might no longer be issues after respondents filled in the questionnaire, whereas some unexpected problems might emerge during the time when respondents filled in the questionnaire. The skip pattern was an example of this issue that showed up during the interviews. People who did not fill in the questionnaire usually sensed that the skip pattern looked complicated; however, people who finished the questionnaire deemed that the skip pattern was helpful and easy to follow. Unavoidably, the discrepancy between respondents with/without filling in the questionnaire might have an impact on the final product of the questionnaire. Likewise, the differences in the extent to which interviewees attached importance to our study had an effect on the improvement of the questionnaire.

Last but not least, the third aspect of potential limitations came from the use of phone interviews in the cognitive interviews. We are conscious that the relatively good interviewing results (no new problems/issues identified) are related to at least two factors. The first one is the progress we made to the development of the questionnaire: major problems have already been identified, and no new issues or mistakes can be raised by participants. However, we cannot rule out the other factor—the effect of the mixed interviewing methods. Face-to-face interviews are usually more intensive interviews. They often take a longer time than phone interviews, and provide more non-verbal communication, e.g., behavior observation, between the interviewer and the respondent. Thus, phone interviews, depending more on verbal communication, might identify fewer questions/issues. As a result, there is a possibility in our study that unexpected
issues could be raised by respondents if the last round of interviews were conducted as face-to-face interviews.

5.3 Next steps

5.3.1 Field pilot test

Usually, a pilot test under realistic field conditions is recommended after cognitive interviews, due to the limitations of cognitive interviewing as a way of evaluating questionnaires (106). However, it is also suggested in the literature that the value of a field pilot test cannot be overstated (107). In our case, by balancing the advantages and disadvantages of a pilot test, it is likely that a field pilot test will provide little additional insight into the questionnaire development and the administration process of the eventual large scale survey. The reasons are presented below.

The results of the thesis suggest that adequate pretesting of the questionnaire has been done. No fixed number of interviews was predefined. Rounds of interviews kept being conducted until no problems were identified, indicating that a) the sampled individuals were satisfied with the content and the structure of the questionnaire finally, and b) the development and evaluation of the questionnaire were completed because the objectives of the study were met.

Therefore, the intention for a field pilot test is to complement the shortcoming of cognitive interviewing. The potential objectives of a pilot test include a) questionnaire
development: trying to identify new issues or problems, and b) survey administration: gaining some insight into how the target population will react to the mailed questionnaire.

Cognitive interviewing may underestimate the severity of some problems. A field pilot test can help identify potential problems on whether or not the questionnaire is understandable by soliciting respondents’ written comments. These comments can be written at the margins of particular questions, and can also be marked on an additional self-administered questionnaire which collects questions regarding the design of the proposed questionnaire. Moreover, missing data regarding specific questions can also be estimated. These questions may, in turn, require revision or deletion.

Regarding survey administration, a field pilot test may help identify which administration procedures are most likely to be effective. In addition, an expected response rate can be estimated roughly. This response rate can help determine whether the cover letter should be modified, whether multiple modes of survey administration should be applied, and whether cash incentive should be included in the succeeding large scale survey.

Many factors can limit the potential of a field pilot test to provide the above information. Related to questionnaire design, determining the causes of problems would be challenging. Respondents’ handwriting may be difficult to read. Respondent debriefings formulated as an enclosed self-administered questionnaire places an additional burden on participants, which may result in an increase of non-response rate to either the debriefing questionnaire or the proposed questionnaire, or both. The level of detail and the accuracy of responses in debriefings may be
difficult to control as well. Although questions involving missing data can be identified, we can never be confident of the reasons for non-response. In addition, even if some substantive problems can be identified, a field pilot test per se cannot ensure that the revision of the problematic questions will do better in the large scale survey. In the administration of a pilot test, resources are always a concern. Time and cost may be restrained. Finally, the target population is a small, specialized population. The more who participate in developmental studies, the fewer they will be available to participate in the large scale survey.

5.3.2 Issues to consider when implementing the eventual large scale survey

Although this thesis excludes the administration of the large scale cross-sectional survey, issues around its implementation are presented below.

5.3.2.1 Sample size

At the survey planning stage, the relationships between variables and sample size calculation are usually determined. To assist this process for the proposed questionnaire, the table below presents variables grouped by the potential outcomes of interest and their corresponding potential analyses. They represent the fundamental analyses, although it is possible that future investigators could have additional research questions they would like to explore.
Table 31: Potential outcomes of interest and the corresponding variables

<table>
<thead>
<tr>
<th>The potential outcome of interest</th>
<th>Dimension</th>
<th>Question # of the variable (final version)</th>
<th>Potential Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physicians’ attitudes towards Oncotype DX™</td>
<td>General willingness to use Oncotype DX™</td>
<td>9, 19</td>
<td>Frequency table</td>
</tr>
<tr>
<td></td>
<td>Concerns</td>
<td>10, 20</td>
<td>Frequency table</td>
</tr>
<tr>
<td></td>
<td>Usefulness of Oncotype DX™ information</td>
<td>35-37</td>
<td>Frequency table</td>
</tr>
<tr>
<td></td>
<td>Different attitudes for distinct scenarios *</td>
<td>9, 19</td>
<td>Hypothesis test</td>
</tr>
<tr>
<td></td>
<td>To compare physicians’ attitudes by their familiarity with Oncotype DX™</td>
<td>9/19, 38</td>
<td>Parameter estimation</td>
</tr>
<tr>
<td></td>
<td>To compare physicians’ attitudes towards Oncotype DX™ by years of their clinical experience</td>
<td>9/19, 39</td>
<td>Parameter estimation</td>
</tr>
<tr>
<td></td>
<td>Risk classification *</td>
<td>22</td>
<td>Frequency table</td>
</tr>
<tr>
<td></td>
<td>Average recurrence rate</td>
<td>23, 24</td>
<td>Frequency table</td>
</tr>
<tr>
<td></td>
<td>Individual recurrence rate</td>
<td>25, 26</td>
<td>Frequency table</td>
</tr>
<tr>
<td></td>
<td>Specific test results of Oncotype DX™ (may contain subcategories)</td>
<td>27</td>
<td>Frequency table</td>
</tr>
<tr>
<td></td>
<td>Information about Oncotype DX™ itself (may contain subcategories)</td>
<td>28</td>
<td>Frequency table</td>
</tr>
<tr>
<td></td>
<td>Presentation/formats of risk information (may contain subcategories)</td>
<td>30 (model) 31 (number vs. word) 32 (95% CI) 33 (recurrence vs. survival rate) 34 (visual formats)</td>
<td>Frequency table</td>
</tr>
<tr>
<td>The impact of Oncotype DX™ on physicians’ treatment recommendation</td>
<td>Certainty of the risk estimate</td>
<td>5, 15</td>
<td>Frequency table</td>
</tr>
<tr>
<td></td>
<td>Usefulness in treatment recommendation</td>
<td>6, 16</td>
<td>Frequency table</td>
</tr>
<tr>
<td></td>
<td>Changes in treatment recommendation</td>
<td>7, 8, 17, 18</td>
<td>Frequency table</td>
</tr>
<tr>
<td></td>
<td>Different impacts on distinct scenarios</td>
<td>5-8, 15-18</td>
<td>Hypothesis test</td>
</tr>
</tbody>
</table>

*: will be used in the following paragraphs as examples for discussion of sample size determination.

The following discussion regarding sample size calculation assumes that our primary outcome would be physicians’ attitudes towards Oncotype DX™ because it was the outcome that
motivated the questionnaire development. In particular, I would like to know whether there would be a shift in physicians’ attitudes from scenario A to scenario B. If a shift did occur, it would demonstrate that physicians’ attitudes towards Onco\textit{type} DX\textsuperscript{TM} varied according to particular situations, providing evidence of construct validity of the questionnaire. In addition, between the two scenarios provided, an observed shift in attitudes would identify the scenario under which Onco\textit{type} DX\textsuperscript{TM} would be more useful, which in turn would facilitate the optimization of the clinical use of Onco\textit{type} DX\textsuperscript{TM}.

The sample size calculation would be driven by this primary research question. However, the calculation could not derive directly from the abstract research question or from a summary of particular questions which measure physicians’ attitudes. Instead, we would compare the answer to question 9 with the answer to question 19 to shed some light on the magnitude of attitude shift. These two questions were selected as they directly measure physicians’ willingness of using Onco\textit{type} DX\textsuperscript{TM}. To simplify the sample size calculation, we would collapse 5 response options into “positive” (“definitely would”/“probably would”) and “not positive” (“unsure or neutral”/“probably would not”/“definitely would not”). So the answers to these questions would be dichotomous. One could calculate the proportion of physicians’ positive answer to question 9, and to question 19, however the difference in the two proportions would not be sufficient for studying the shift in physicians’ attitudes, because it would result in serious loss of information. For example, it is possible that there could be no difference in the proportions of positive answers to the two questions, but actually physicians switch their answers between positive and not
positive in either direction with equal proportions. Therefore, the proportion of discordant responses \((p)\) is a better measure of physicians’ shifting attitudes. That is, the proportion of discordant responses \((p)\) would measure the proportion of the surveyed individuals who had discordant answers between question 9 and question 19.

For the purpose of estimating a sample size requirement, we would be interested to determine if the discordant response proportion \((p)\) would be greater than 10% because we would not consider a discordant response proportion at 10% or less to be clinically significant. This level of 10% was reached by consensus discussion among the thesis committee members. In our case, we deemed that a difference greater than 10% in discordant responses would provide evidence that, at least to some clinicians, the context in which Oncotype DX™ was used would be important. In other words, under some contexts, Oncotype DX™ should be clinically useful by adding a new piece of information in clinical decision-making process. Even though we preferred a difference at 10% to be the threshold for clinical relevance, future investigators would be able to choose a different threshold based on their outcomes of interest.

The sample size calculation for detecting a difference in proportion greater than 10% would be based on a one-sample test for a binomial proportion with the normal theory method and a one-sided alternative (108):

**Ho:** The discordant response proportion \((p)\) is equal to or smaller than 10% \((p = p_0 = .10)\).

**H_1:** The discordant response proportion \((p)\) is greater than 10% \((p > .10)\).
Traditionally, a one-sided hypothesis test at 95% confidence level ($\alpha=.05, Z_{1-\alpha}=1.645$) is accepted as a satisfactory certainty. In our case, in order to find a discordant response proportion greater than 10% significantly, $Z$ score in equation 5.1 should be greater than 1.645. However, the anticipated sample size could not be generated directly from equation 5.1, because the power of the selected one-sample test should also be considered. Conventionally, 80% power ($1-\beta=.80, Z_{1-\beta}=.84$) is widely used in clinical studies to correctly reject the null hypothesis ($H_0$) when a specific alternative hypothesis ($H_a$) is true.

Therefore, the sample size needed to conduct a one-sided hypothesis test would be (108): 

$$n = \frac{p_0(1-p_0)\left(Z_{1-\alpha} + Z_{1-\beta} \sqrt{\frac{p(1-p)}{p_0(1-p_0)}}\right)^2}{(p-p_0)^2}$$

(Equation 5.2)

Where:

- $n$ - The sample size
- $p_0 = .10$
- $Z_{1-\alpha}$ - Z score for a certain confidence level.
- $Z_{1-\beta}$ - Z score for a certain power level.

Since it is unknown what the magnitude of the discordant response proportion ($p$) would be, several sample size calculations are provided in the following table for a range of examples of “$p$”s using equation 5.2.
Table 32: Sample size determination

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted discordant response proportion (p)</td>
<td>.30</td>
</tr>
<tr>
<td>Minimum sample size needed n</td>
<td>20</td>
</tr>
</tbody>
</table>

From the above table, we would be 95% certain that the calculated sample size would be sufficient to have 80% power to ensure that we could detect a true shift in the target population’s attitudes towards the two scenarios of greater than 10% for the corresponding predicted proportion. Specifically, a sample size of 69 would be required if a discordant response proportion at 20% or higher would be anticipated. To be more conservative, if the predicted proportion would fall in the range of 15% to 20%, a sample size of 252 would be required. Likewise, if a proportion less than 15% would be expected, a much larger sample size would be needed. However, to recruit such a large number of participants for the detection of a relatively small difference would not be necessary in our case.

The impact of the sample size on the precision of the proportion estimate (two-sided) could also be described. The desired precision $d$, which measures the margin of error, could be expressed in the following equation (66):

$$d = Z_{1-\alpha/2} \times SE = Z_{1-\alpha/2} \sqrt{p(1-p)/n}$$

(Equation 5.2)

Conservative estimates of the desired precision would be situations with low precision, that is, to choose a wide margin of error. $Z_{1-\alpha/2}$ is a fixed number in the equation 5.2. For 95% confidence interval level, $\alpha = .05$, $Z_{1-\alpha/2} = 1.96$. $p(1-p)$ attains its maximum at $p = .50$. That is, for the binomial distribution, when $p = .50$, the standard deviation of the population is the largest, and the confidence interval around $.50 (p = .50)$ is the widest. The confidence interval becomes
tighter and the standard deviation becomes smaller as \( p \) moves away from .50 in either direction.

In other words, a proportion other than .50 would result in a higher precision. The calculated widest possible confidence intervals are provided in the following table for each sample size computed above:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size determined previously</td>
<td>20</td>
</tr>
<tr>
<td>The possible biggest standard error (( p = .50 ))</td>
<td>.11 .06</td>
</tr>
<tr>
<td>The widest possible 95% Confidence interval for predicted discordant response proportion (( p ))</td>
<td>( p \pm .22 ) ( p \pm .12 ) ( p \pm .06 )</td>
</tr>
</tbody>
</table>

The precisions estimate of a proportion can also be applied to other potential survey variables. For example, the proportion of the target population who would like to know the risk classification information from Oncoty”pe DX™ test report.

As a researcher studying the implementation of a novel test, I would choose a sample size around 69. From table 32 and table 33, such a sample size would have 80% power to detect a difference in discordant response proportion greater than 10% at the 95% confidence level, as well as providing an acceptable margin of error around 10%. The margin of error at 10% should be clinically reasonable, because it ensured that the corresponding confidence interval around a proportion would not be too wide or unnecessarily narrow. A margin of error larger than .10 would result in a wider confidence interval, which might not be precise enough to inform information about the shift in physicians’ attitudes, or the usefulness of Oncoty”pe DX™ test results. A margin of error smaller than .05 would lead to a much narrower confidence interval;
however it would less likely to be clinically significant. In addition, the benefit of having a small
margin of error might not outweigh the dramatic increase of required sample size (as shown in
table 32). However, future investigators could choose a different desired precision level based on
their outcomes of interest.

Additionally, the computed sample size should be adjusted for the number of the target
population and an estimated sample effect, if the data collection method would be other than
simple random sampling. Moreover, the final sample size should be adjusted for the expected
response rate. Conservatively, we would assume the response rate may reach 60-70%, after
consulting other surveys targeting medical oncologists (109-111). As a result, future investigators
would have to consider a sample size increment to increase their likelihood of getting the targeted
number of responses.

5.3.2.2 Other issues

To conduct the large scale survey, the target population would be Canadian medical
oncologists who treat breast cancer. However, the surveyed sample might expand to breast cancer
medical oncologists from the USA, Europe or/and Australia, allowing a more general survey of
oncologists’ attitudes or concerns about Oncotype DX™, and also allowing comparisons between
geographically-defined groups.

Order effects are often evident in surveys (112, 113). In order to minimize possible order
effects, several versions of the questionnaire would need to be created. The changes should be
made only to the order of questions, and the logical flow should not be affected. Then the
responses across the different orders can be averaged to balance possible order effects. Potential modifications of the question order include: switching the locations of two scenarios, question order in the “test results” section, in the “format” section, and in the “usefulness” sections, respectively.

Future investigators may consider assessing test-retest reliability. The analysis of the reliability may help determine whether the target population could understand issues and questions the same way. This type of reliability is very important in determining the reliability of responses in the large scale survey.

In the data collection stage, probability sampling is preferred (59). Simple random sampling, systematic random sampling or other complex sampling methods could be employed.

During survey administration, Dillman’s guidelines for tailored design are recommended (61, 88). Multiple modes of questionnaire administration (e.g., paper-and-pencil and web-based self-administered questionnaires) may be considered in order to maximize the response rate.

5.4 Contribution of the study

Given the great potential of new genomic tests, this study is an initial attempt to design a questionnaire to address physicians’ attitudes, concerns, and refine the test report as friendly, as conveniently, and as efficiently as possible. The study is devoted to helping optimize the clinical application of new genomic tests and their results, and tries to achieve benefits for all, including
bio-tech laboratories, physicians and patients. Although these new genomic tests have not been fully validated, and some of them are merely under development, it is expected that, in the near future, the new genomic tests will be mature and show their full value in clinical medicine. The combination of new genomic tests and conventional prognostic tests will provide more accurate prognosis for cancer patients, reduce uncertainty in treatment decision-making, improve cancer management by tailoring treatment to individuals, and increase the understanding of cancer etiology. Until then, more and more studies similar to this thesis may emerge, and target the implementation interface.

I hope that, through this study and other similar studies, physicians’ perspectives on new genomic tests will be addressed completely and effectively, and that conveying test information to physicians will be facilitated. I also hope that such studies can have an influence on future research of new genomic tests, with the purpose of assessing the quality of the implementation of such tests.

Although the thesis focuses on new genomic test, the research approach may be adapted to other novel tests. It is likely that in the implementation of future prognostic tests, or/and maybe other medical tests as well, studies similar to this thesis may be required to gain some insight into physicians’ perceptions of these tests. Then to design a test- or/and scenario-based questionnaire, followed by cognitive interviewing, may be a decent choice.
5.5 Conclusions

In conclusion, this thesis has successfully developed a questionnaire to address the objectives it intended to do. The stopping criterion of cognitive interviews was met, indicating that the questionnaire should be comprehensive, have face validity, be interesting, and would place a reasonable burden on respondents. These findings suggest that the questionnaire should be ready for the field. A large scale survey should be conducted to fully understand the implementation of new genomic tests from physicians’ perspectives.
References


43. Rimer BK. The cancer risk communication meeting in perspective. Journal of the National Cancer Institute Monographs 1999;25.


Appendix A
The panel of 21 genes and the Recurrence-Score algorithm

“The recurrence score on a scale from 0 to 100 is derived from the reference-normalized expression measurements in four steps. First, expression for each gene is normalized relative to the expression of the five reference genes (ACTB [the gene encoding β-actin], GAPDH, GUS, RPLPO, and TFRC). Reference-normalized expression measurements range from 0 to 15, with a 1-unit increase reflecting approximately a doubling of RNA. Genes are grouped on the basis of function, correlated expression, or both. Second, the GRB7, ER, proliferation, and invasion group scores are calculated from individual gene-expression measurements, as follows: GRB7 group score = 0.9 x GRB7 + 0.1 x HER2 (if the result is less than 8, then the GRB7 group score is considered 8); ER group score = (0.8 x ER + 1.2 x PGR + BCL2 + SCUBE2) ÷ 4; proliferation
group score = (Survivin + KI67 + MYBL2 + CCNB1 [the gene encoding cyclin B1] + STK15) ÷ 5
(if the result is less than 6.5, then the proliferation group score is considered 6.5); and invasion group score = (CTSL2 [the gene encoding cathepsin L2] + MMP11 [the gene encoding stromolysin 3]) ÷ 2. The unscaled recurrence score (RSU) is calculated with the use of coefficients that are predefined on the basis of regression analysis of gene expression and recurrence in the three training studies: RSU = + 0.47 x GRB7 group score – 0.34 x ER group score + 1.04 x proliferation group score + 0.10 x invasion group score + 0.05 x CD68 – 0.08 x GSTM1 – 0.07 x BAG1. A plus sign indicates that increased expression is associated with an increased risk of recurrence, and a minus sign indicates that increased expression is associated with a decreased risk of recurrence. Fourth, the recurrence score (RS) is rescaled from the unscaled recurrence score, as follows: RS=0 if RSU<0; RS=20x(RSU–6.7) if 0≤RSU≤100; and RS=100 if RSU>100.” (67)
Appendix B

Do’s and Don’t’s of questionnaire design

<table>
<thead>
<tr>
<th>Do</th>
<th>Don’t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assign numbers to each question</td>
<td>Don’t leave off the question number</td>
</tr>
<tr>
<td>User letters to indicate subparts of a question when it has more than one part</td>
<td>Don’t leave off the letter for subparts of a question</td>
</tr>
<tr>
<td>Use a vertical response format for closed-end responses</td>
<td>Don’t list closed-end responses horizontally</td>
</tr>
<tr>
<td>Use numerical codes for closed-end responses</td>
<td>Don’t use alphabetical codes or blank lines to place X or check on, for closed-end responses</td>
</tr>
<tr>
<td>Use consistent numerical codes and formats</td>
<td>Don’t use different numerical codes and formats for comparable responses to different questions</td>
</tr>
<tr>
<td>Align response codes</td>
<td>Don’t vary alignment of response codes on a page</td>
</tr>
<tr>
<td>Provide clear instructions for open-ended items</td>
<td>Don’t just leave a space with no instructions for the answer</td>
</tr>
<tr>
<td>Provide clear special instructions</td>
<td>Don’t just leave a space with no instructions for the answer</td>
</tr>
<tr>
<td>Provide clear skip instructions</td>
<td>Don’t have instructions about how to answer questions in same typeface and format as question</td>
</tr>
<tr>
<td>Phrase full and complete questions</td>
<td>Don’t simply use words or headings to elicit information from respondents</td>
</tr>
<tr>
<td>Use a forced-choice format for a list</td>
<td>Don’t ask respondent to indicate “all that apply” if he or she could indicate more than one response</td>
</tr>
<tr>
<td>Use a column format for a series with the same response categories</td>
<td>Don’t repeat a string of questions with the same response categories</td>
</tr>
<tr>
<td>Use a column format for a series with comparable skip patterns</td>
<td>Don’t fail to clearly link a series of questions to subsequent dependent items</td>
</tr>
<tr>
<td>Put all parts of a question on the same page</td>
<td>Don’t split a question between pages, particularly when skip instructions are part of the question</td>
</tr>
<tr>
<td>Allow plenty of space on the questionnaire</td>
<td>Don’t crowd the questions and space for recording the answers</td>
</tr>
<tr>
<td>Carefully consider the appearance of the questionnaire</td>
<td>Don’t just start the questions on page 1 without introducing the study, identifying the sponsoring organization, and so on</td>
</tr>
<tr>
<td>End the questionnaire with a thank you</td>
<td>Don’t forget to thank the respondent</td>
</tr>
<tr>
<td>Consider how the data will be processed</td>
<td>Don’t fail to anticipate how you will code and process the data as you design the questionnaire</td>
</tr>
</tbody>
</table>
Appendix C

The invitation email to local oncologists

Subject line: Dr. (interviewee name), We need your feedback on a draft questionnaire

Main body of the email:

Dear Dr. (interviewee name),

I am a master’s student studying what oncologists think about new genomic tests estimating breast cancer recurrence risk (such as Oncotype DX or MammaPrint) and how oncologists wish them to be implemented. Eventually, we will use a mailed survey to obtain those opinions.

This email is a follow-up to the presentation that I made in the Breast Cancer Site Group in July about my study. The group suggested that I email each physician individually.

I would like to interview you to obtain your feedback on my draft questionnaire regarding:

1. what questions should be included in the survey;
2. how these questions should be asked.

The interview should last no more than one hour, and I am happy to meet you at your earliest convenience. It would be greatly appreciated if you could provide me with the date and time convenient for the interview. Thank you!

If you would like more information about the study, please contact me, Jing Jin, at (613)484-8268, or through email at 5jj14@queensu.ca.

Sincerely,

Jing Jin

Supervised by
Dr. Michael Brundage
Dr. Julia Brettschneider
Dr. Deb Feldman-Stewart
Cancer Research Institute, Queen's University.
Appendix D

The sequence of cognitive interviewing activities

**Preparation for Interviewing**

- Expert Appraisal:
  - Review questionnaire, and make suggestions for modifications prior to testing

- Develop basic probes to use in first round of interviewing

**Cognitive Interviewing Round**

- In-depth individual interviews are conducted with 3 to 4 subjects
  - (In early rounds, emphasis is on general concepts)
  - (In later rounds, emphasis is on question wording, format)

- Comments are written up when each interview is done

- Interviewer meet with researchers to discuss results, propose potential modifications

- The questionnaire is revised by designers

- Is another round of cognitive testing to be conducted?
  - **YES**
  - **NO**

**Field Pretesting Phase (optional)**

- Survey is administered in the field
Appendix E

Ethics approval
June 4, 2008

Ms. Jing Jin
15 MacPherson Avenue, Unit 15-104
Kingston, ON K7M 2W8

Dear Ms. Jin,

Study Title: Communicating Results of New Genomic Tests to Physicians: Exploring Physicians' Attitudes, Needs and Concerns About New Genomic Tests

Co-Investigators: Dr. J. Brettschneider, Dr. M. Brundage, Dr. D. Feldman-Stewart

I am writing to acknowledge receipt of your recent ethics submission. We have examined the protocol, information letter and questionnaire for your project (as stated above) and consider it to be ethically acceptable. This approval is valid for one year from the date of the Chair's signature below. This approval will be reported to the Research Ethics Board. Please attend carefully to the following list of ethics requirements you must fulfill over the course of your study:

- **Reporting of Amendments**: If there are any changes to your study (e.g. consent, protocol, study procedures, etc.), you must submit an amendment to the Research Ethics Board for approval. (see http://www.queensu.ca/vpr/reb.htm).

- **Reporting of Serious Adverse Events**: Any unexpected serious adverse event occurring locally must be reported within 2 working days or earlier if required by the study sponsor. All other serious adverse events must be reported within 15 days after becoming aware of the information.

- **Reporting of Complaints**: Any complaints made by participants or persons acting on behalf of participants must be reported to the Research Ethics Board within 7 days of becoming aware of the complaint. Note: All documents supplied to participants must have the contact information for the Research Ethics Board.

- **Annual Renewal**: Prior to the expiration of your approval (which is one year from the date of the Chair's signature below), you will be reminded to submit your renewal form along with any new changes or amendments you wish to make to your study. If there have been no major changes to your protocol, your approval may be renewed for another year.

Yours sincerely,

[Signature]

Chair, Research Ethics Board

Date

Study Code: EPID-263-08

Investigators please note that if your trial is registered by the sponsor, you must take responsibility to ensure that the registration information is accurate and complete.
The membership of this Research Ethics Board complies with the membership requirements for Research Ethics Boards as defined by the Tri-Council Policy Statement; Part C Division 5 of the Food and Drug Regulations, OHRP, and U.S DHHS Code of Federal Regulations Title 45, Part 46 and carries out its functions in a manner consistent with Good Clinical Practices.

Federalwide Assurance Number: #FWA00004184
#IRB00001173

Current 2008 membership of the Queen's University Health Sciences & Affiliated Teaching Hospitals Research Ethics Board

Dr. A.F. Clark Emeritus Professor, Department of Biochemistry, Faculty of Health Sciences, Queen's University (Chair)
Dr. H. Abdollah Professor, Department of Medicine, Queen's University
Dr. C. Cline Assistant Professor, Department of Medicine
Director, Office of Bioethics, Queen's University
Clinical Ethicist, Kingston General Hospital
Rev. T. Deline Community Member
Dr. M. Evans Community Member
Dr. S. Irving Psychologist, Providence Care, St. Mary's of the Lake Hospital Site
Prof. L. Keeping-Burke Assistant Professor, School of Nursing, Queen's University
Dr. J. Low Emeritus Professor, Department of Obstetrics and Gynaecology, Queen's University and Kingston General Hospital
Dr. W. Racz Emeritus Professor, Department of Pharmacology & Toxicology, Queen's University
Dr. H. Richardson Assistant Professor, Department of Community Health & Epidemiology
Project Coordinator, NCIC CTG, Queen's University
Dr. B. Simechison Assistant Professor, Department of Anesthesiology, Queen's University
Dr. A.N. Singh WHO Professor in Psychosomatic Medicine and Psychopharmacology
Professor of Psychiatry and Pharmacology
Chair and Head, Division of Psychopharmacology, Queen's University
Director & Chief of Psychiatry, Academic Unit, Quinte Health Care, Belleville General Hospital
Dr. E. Tsai Assistant Professor, Department of Paediatrics and Office of Bioethics, Queen's University
Rev. J. Warren Community Member
Ms. K. Weisbaum LL.B. and Adjunct Instructor, Department of Family Medicine (Bioethics)
Dr. S. Wood Director, Office of Research Services (Ex-Officio)
Appendix F

Introduction letter (cover letter)

Dear Dr. (interviewee name),

Thank you for your interest in the project entitled “Communicating Results of New Genomic Tests to Physicians: Exploring physicians’ attitudes, needs and concerns about new genomic tests”. This study is conducted by Jing Jin under the supervision of Dr. Julia Brettschneider, Dr. Michael Brundage, and Dr. Deb Feldman-Stewart at Queen’s University.

The project is focusing on new genomic tests estimating cancer recurrence risk. The purpose of the project is to develop a mail-out questionnaire to determine:

- a) oncology physicians’ attitudes towards these tests,
- b) what information physicians would like included in the test reports, and
- c) how they think new genomic test results would impact their treatment recommendations.

A customized genomic test will be provided in the questionnaire to represent the whole series of tests we are interested in.

The goal of this interview is to ask for oncology physicians’ feedback on the development of the questionnaire. We want to ensure that:

- a) every part of the questionnaire is clear to you;
- b) you are satisfied with the questions and the answer options; and
- c) we didn’t miss any issues that would be considered important by you and your colleagues.

As an oncology physician, your feedback is very valuable.

During the interviewing process, Jing Jin, the student investigator will present the drafted questionnaire to you, and ask some questions regarding its content. It will take approximately one hour of your time.

As you would expect, all information obtained during the course of this study is strictly confidential and your anonymity will be protected at all times. Data will be stored in locked files and will be available only to principal investigators of this study. You will not be identified in any publication or reports.

This study is being supported by Canadian Institutes of Health Research (CIHR) and Natural Sciences and Engineering Research Council of Canada (NSERC). Additionally, this study has been approved by the Queen’s University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board.
If you would like further information about this study, please contact Jing Jin at (613) 484-8268. If you have further questions or concerns, please contact Dr. Michael Brundage at (613) 544-2631 ext. 4144 or Dr. Deb Feldman-Stewart at (613) 533-6000, ext. 78516. If I have questions regarding ethics issues in this study, please contact Dr. Albert Clark, Chair, Queen’s University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board at (613) 533-6081.

Sincerely,

_________________________________________   ___________________________ (Date)

Jing Jin

Master of Science candidate, Department of Community Health and Epidemiology, Queen’s University
Appendix G

The first version of the questionnaire in the interviews (Version 1)
Survey of Physicians’ Attitudes Toward Genomic Testing

Conducted by

Queens UNIVERSITY

Supported by

Canadian Institutes of Health Research (CIHR) and
Natural Sciences and Engineering Research Council of Canada (NSERC)
Direction

Thank you for your participation in this survey. It consists of 4 parts, and it will take about 20 mins to complete.

Part I presents a case scenario, followed by questions relevant to it. Then a specific new genomic test estimating breast cancer recurrence risk will be reviewed, followed by a few questions.

Part II presents another case scenario, followed by scenario-related questions similar to those in part I.

Part III asks you general questions about genomic testing, not relevant to either scenario.

Part IV asks for some background information and any comments you may have.

If you have questions or concerns, please contact:

Jing Jin
Queen's Cancer Research Institute
Suite 300, 10 Stuart St., Kingston, ON, Canada, K7L 3N6
Phone number:  (613) 484-8268
Email: 5jj14@queensu.ca
Part I  Scenario A

Case A: The patient is a 45-year-old premenopausal woman. She underwent a right modified radical mastectomy 2 weeks ago. She was found to have:

- Stage IIA breast cancer, node-negative (15 of 15), M0
- 1.5cm (T1c)
- Tubular carcinoma
- Grade 1, well differentiated
- Estrogen receptor (ER)-positive
- Progesterone receptor (PgR)-positive
- Her2-negative
- Normal chest x-ray, bone scan, and liver function tests
- No first-degree relatives with breast cancer
- No cardiovascular disease or other nonmalignant systemic disease

Please answer the following questions about this patient (Please mark the most appropriate box ☐):

1. How would you classify this patient’s breast cancer recurrence risk?
   - ☐ Low risk
   - ☐ Intermediate risk
   - ☐ High risk
   - ☐ Unsure
   - ☐ Other, please specify: ___________

2. What would you estimate this patient’s breast cancer recurrence risk at 10 years to be?
   - ☐ 0% to 10%
   - ☐ 11% to 20%
   - ☐ More than 20%
   - ☐ Unsure
   - ☐ Other, please specify: ___________

3. What treatment would you usually recommend for this patient at this point?
   - ☐ None
   - ☐ Hormonal Therapy only
   - ☐ Chemotherapy Therapy only
   - ☐ Chemotherapy and Hormonal Therapy
   - ☐ Other, please specify: ___________

4. How confident are you about your treatment recommendation?
   - ☐ Not at all confident
   - ☐ Not very confident
   - ☐ Fairly confident
   - ☐ Very confident
   - ☐ Other, please specify: ___________
Introduction to Oncotype DX™:

The clinicopathological features in “scenario A” suggest that the patient is a good candidate for a new genomic prognostic test -- Oncotype DX™. Below is a brief introduction to this test as it is currently developed. Further information about Oncotype DX™ is provided in the Appendix at the end of this questionnaire.

Target Population:  • Stage I or II, N-, ER+ breast cancer patients who receive tamoxifen

What the test does:  • Predicts the risk of breast cancer recurrence within 10 years

Nature of the test:  • Measures the expression of 21 genes (16 cancer genes including the HER2 gene, and 5 reference genes) in samples of tissue from primary breast cancer

Validation:  • The test was validated retrospectively in clinical trial populations (NSABP B-14) and in studies done by Kaiser Permanente. The clinical trial for the prospective validation is ongoing.

Result output:  • Produces a Recurrence Score (RS), a number between 0 and 100, that is correlated to an estimated likelihood of breast cancer recurrence within 10 years of initial diagnosis

Oncotype DX™ test result for the patient in Scenario A:

Below is the test result from Oncotype DX™ showing that the Recurrence Score of the patient in Scenario A is 10 (low risk). Interpretation: patients with a Recurrence Score of 10 in the clinical validation study had an average rate of distant recurrence at 10 years of 7% (95% CI: 4%-9%).
Considering the Oncotype DX™ test result for the patient in Scenario A:

(Please mark the most appropriate box ☑):

5. How does the test result influence your certainty about the recurrence risk estimate you made?
   - Increases
   - Decreases
   - Remains the same
   - Not sure, comment: _____________

6. Would the test be helpful in making a treatment recommendation?
   - Definitely would not
   - Probably would not
   - Unsure or neutral
   - Probably would
   - Definitely would

7. Would the test change your treatment recommendation? (If your answer is no change (Definitely would not/Probably would not/Unsure), please proceed to question 9.)
   - Definitely would not
   - Probably would not
   - Unsure or neutral
   - Probably would
   - Definitely would

8. If “Probably would” or “Definitely would”, what treatment would you recommend now for this patient?
   - None
   - Hormonal Therapy only
   - Chemotherapy Therapy only
   - Chemotherapy and Hormonal Therapy
   - Other, please specify: _____________

Proceed to question 9
9. Would you be generally willing to recommend that Oncotype DX™ be used in this type of patient? (If your answer is positive (Probably would/Definitely would), please proceed to Part II.)

☐ Definitely would not
☐ Probably would not
☐ Unsure or neutral
☐ Probably would
☐ Definitely would

Proceed to Part II

10. If your answer is “Definitely would not”, “Probably would not” or “Unsure or neutral”, what are your concerns? (You can mark as many options as you choose to)

☐ No concerns
☐ I am sure about my estimate of this patient's recurrence risk, and I can make the treatment recommendation without this test
☐ The patient may have concerns about the test
☐ Accuracy or the predictive power of the test may not be high enough
☐ Reliability of the test may not be good enough
☐ The test may not be completely clinically validated
☐ The test may not provide any new information beyond the current knowledge
☐ This test may have not been used widely
☐ Insurance coverage and the financial burden may be an issue
☐ Other, please specify: ______________

Proceed to Part II
Part II  Scenario B

Case B: The patient is also a 45-year-old premenopausal woman. She underwent a left modified radical mastectomy 2 weeks ago. She was found to have:

– Stage IIA breast cancer, node-negative (15 of 15), M0
– 3 cm (T2)
– Infiltrative ductal carcinoma
– Grade 1, moderately differentiated
– Estrogen receptor (ER)-positive
– Progesterone receptor (PgR)-positive
– Her2-negative
– Normal chest x-ray, bone scan, and liver function tests
– No first-degree relatives with breast cancer
– No cardiovascular disease or other nonmalignant systemic disease

Please answer the following questions about this patient: *(Please mark the most appropriate box □):*

11. **How would you classify this patient’s breast cancer recurrence risk?**
   - □ Low risk
   - □ Intermediate risk
   - □ High risk
   - □ Unsure
   - □ Other, please specify:
     ______________

12. **What would you estimate this patient’s breast cancer recurrence risk at 10 years to be?**
   - □ 0% to 10%
   - □ 11% to 20%
   - □ More than 20%
   - □ Unsure
   - □ Other, please specify:
     ______________

13. **What treatment would you usually recommend for this patient at this point?**
   - □ None
   - □ Hormonal Therapy only
   - □ Chemotherapy Therapy only
   - □ Chemotherapy and Hormonal Therapy
   - □ Other, please specify:
     ______________

14. **How confident are you about your treatment recommendation?**
   - □ Not at all confident
   - □ Not very confident
   - □ Fairly confident
   - □ Very confident
   - □ Other, please specify:
     ______________
**Oncotype DX™ test result for the patient in Scenario B:**

Below is the test result from Oncotype DX™ showing that the Recurrence Score of the patient in Scenario B is 30 (high risk). Interpretation: patients with a Recurrence Score of 30 in the clinical validation study had an average rate of distant recurrence at 10 years of 20% (95% CI: 15%-24%).

![Recurrence Score Graph](image)

**Considering the Oncotype DX™ test result for the patient in Scenario B:**

*(Please mark the most appropriate box [ ]):*

15. **How does the test result influence your certainty about the recurrence risk estimate you made?**
   - [ ] Increases
   - [ ] Decreases
   - [ ] Remains the same
   - [ ] Not sure, comment: ______________

16. **Would the test be helpful in making a treatment recommendation?**
   - [ ] Definitely would not
   - [ ] Probably would not
   - [ ] Unsure or neutral
   - [ ] Probably would
   - [ ] Definitely would
17. Would the test change your treatment recommendation? (If your answer is **no change** (Definitely would not/Probably would not/Unsure), please proceed to question 19.)

- Definitely would not
- Probably would not
- Unsure or neutral
- Probably would
- Definitely would

18. If “**Probably would”** or “**Definitely would**”, what treatment would you recommend now for this patient?

- None
- Hormonal Therapy only
- Chemotherapy Therapy only
- Chemotherapy and Hormonal Therapy
- Other, please specify: _____________

19. Would you be generally willing to recommend Oncotype DX™ be used in this type of patient? (If your answer is **positive** (Probably would/Definitely would), please proceed to Part III.)

- Definitely would not
- Probably would not
- Unsure or neutral
- Probably would
- Definitely would

Proceed to Part III

20. If your answer is “**Definitely would not”**, “**Probably would not”** or “**Unsure or neutral”**, what are your concerns? (**You can mark as many options as you choose to**)

- No concerns
- I am sure about my estimate of this patient's recurrence risk, and I can make the treatment recommendation without this test
- The patient may have concerns about the test
- Accuracy or the predictive power of the test may not be high enough
- Reliability of the test may not be good enough
- The test may not be completely clinically validated
- The test may not provide any new information beyond the current knowledge
- This test may have not been used widely
- Insurance coverage and the financial burden may be an issue
- Other, please specify: _____________

Proceed to Part III
Part III  What would you like to see in the test report

TEST RESULTS

21. Which type(s) of classification of recurrence risk would you prefer in the result report? (You can mark as many options as you choose to)
   - [ ] None
   - [ ] Low / high risk group
   - [ ] Low / intermediate / high risk group
   - [ ] 10 point scale (e.g., 1, 2, 3,…, 9, 10)
   - [ ] 100 point scale (e.g., 1, 2, 3,…, 99, 100)
   - [ ] Other, please specify: ___________

22. If the test result could be provided as risk categories (e.g., low/high risk group or low/intermediate/high risk group), would you like to know the average rate of recurrence risk in each category (e.g., the average rate of recurrence risk for patients in low risk group is 5%)? If your answer is negative (Definitely would not/Probably would not/Unsure), please proceed to question 24.
   - [ ] N/A, I don’t want the risk categories
   - [ ] Definitely would not
   - [ ] Probably would not
   - [ ] Unsure or neutral
   - [ ] Probably would
   - [ ] Definitely would

23. If “Probably would” or “Definitely would”, what would you like included in the Oncotype DX™ test report? (You can mark as many options as you choose to)
   - [ ] The average rate of recurrence risk in each category at 5 years
   - [ ] The average rate of recurrence risk in each category at 10 years
   - [ ] Other, please specify: ____________

Proceed to question 24
24. Would you like to know the individual recurrence risk of a patient? If your answer is negative (Definitely would not/Probably would not/Unsure), please proceed to question 26.

☐ Definitely would not
☐ Probably would not
☐ Unsure or neutral
☐ Probably would
☐ Definitely would

25. If “Probably would” or “Definitely would”, what would you like included in the Oncotype DX™ test report? (You can mark as many options as you choose to)

☐ The individual recurrence risk of a patient at 5 years
☐ The individual recurrence risk of a patient at 10 years
☐ Other, please specify: _____________

26. To what extent would you like the results information listed below included in the Oncotype DX™ test report? (Please mark the most appropriate box ☑ for each item)

<table>
<thead>
<tr>
<th></th>
<th>Definitely would not</th>
<th>Probably would not</th>
<th>Unsure or neutral</th>
<th>Probably would</th>
<th>Definitely would</th>
</tr>
</thead>
<tbody>
<tr>
<td>The RNA expression results of some well-known genes (Her-2/neu, Ki67, ER, PR gene, etc.) that are included in the test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The RNA expression results of all 21 genes of the test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The ER status determined by the test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The PR status determined by the test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The HER-2/neu gene amplification status determined by the test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
27. To what extent would you like the following information about the Oncotype DX™ test itself to be included in the test report?

<table>
<thead>
<tr>
<th>Information</th>
<th>Definitely would not</th>
<th>Probably would not</th>
<th>Unsure or neutral</th>
<th>Probably would</th>
<th>Definitely would</th>
</tr>
</thead>
<tbody>
<tr>
<td>The type of the platform used (e.g. bicolour cDNA microarray or high-throughput RT-PCR)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Names of the 21 genes measured</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Validity / reliability of the measurement process</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The accuracy of the test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The model used in this test to estimate the recurrence risk</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The accuracy of the model of this test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Scientific background of Oncotype DX™ (e.g., summaries of the development of the test or the clinical validation trials)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Oncotype DX™'s comparison with existing risk classifiers, such as Adjuvant! Online, St Gallen criteria, or NCCN guideline</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

28. Is there anything else you would like included in the test report?

__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
29. What would you prefer the individual recurrence risk of a patient to be presented? A result which integrates Oncotype DX™ test result with clinicopathological information? Or a separated estimate for each indicator? (Please mark the most appropriate box ☒)

☐ I prefer an integrated result
☐ I prefer to look at them separately
☐ I would like to see both, an integrated result from all indicators as well as the separated estimates
☐ Unsure or neutral
☐ OTHER, please specify: ___________

30. If the test indicates a patient classified into the low risk group with a recurrence rate of 6%, how would you prefer that be described in the report?

☐ It doesn’t matter. I don’t care
☐ “The chance that the cancer will recur is low.”
☐ “The chance that the cancer will recur is 6 in 100.”
☐ “The chance that the cancer will recur is 6 in 100. This is a low chance.”
☐ “The chance that the cancer will recur is 6%.”
☐ “The chance that the cancer will recur is 6%. This is a low chance.”
☐ Other, please specify: ___________

31. If the test indicates a patient classified into the high risk group with a recurrence rate of 50%, what would you prefer that be described in the report?

☐ It doesn’t matter. I don’t care
☐ “The chance that the cancer will recur is high.”
☐ “The chance that the cancer will recur is 50 in 100.”
☐ “The chance that the cancer will recur is 50 in 100. This is a high chance.”
☐ “The chance that the cancer will recur is 50%.”
☐ “The chance that the cancer will recur is 50%. This is a high chance.”
☐ Other, please specify: ___________

32. If the risk information is denoted by percentage, would you like to additionally know the 95% confidence interval of that percentage? (The 95% confidence interval is the risk range within which the true risk will lie 95 times out of 100)

☐ Definitely would not
☐ Probably would not
☐ Unsure or neutral
☐ Probably would
☐ Definitely would
☐ Other, please specify: ___________
33. What would you prefer about the risk expression? Recurrence rate or recurrence-free survival rate?

- [ ] It doesn’t matter. I don’t care
- [ ] Recurrence rate
- [ ] Recurrence-free survival rate
- [ ] Both
- [ ] Other, please specify: ___________

34. What visual format would you prefer the risk information to be presented?

(There will be 5 formats on the following 2 pages. You can mark as many options as you choose to; Use the tick box to the top left of the option)

<table>
<thead>
<tr>
<th>Patient</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk category</td>
<td>Low risk</td>
</tr>
<tr>
<td>Recurrence score</td>
<td>30</td>
</tr>
<tr>
<td>Predicted recurrence rate at 10 years</td>
<td>20%</td>
</tr>
</tbody>
</table>

Reference [Risk category (recurrence score)]:
Low risk (0-10), Intermediate risk (10-25), High risk (25-100)
35. To what extent would each of the following pieces of information from the Oncotype DX™ test be useful to you in deciding on treatment? (Please mark the most appropriate box ☐ for each item)

<table>
<thead>
<tr>
<th>Information</th>
<th>Not useful at all</th>
<th>Not very useful</th>
<th>Unsure or neutral</th>
<th>Somewhat Useful</th>
<th>Very Useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer recurrence risk category</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Average rate of recurrence in each category</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Cancer recurrence risk for individual patient</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Results of specific well-known genes measured in the test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Results of all 21 genes measured</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The ER and PR status determined by the test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The type of the platform used (e.g. bicolour cDNA microarray or high-throughput RT-PCR)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Validity / reliability of the measurement process</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The accuracy of the test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The model used in this test to estimate the recurrence risk</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The accuracy of the model of this test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Scientific background of Oncotype DX™ (e.g., summaries of the development of the test or the clinical validation trials)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Oncotype DX™’s comparison with existing risk classifiers, such as Adjuvant! Online, St Gallen criteria, or NCCN guideline</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
36. To what extent would each of the following pieces of information from the Oncotype DX™ test be useful to you in communicating with the patient?

<table>
<thead>
<tr>
<th>Information</th>
<th>Not useful at all</th>
<th>Not very useful</th>
<th>Unsure or neutral</th>
<th>Somewhat Useful</th>
<th>Very Useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer recurrence risk category</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average rate of recurrence in each category</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Cancer recurrence risk for individual patient</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Results of specific well-known genes measured in the test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Results of all 21 genes measured</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The ER and PR status determined by the test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The type of the platform used (e.g. bicolour cDNA microarray or high-throughput RT-PCR)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Validity / reliability of the measurement process</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The accuracy of the test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The model used in this test to estimate the recurrence risk</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The accuracy of the model of this test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Scientific background of Oncotype DX™ (e.g., summaries of the development of the test or the clinical validation trials)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Oncotype DX™’s comparison with existing risk classifiers, such as Adjuvant! Online, St Gallen criteria, or NCCN guideline</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
Part IV  Background Information & Comments

37. What is your professional background? *(Please mark the most appropriate box ☑)*
   - ☐ Radiation Oncologist
   - ☐ Medical Oncologist
   - ☐ Surgical Oncologist
   - ☐ Others, please specify: ______________________________________

38. How many years of clinical experience do you have (including residency etc.)?
   ________________________________________________________________

39. Do you have any additional comments/questions about this questionnaire/this study?
   ________________________________________________________________
   ________________________________________________________________
   ________________________________________________________________
   ________________________________________________________________
   ________________________________________________________________
   ________________________________________________________________
   ________________________________________________________________
   ________________________________________________________________
   ________________________________________________________________

Thank you for your time and your valuable input!
Appendix

What else the test may do:

- To identify patients who will benefit most from adjuvant tamoxifen and may not need adjuvant chemotherapy.
- To identify patients who may obtain more benefit from adjuvant chemotherapy (specifically (C)MF) than from tamoxifen.

Genes included in the Onco
type DX™ and Recurrence Score algorithm:

* Genomic Health Recurrence Score (RS) Algorithm (Onco
type DX™)

<table>
<thead>
<tr>
<th>Group</th>
<th>Genes</th>
<th>Adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proliferation</td>
<td>Ki67, STK15, survivin, cyclin B1, MYB2</td>
<td>+1.04</td>
</tr>
<tr>
<td>Her2</td>
<td>Her2, Grb7</td>
<td>+0.47</td>
</tr>
<tr>
<td>ER</td>
<td>ER, PR, BCL-2, SCUBE2</td>
<td>-0.34</td>
</tr>
<tr>
<td>Invasion</td>
<td>Stromelysin-3, CAT</td>
<td>+0.10</td>
</tr>
<tr>
<td>CD68</td>
<td>CD68</td>
<td>+0.05</td>
</tr>
<tr>
<td>BAG-</td>
<td>BAG-1</td>
<td>-0.07</td>
</tr>
<tr>
<td>GSTM-1</td>
<td>GSTM-1</td>
<td>-0.08</td>
</tr>
</tbody>
</table>

The development of the Oncotype DX™ test:

- The final 21 genes in the Oncotype DX™ gene panel were selected from 250 candidate genes.
- Three studies analyzed retrospectively clinical material of 440 patients from three groups: 224 patients with node-negative, ER-positive disease treated with tamoxifen, 79 patients with 10 or more positive axillary nodes, and 146 additional patients with operable breast cancer.

The validation of the Oncotype DX™ test:

NSABP B-14 Trial study:

- 668 People in NSABP trial B-14, who have been treated with tamoxifen.
- Cox proportional hazard model indicated that the RS was the only factor that emerged as a significant prognostic factor at 10 years (hazard ratio=3.21 [2.23- 4.61]; p< 0.00001). The other two factors used in the analysis were age (hazard ratio=0.71 [0.48, 1.05]) and tumor size (hazard ratio=1.26 [0.86, 1.85]).
- The results of the analysis:

* Genomic Health Recurrence Score in B-14 Trial (N=668)

<table>
<thead>
<tr>
<th>Recurrence Score (1–100)</th>
<th>Risk group</th>
<th>No. (%)</th>
<th>10-year distant recurrence rate (95% C.I.)</th>
<th>5-year distant recurrence rate (95% C.I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 18</td>
<td>Low</td>
<td>338 (51%)</td>
<td>6.8 % (4–9.6%)</td>
<td>2.1% (0.6-3.7%)</td>
</tr>
<tr>
<td>18 –30</td>
<td>Intermediate</td>
<td>149 (22%)</td>
<td>14.3% (8.3–20.3%)</td>
<td>9.2% (4.4-14%)</td>
</tr>
<tr>
<td>&gt;/= 31</td>
<td>High</td>
<td>181 (27%)</td>
<td>30.5% (23.6–37.4%)</td>
<td>22.1% (15.9-28.2%)</td>
</tr>
</tbody>
</table>
Kaiser Permanente study:
- Population-based case-control study among 4964 Kaiser Permanente patients (227 eligible cases who died of breast cancer and 446 eligible controls who did not die of breast cancer were matched for clinical characteristics).
- The results indicated that RS score was the strongest predictor of breast cancer death in multivariate analysis (odds ratio for death 6.5, p=0.002), that RS significantly predicted breast cancer death in patients treated with or without tamoxifen, and that the risk of breast cancer death at 10 years was similar in this population as in the B-14 population.

Chemotherapy benefit:
- A retrospective study of patient samples from the NSABP B-20 Study, which assessed the clinical benefit of adjuvant CMF (cyclophosphamide, methotrexate, and 5-fluorouracil), revealed preliminarily that Oncotype DX™ also predicts the magnitude of chemotherapy benefit.
- Only patients with high recurrence scores have statistically significant benefit from tamoxifen + chemo (28% absolute benefit) (Relative risk for high risk group: 0.26 [0.13 to 0.53], absolute increase in proportion distant recurrence free at 10 years for high risk group: improved from 60% to 88%).

Relationship with other risk classifiers:
- The Oncotype DX™ Recurrence Score is more accurate in the identification of the risk of distant recurrence than the NCCN and St Gallen’s guidelines which rely on age, tumor size, and tumor grade, based on the results from NSABP B-14 trial.
- Oncotype DX™ test is different from Adjuvant!Online (R²=0.43).

Biological technology:
- This new test is a Reverse Transcription Polymerase Chain Reaction (RT-PCR) assay that measures the expression of 21 genes in RNA extracted from formalin-fixed, paraffin-embedded (FFPE) samples of tissue from primary breast cancer.

Limitations:
- No treatment recommendations for breast cancer patients with intermediate recurrence risk.

Certification and regulation:
- The Oncotype DX™ test has received Clinical Laboratory Improvement Amendments (CLIA) approval in the United States.
- This Oncotype DX™ has been recommended by American Society of Clinical Oncology as a tumor marker to predict the risk of breast cancer recurrence in predefined population.
Appendix H

The final product of the questionnaire
Survey of Physicians’ Attitudes Toward Genomic Testing In Breast Cancer

Conducted by

Queen's University

Supported by
Canadian Institutes of Health Research (CIHR) and
Natural Sciences and Engineering Research Council of Canada (NSERC)
**Direction**

Thank you for your participation in this survey. It consists of 4 parts, and it will take about 20 mins to complete.

Part I presents a case scenario, followed by questions relevant to it. Then a specific new genomic test estimating the risk for distant recurrence of breast cancer will be reviewed, followed by a few questions.

Part II presents another case scenario, followed by scenario-related questions similar to those in part I.

Part III asks you general questions about genomic testing, not relevant to either scenario.

Part IV asks for some background information and any comments you may have.

If you have questions or concerns, please contact:

Jing Jin

Queen's Cancer Research Institute

Suite 300, 10 Stuart St., Kingston, ON, Canada, K7L 3N6

Phone number: (613) 484-8268

Email: 5jj14@queensu.ca
Case A: The patient is a 45-year-old premenopausal woman. She underwent a right breast lumpectomy 2 weeks ago. She was found to have:

- Stage I breast cancer
- 1.0 cm (T1b)
- Node negative (0/15 nodes)
- M0
- Tubular carcinoma
- Grade 1, well differentiated
- Estrogen receptor (ER) positive
- Progesterone receptor (PgR) positive
- Her2 negative
- No DCIS, and surgical margins are clear
- Normal chest x-ray, bone scan, and liver imaging
- No first-degree relatives with breast cancer
- No cardiovascular disease or other nonmalignant systemic disease

Please answer the following questions about this patient (Please mark the most appropriate box ☐):

1. How would you quantify this patient’s risk of distant recurrence at 10 years without adjuvant treatment? (e.g., percentage, range etc.)

________________________________________________________________________
________________________________________________________________________

2. Given your above estimate of risk, how would you classify this patient’s risk of distant recurrence, using the following categories?
   - Low risk
   - Intermediate risk
   - High risk
   - Unsure
   - Other, please specify:

________________________________________________________________________

3. What systemic treatment would you usually recommend for this patient at this point?
   - None
   - Hormonal Therapy only
   - Chemotherapy Therapy only
   - Chemotherapy and Hormonal Therapy
   - Other, please specify:

________________________________________________________________________

4. How confident are you about your systemic treatment recommendation?
   - Not at all confident
   - Not very confident
   - Fairly confident
   - Very confident
   - Other, please specify:

________________________________________________________________________
**Introduction to Oncotype DX™:**

The clinicopathological features in “scenario A” suggest that the patient is a good candidate for a new genomic prognostic test -- Oncotype DX™. Below is a brief introduction to this test as it is currently developed.

**Target Population:**
- Stage I or II, N-, ER+ breast cancer patients who receive tamoxifen

**What the test does:**
- Predicts the risk of distant recurrence within 10 years

**Nature of the test:**
- Measures the expression of 21 genes (16 cancer genes including the HER2 gene, and 5 reference genes) in samples of tissue from primary breast cancer

**Validation:**
- The test was validated retrospectively in clinical trial populations (NSABP B-14) and in studies done by Kaiser Permanente. The clinical trial for the prospective validation is ongoing.

**Result output:**
- Produces a Recurrence Score (RS), a number between 0 and 100, that is correlated to an estimated likelihood of distant recurrence within 10 years of initial diagnosis

**Further information**
- Provided in the Appendix at the end of this questionnaire

---

**Oncotype DX™ test result for the patient in Scenario A:**

Below is the test result from Oncotype DX™ showing that the Recurrence Score of the patient in Scenario A is 10 (low risk). Interpretation: patients with a Recurrence Score of 10 in the clinical validation study had an average rate of distant recurrence at 10 years of 7% (95% CI: 4%-9%).

![Graph showing risk levels](image)

Note: the schema of Oncotype DX™ in this questionnaire retains as originally published.
Considering the Oncotype DX™ test result for the patient in Scenario A, when you just saw her: (Please mark the most appropriate box □):

5. How does the test result influence your certainty about the estimate of distant recurrence risk you made?
   - Increases
   - Decreases
   - Remains the same
   - Not sure, comment: ______________

6. Would the test be helpful in making a treatment recommendation?
   - Definitely would not
   - Probably would not
   - Unsure or neutral
   - Probably would
   - Definitely would

7. Would the test change your treatment recommendation you made before? (If your answer is “Definitely would not”, “Probably would not”, or “Unsure/neutral”, please go to question 9.)
   - Definitely would not
   - Probably would not
   - Unsure or neutral
   - Probably would
   - Definitely would

8. If “Probably would” or “Definitely would”, what treatment would you recommend now for this patient?
   - None
   - Hormonal Therapy only
   - Chemotherapy Therapy only
   - Chemotherapy and Hormonal Therapy
   - Other, please specify: ______________
9. **Would you be generally willing to recommend that Oncotype DX™ be used in this type of patient?** (If your answer is **positive** (Probably would/Definitely would), please go to Part II.)

   - □ Definitely would not
   - □ Probably would not
   - □ Unsure or neutral
   - □ Probably would
   - □ Definitely would

**Go to Part II**

10. **If your answer is “Definitely would not”, “Probably would not” or “Unsure or neutral”, what are your concerns?** (Mark as **many options** as you want)

   - □ No concerns
   - □ I am sure about my estimate of this patient's distant recurrence risk, and I can make the treatment recommendation without this test
   - □ The patient may have concerns about the test
   - □ Accuracy or the predictive power of the test may not be high enough
   - □ Reliability of the test may not be good enough
   - □ The test has not been completely clinically validated
   - □ The test may not provide any new information beyond the current knowledge
   - □ This test has not been used widely
   - □ This test is not available as a routine test in Canada
   - □ The cost and the insurance coverage of the test may be an issue
   - □ Other, please specify: _____________
Part II  Scenario B

Case B: The patient is a 60-year-old postmenopausal woman. She underwent a left breast lumpectomy 2 weeks ago. She was found to have:

- Stage I breast cancer
- 1.5 cm (T1c)
- Node negative (0/15 nodes)
- M0
- Infiltrative ductal carcinoma
- Grade 1, well differentiated
- Estrogen receptor (ER) positive
- Progesterone receptor (PgR) positive
- Her2 negative
- No DCIS, and surgical margins are clear
- Normal chest x-ray, bone scan, and liver imaging
- No first-degree relatives with breast cancer
- No cardiovascular disease or other nonmalignant systemic disease

Please answer the following questions about this patient:

(Please mark the most appropriate box □):

11. How would you quantify this patient’s risk of distant recurrence at 10 years without adjuvant treatment? (e.g., percentage, range, etc.)

__________________________________

__________________________________

__________________________________

12. Given your above estimate of risk, how would you classify this patient’s risk of distant recurrence, using the following categories?

- Low risk
- Intermediate risk
- High risk
- Unsure
- Other, please specify:

______________________________

13. What systemic treatment would you usually recommend for this patient at this point?

- None
- Hormonal Therapy only
- Chemotherapy Therapy only
- Chemotherapy and Hormonal Therapy
- Other, please specify:

______________________________

14. How confident are you about your systemic treatment recommendation?

- Not at all confident
- Not very confident
- Fairly confident
- Very confident
- Other, please specify:

______________________________
Onco
type DX™ test result for the patient in Scenario B:

Below is the test result from Onco
type DX™ showing that the Recurrence Score of the patient in Scenario B is 41 (high risk). Interpretation: patients with a Recurrence Score of 41 in the clinical validation study had an average rate of distant recurrence at 10 years of 28% (95% CI: 21%-34%).

Considering the Onco
type DX™ test result for the patient in Scenario B, when you just saw her:  (Please mark the most appropriate box □):

15. How does the test result influence your certainty about the estimate of distant recurrence risk you made?
   □ Increases
   □ Decreases
   □ Remains the same
   □ Not sure, comment: ______________

16. Would the test be helpful in making a treatment recommendation?
   □ Definitely would not
   □ Probably would not
   □ Unsure or neutral
   □ Probably would
   □ Definitely would
17. Would the test change your treatment recommendation you made before? (If your answer is “Definitely would not”, “Probably would not”, or “Unsure/neutral”, please go to question 19.)
- [ ] Definitely would not
- [ ] Probably would not
- [ ] Unsure or neutral
- [ ] Probably would
- [ ] Definitely would

18. If “Probably would” or “Definitely would”, what treatment would you recommend now for this patient?
- [ ] None
- [ ] Hormonal Therapy only
- [ ] Chemotherapy Therapy only
- [ ] Chemotherapy and Hormonal Therapy
- [ ] Other, please specify: _____________

19. Would you be generally willing to recommend Oncotype DX™ be used in this type of patient? (If your answer is positive (Probably would/Definitely would), please go to question 21.)
- [ ] Definitely would not
- [ ] Probably would not
- [ ] Unsure or neutral
- [ ] Probably would
- [ ] Definitely would

20. If your answer is “Definitely would not”, “Probably would not” or “Unsure or neutral”, what are your concerns? (Mark as many options as you want)
- [ ] No concerns
- [ ] I am sure about my estimate of this patient’s distant recurrence risk, and I can make the treatment recommendation without this test
- [ ] The patient may have concerns about the test
- [ ] Accuracy or the predictive power of the test may not be high enough
- [ ] Reliability of the test may not be good enough
- [ ] The test has not been completely clinically validated
- [ ] The test may not provide any new information beyond the current knowledge
- [ ] This test has not been used widely
- [ ] This test is not available as a routine test in Canada
- [ ] The cost and the insurance coverage of the test may be an issue
- [ ] Other, please specify: _____________

Go to question 21
21. Is there any case scenario you believe Oncoype DX™ would be most helpful in making a treatment recommendation?

_______________________________________________________________________________

_______________________________________________________________________________

_______________________________________________________________________________

_______________________________________________________________________________

Part III  Oncoype DX™ test report: what do you prefer

TEST RESULTS-------------------------------------------------------------------------------------------------------------------------------------

22. Which type(s) of classification of distant recurrence risk would you prefer in the test report? (Mark as many options as you want)
   - None (percentage of distant recurrence risk only)
   - Low / high risk group
   - Low / intermediate / high risk group
   - Low / intermediate / high / very high risk group
   - 10 point/score scale (e.g., 1, 2, 3,…, 9, 10)
   - 100 point/score scale (e.g., 1, 2, 3,…, 99, 100)
   - Other, please specify: ___________

23. If relative risk categories (e.g., low/high risk group) are provided, would you like to know the average rate of distant recurrence for each risk category (e.g., the average rate of distant recurrence for patients in low risk group is 5%)? (If your answer is “N/A”, “Definitely would not”, “Probably would not” or “Unsure or neutral”, please go to question 25.)
   - N/A, I don’t want the risk categories
   - Definitely would not
   - Probably would not
   - Unsure or neutral
   - Probably would
   - Definitely would

Go to question 25

24. If “Probably would” or “Definitely would”, what would you like included in the Oncoype DX™ test report? (Mark as many options as you want)
   - The average rate of distant recurrence for each category at 5 years
   - The average rate of distant recurrence for each category at 10 years
   - The average rate of distant recurrence for each category at 20 years
   - Other, please specify: ___________

- 8 -
25. Would you like to know the risk of distant recurrence in individual patients? (If your answer is “Definitely would not”, “Probably would not” or “Unsure/neutral”, please go to question 27.)
- [ ] Definitely would not
- [ ] Probably would not
- [ ] Unsure or neutral
- [ ] Probably would
- [ ] Definitely would

26. If “Probably would” or “Definitely would”, what would you like included in the Onco
type DX™ test report? (Mark as **many options** as you want)
- [ ] The risk of distant recurrence in a patient at **5** years
- [ ] The risk of distant recurrence in a patient at **10** years
- [ ] The risk of distant recurrence in a patient at **20** years
- [ ] Other, please specify: _____________

27. To what extent would you like the results information listed below included in the Onco
type DX™ test report? (Mark the most appropriate box ☒ for each item)

<table>
<thead>
<tr>
<th></th>
<th>Definitely would not</th>
<th>Probably would not</th>
<th>Unsure or neutral</th>
<th>Probably would</th>
<th>Definitely would</th>
</tr>
</thead>
<tbody>
<tr>
<td>The RNA expression results of only well-known genes (ER gene, PR gene, Her-2/neu, Ki67, Bcl2, etc.) that are included in the test</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>The RNA expression results of all 21 genes of the test</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
</tbody>
</table>
| The ER status (+/−) determined by Onco
type DX™ * | [ ] | [ ] | [ ] | [ ] | [ ] |
| The PR status (+/−) determined by Onco
type DX™ * | [ ] | [ ] | [ ] | [ ] | [ ] |
| The HER-2/neu status (+/−) determined by Onco
type DX™ * | [ ] | [ ] | [ ] | [ ] | [ ] |

*: May differ from the status arrived at using conventional methods, e.g., immunohistochemistry (IHC) or fluorescent in situ hybridization (FISH).
### TEST ITSELF

#### 28. To what extent would you like the following information about the Oncotype DX™ test itself to be included in the test report?

<table>
<thead>
<tr>
<th>Information</th>
<th>Definitely would not</th>
<th>Probably would not</th>
<th>Unsure or neutral</th>
<th>Probably would</th>
<th>Definitely would</th>
</tr>
</thead>
<tbody>
<tr>
<td>The type of the platform used (e.g., bicolour cDNA microarray or high-throughput RT-PCR)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Names of the 21 genes measured in this test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Validity / reliability of the measurement process</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The accuracy of the test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The statistical model used to estimate the risk of distant recurrence</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The accuracy of this statistical model</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Scientific background of this test (e.g., summaries of the test development or the clinical validation trials)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>This test’s comparison with existing risk classifiers, such as Adjuvant! Online, St Gallen criteria, or NCCN guideline</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

#### 29. Is there anything else you would like included in the test report?

_______________________________________________________________________________
_______________________________________________________________________________
_______________________________________________________________________________
_______________________________________________________________________________
30. How would you prefer the risk of distant recurrence in a patient to be presented? A result which integrates Oncotype DX™ test result with clinicopathological information? Or a separated estimate for each indicator? (Please mark the most appropriate box □)
□ I prefer an integrated result
□ I prefer to look at them separately
□ I would like to see both, an integrated result from all indicators as well as the separated estimates
□ Unsure or neutral
□ Other, please specify: ___________

31. If the test indicates a patient classified into the low risk group with a distant recurrence risk of 6%, how would you prefer that be described in the report?
□ It doesn’t matter. I don’t care.
□ “The chance that the cancer will recur and spread is low.”
□ “The chance that the cancer will recur and spread is 6 in 100.”
□ “The chance that the cancer will recur and spread is 6 in 100. This is a low chance.”
□ “The chance that the cancer will recur and spread is 6%.”
□ “The chance that the cancer will recur and spread is 6%. This is a low chance.”
□ Other, please specify: ___________

32. If the risk information is denoted by percentage, would you like to additionally know the 95% confidence interval of that percentage? (The 95% confidence interval is the risk range within which the true risk will lie 95 times out of 100)
□ Definitely would not
□ Probably would not
□ Unsure or neutral
□ Probably would
□ Definitely would
□ Other, please specify: ___________

33. What would you prefer about the risk expression? Distant recurrence rate or distant recurrence-free survival rate?
□ It doesn’t matter. I don’t care
□ Distant recurrence rate
□ Distant recurrence-free survival rate
□ Both
□ Other, please specify: ___________
34. In which visual format would you prefer the risk information to be presented? (There are 5 formats on the following 2 pages. You can mark as many options as you want; Use the tick box to the top left of the option)

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Risk category</th>
<th>Recurrence score</th>
<th>Predicted recurrence rate at 10 years</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>High risk</td>
<td>34</td>
<td>23%</td>
<td>Recurrence Score (RS): Low risk (RS &lt; 18), Intermediate risk (18 ≤ RS &lt;31), High risk (RS ≥ 31)</td>
</tr>
</tbody>
</table>
Patient ID

Other, please specify:

_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
To what extent would each of the following pieces of information from the Oncotype DX™ test be useful to you in deciding on treatment? (Please mark the most appropriate box ☑️ for each item)

<table>
<thead>
<tr>
<th>Information</th>
<th>Not useful at all</th>
<th>Not very useful</th>
<th>Unsure or neutral</th>
<th>Somewhat Useful</th>
<th>Very Useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categories of distant recurrence risk</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Average rate of distant recurrence for each category</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Distant recurrence risk in individual patients</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The RNA expression results of only specific well-known genes measured in the test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The RNA expression results of all 21 genes measured</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The ER, PR and HER-2/neu status determined by the test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The type of the platform used (e.g., bicolour cDNA microarray or high-throughput RT-PCR)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Validity / reliability of the measurement process</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The accuracy of the test</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The statistical model used in this test to estimate the risk of distant recurrence</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The accuracy of this model</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Scientific background of Oncotype DX™ (e.g., summaries of the development of the test or the clinical validation trials)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Oncotype DX™’s comparison with existing risk classifiers, such as Adjuvant! Online, St Gallen criteria, or NCCN guideline</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
36. To what extent would each of the following pieces of information from the **Oncotype DX™** test be useful to you in communicating with the patient?

<table>
<thead>
<tr>
<th>Information</th>
<th>Not useful at all</th>
<th>Not very useful</th>
<th>Unsure or neutral</th>
<th>Somewhat Useful</th>
<th>Very Useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categories of distant recurrence risk</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>Average rate of distant recurrence for each category</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>Distant recurrence risk in individual patients</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>The RNA expression results of only specific well-known genes measured in the test</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>The RNA expression results of all 21 genes measured</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>The ER, PR and HER-2/neu status determined by the test</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>The type of the platform used (e.g. bicolour cDNA microarray or high-throughput RT-PCR)</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>Validity / reliability of the measurement process</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>The accuracy of the test</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>The statistical model used in this test to estimate the risk of distant recurrence</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>The accuracy of this model</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>Scientific background of <strong>Oncotype DX™</strong> (e.g., summaries of the development of the test or the clinical validation trials)</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td><strong>Oncotype DX™</strong>'s comparison with existing risk classifiers, such as Adjuvant! Online, St Gallen criteria, or NCCN guideline</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>
37. Do you have any additional comments on how Oncotype DX™ would be more clinically useful?
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________

Part IV  Background Information & Comments

38. How much did you know about Oncotype DX™ before you saw this questionnaire?
☐ I didn’t know anything about it
☐ I knew only a little about it
☐ I knew something about it
☐ I knew it very well
☐ Others, please specify: ________________________________________________

39. How many years of clinical experience do you have (including residency etc.)?
_________________________________________________________________________
_________________________________________________________________________

40. Do you have any additional comments/questions about this questionnaire/this study?
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________

Thank you for your time and your valuable input!
Appendix  Further information about Oncotype DX™

What else the test may do:

• To identify patients who will benefit most from adjuvant tamoxifen and may not need adjuvant chemotherapy.
• To identify patients who may obtain more benefit from adjuvant chemotherapy (specifically (C)MF) than from tamoxifen.

Biological technology:

• This new test is a Reverse Transcription Polymerase Chain Reaction (RT-PCR) assay that measures the expression of 21 genes in RNA extracted from formalin-fixed, paraffin-embedded (FFPE) samples of tissue from primary breast cancer.

Genes included in the Oncotype DX™ and Recurrence Score algorithm:

| Recurrence Score = |
| + 0.47 * HER2 Group Score |
| - 0.34 * ER Group Score |
| + 1.04 * Proliferation Group Score |
| + 0.10 * Invasion Group Score |
| + 0.05 * CD68 |
| - 0.08 * GSTM1 |
| - 0.07 * BAG1 |

<table>
<thead>
<tr>
<th>Group</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proliferation</td>
<td>Ki67, STK15, survivin, cyclin B1, MYB2</td>
</tr>
<tr>
<td>Her2</td>
<td>Her2, Grb7</td>
</tr>
<tr>
<td>ER</td>
<td>ER, PR, BCL-2, SCUBE2</td>
</tr>
<tr>
<td>Invasion</td>
<td>Stromelysin-3, CAT</td>
</tr>
<tr>
<td>CD68</td>
<td>CD68</td>
</tr>
<tr>
<td>BAG-</td>
<td>BAG-1</td>
</tr>
<tr>
<td>GSTM-1</td>
<td>GSTM-1</td>
</tr>
</tbody>
</table>

• The Recurrence Score is calculated from the expression of 16 cancer-related genes (listed above) and five reference genes (Beta-actin, GAPDH, RPLOP, GUS, and TFRC) used to normalize the expression of those cancer genes.
• A score for each group was calculated using a weighted average based on the expressions of individual genes in this group.

The development of the Oncotype DX™ test:

• The final 21 genes in the Oncotype DX™ gene panel were selected from 250 candidate genes.
• Three studies analyzed retrospectively clinical material of 440 patients from three groups: 224 patients with node-negative, ER-positive disease treated with tamoxifen, 79 patients with 10 or more positive axillary nodes, and 146 additional patients with operable breast cancer.

The validation of the Oncotype DX™ test:

NSABP B-14 Trial study:

• 668 People in NSABP trial B-14, who have been treated with tamoxifen.
• Cox proportional hazard model indicated that the RS was the only factor that emerged as a significant prognostic factor at 10 years (hazard ratio=3.21 [2.23- 4.61]; p< 0.00001). The other two factors used in the analysis were age (hazard ratio=0.71 [0.48, 1.05]) and tumor size (hazard ratio=1.26 [0.86, 1.85]).
The results of the analysis:

<table>
<thead>
<tr>
<th>Recurrence Score (1–100)</th>
<th>Risk group</th>
<th>No. (%)</th>
<th>10-year distant recurrence rate (95% C.I.)</th>
<th>5-year distant recurrence rate (95% C.I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 18</td>
<td>Low</td>
<td>338 (51%)</td>
<td>6.8 % (4–9.6%)</td>
<td>2.1% (0.6-3.7%)</td>
</tr>
<tr>
<td>18 –30</td>
<td>Intermediate</td>
<td>149 (22%)</td>
<td>14.3% (8.3–20.3%)</td>
<td>9.2% (4.4-14%)</td>
</tr>
<tr>
<td>&gt;= 31</td>
<td>High</td>
<td>181 (27%)</td>
<td>30.5% (23.6–37.4%)</td>
<td>22.1% (15.9-28.2%)</td>
</tr>
</tbody>
</table>

Kaiser Permanente study:
- Population-based case-control study among 4964 Kaiser Permanente patients (227 eligible cases who died of breast cancer and 446 eligible controls who did not die of breast cancer were matched for clinical characteristics).
- The results indicated that RS score was the strongest predictor of breast cancer death in multivariate analysis (odds ratio for death 6.5, p=0.002), that RS significantly predicted breast cancer death in patients treated with or without tamoxifen, and that the risk of breast cancer death at 10 years was similar in this population as in the B-14 population.

Chemotherapy benefit:
- A retrospective study of patient samples from the NSABP B-20 Study, which assessed the clinical benefit of adjuvant CMF (cyclophosphamide, methotrexate, and 5-fluorouracil), revealed preliminarily that Oncotype DX™ also predicts the magnitude of chemotherapy benefit.
- Only patients with high recurrence scores have statistically significant benefit from tamoxifen + chemo (28% absolute benefit) (Relative risk for high risk group: 0.26 [0.13 to 0.53], absolute increase in proportion distant recurrence free at 10 years for high risk group: improved from 60% to 88%).

Relationship with other risk classifiers:
- The Oncotype DX™ Recurrence Score is more accurate in the identification of the risk of distant recurrence than the NCCN and St Gallen’s guidelines which rely on age, tumor size, and tumor grade, based on the results from NSABP B-14 trial.
- Oncotype DX™ test is different from Adjuvant!Online (R²=0.43).

Limitations:
- No treatment recommendations for breast cancer patients with an intermediate risk of distant recurrence.

Certification and regulation:
- The Oncotype DX™ test has received Clinical Laboratory Improvement Amendments (CLIA) approval in the United States.
- This Oncotype DX™ has been recommended by American Society of Clinical Oncology as a tumor marker to predict the distant recurrence risk of breast cancer in predefined population.