NIPH
Norwegian Institute of Public Health

REPORT

2019

# SINGLE TECHNOLOGY ASSESSMENT:

## Prosigna Gene Signature to Assess Expected Benefit from Chemotherapy in Breast Cancer. Assessment of manufacturer's submission

| | |
|---|---|
| **Title** | Prosigna Gene Signature to Assess Expected Benefit from Chemotherapy in Breast Cancer. Assessment of manufacturer's submission |
| **Norwegian title** | Prosigna genprofiltest for å vurdere forventet nytte av kjemoterapi ved brystkreft. Vurdering av innsendt informasjon |
| **Institution** | Norwegian Institute of Public Health (Folkehelseinstituttet) |
| **Authors** | Camilla Stoltenberg, *Director*<br>Beate Charlotte Fagerlund*, *Health economist*<br>Kishan K. Chudasama*, *Senior Advisor (project coordinator)*<br>Kjetil G. Brurberg, *Researcher*<br>Bjarne Robberstad, *Health economist*<br>Chris Rose, *Statistician*<br>Lene K. Juvet, *Department director*<br>Atle Fretheim, *Researcher*<br>*Chudusama and Fagerlund contributed equally to this report |
| **ISBN** | 978-82-8406-020-0 |
| **Type of report** | Single-technology assessment (Hurtigmetodevurdering) |
| **No. of pages** | 68 |
| **Client** | Bestillerforum RHF |
| **Project number** | ID2017_073 |
| **Subject heading (MeSH)** | Breast Neoplasms; Gene Expression Profiling; Biomarker, Tumors; Sequence Analysis, RNA; Systematic Review; Technology Assessment, Biomedical; Cost Benefit Analysis |
| **Citation** | Fagerlund BC*, Chudasama KK*, Brurberg KG, Robberstad B, Rose C, Juvet LK, Fretheim A. Prosigna Gene Signature to Assess Expected Benefit from Chemotherapy in Breast Cancer. Assessment of manufacturer's submission. Oslo: Norwegian Institute of Public Health, 2019. (*Shared first authorship) |

Norwegian Institute of Public Health (NIPH)
Oslo, June 2019

# Executive summary

## Background

Breast cancer can be treated with chemotherapy, hormone therapy and radiation, or a combination of these to prevent the spread of cancer cells, after surgical removal of the tumor. When assessing whether a patient should be offered chemotherapy, information about prognosis is important. Patients at high risk of recurrence should be offered chemotherapy, while patients at low risk are not very likely to gain from such treatment, in which case side effects outweigh the benefits. The assessment of risk of recurrence is based on clinical findings, e.g. tumor size, lymph node involvement, and expression of certain receptors on the cancer cells.

In this health technology assessment, we have considered a molecular profiling panel, Prosigna, which is meant to improve the assessment of recurrence risk among women who have undergone surgical treatment for breast cancer.

Our assessment is based on documentation submitted by the manufacturer of Prosigna, Nanostring.

## Objective

The objective was to investigate the prognostic accuracy, clinical effectiveness, and cost effectiveness of Prosigna in patients diagnosed with breast cancer.

In Norway, the group considered as potentially eligible for the test is patients with breast cancer who had their tumor removed, are node-negative and where the tumor is classified as hormone receptor-positive/human epidermal growth factor receptor 2-negative (HR+/HER2-).

## Methods

### Prognostic accuracy and clinical effectiveness

To validate the submitted evidence we extracted data from the key publications and critically appraised the risk of bias in the findings.

### Health economics

We assessed cost-effectiveness estimates of Prosigna compared with current practice for HR+/ HER2-, node negative patients, provided by the submitter, and similarly for a defined subgroup of patients at higher risk of recurrence ("Luminal B like pT1c-pT2 pN0"). The estimates were based on a hybrid model with a decision tree combined with a Markov model. Cost-effectiveness estimates for Prosigna versus current practice were modelled over a 50-year time horizon, for patients aged 58. We performed separate analyses using the submitted model, adjusting some of the input variables based on revised assumptions. Also, we performed alternative scenario analyses for differences in chemotherapy use with and without use of the Prosigna test, based on various other data sources.

## Results

### Prognostic accuracy

There is convincing evidence of a correlation between the observed risk of recurrence and the risk stratification score generated by the Prosigna test. For patients classified as low risk, the ten-year risk of recurrence is around 4%. For the intermediate risk group, the risk is around 10%, and for the high-risk group around 21%. We also expressed the performance of the test in terms of prognostic sensitivity and specificity. When we merged the intermediate risk group with the low risk group (intermediate test constitutes a "negative" test), we estimated the test's sensitivity to 52%, and its specificity to 77%. If the intermediate group was merged with the high-risk group (intermediate risk constitutes a "positive" test), the sensitivity and specificity were estimated to be 83% and 42%, respectively.

The estimates presented above reflect the test's performance when used as a standalone tool. In practice, it can be anticipated that the test will be used as a supplement to the current risk stratification approaches.

### Clinical effectiveness

We did not identify comparative studies where patients were allocated to risk stratification with or without the Prosigna and followed over time. Without such comparative studies, it is difficult to estimate the clinical utility of Prosigna, i.e. Prosigna's impact on the use of chemotherapy and patient outcomes such as disease-free survival and side effects from chemotherapy. Studies exploring the prognostic value of adding Prosigna to other prognostic variables in multivariate regression models suggest that Prosigna adds prognostic information that may be useful when deciding about further use of chemotherapy. However, these data are sparse, and it remains

unclear to what extent Prosigna will contribute to fewer recurrences or a reduction in the needless use of chemotherapy than current practice.

### Health economics

The incremental cost-effectiveness ratio (ICER) based on the revised economic model for HR+/ HER2-, node negative patients, is calculated to NOK 897,923 per QALY gain in our base-case analysis. The estimate is based on questionable assumptions and is highly sensitive to changes in the chemotherapy use parameter. We estimated the total added costs of implementing Prosigna for this group in Norway, to about NOK 13.5 million in year five.

The calculated incremental cost-effectiveness ratio (ICER) based on the revised economic model for the subgroup of "Luminal B like pT1c-pT2 pN0"-patients (38% of the "all node negative"-population) would be NOK 98,188 per QALY gain. Implementing Prosigna for this subgroup in Norway, would lead to a total cost saving of NOK 9.9 million in year five.

## Discussion

### Clinical efficacy and safety

Several studies have assessed the extent to which the test is able to categorise patients into groups with a low, intermediate or high risk of recurrence. However, the utility of this information for clinical decision making is uncertain. Uncertainties are mainly caused by lack of data regarding the accuracy of procedures that are currently used when selecting patients for chemotherapy, and uncertainties regarding the emphasis clinicians and patients will put on Prosigna when deciding for or against chemotherapy. The relatively low sensitivity of the Prosigna test means that it yields a considerable number of false negative classifications, which entails a risk that patients who could benefit from chemotherapy are not offered the treatment.

Evidence from multivariate regression analyses indicate that Prosigna contributes information of prognostic value beyond tests and assessment tools in current use. The manufacturer of Prosigna has not based any of the analyses in the submission on these data, and we did not see how these results could be used to estimate an expected health gain from introducing Prosigna testing into clinical practice.

### Health economics

Regarding the model input, empirical data on chemotherapy use is lacking. In the submitted model the proportion of chemotherapy use following Prosigna testing was based on the opinions of 11 British oncologists. The proportion of chemotherapy use

in current practice (no test) was assumed to be the same across the different risk of recurrence-groups – a dubious assumption.

We used data derived from a Norwegian study and recommendations in Norwegian clinical practice guidelines for breast cancer management, which we believe yield more trustworthy estimates than those in the submitted analyses.

Further, there is controversy regarding the utility value of "the risk profiling knowledge to patients", which was assumed by the submitter. We do not consider preferences for knowing the test result "health-related", and we therefore consider this parameter irrelevant in this case.

The economic model submitted by NanoString did not incorporate sensitivity and specificity, and we are uncertain what the consequences of this are for the cost-effectiveness-estimates. If we were to prepare a health economic model for a prognostic test such as Prosigna, we would probably have opted for a different approach, and included the test's prognostic sensitivity and specificity into the model.

## Conclusion

There is probably a statistical association between Prosigna's risk prediction and the observed risk of distant recurrence after breast cancer. However, it is uncertain to what extent Prosigna contributes prognostic information that translates into better clinical results in terms of lower recurrence rates and reduced chemotherapy use

Conclusions about the cost-effectiveness of Prosigna cannot be made as we do not have reliable data on chemotherapy use and clinical outcomes for patients who have or have not undergone Prosigna testing.

# Sammendrag (Norwegian summary)

## Bakgrunn

Brystkreft kan behandles med cellegift, hormonterapi og stråling, eller en kombinasjon av disse for å hindre spredning av kreftceller etter kirurgisk fjerning av svulsten. Informasjon om prognose er viktig i vurderingen av om en pasient bør anbefales cellegift. Pasienter med høy risiko for tilbakefall bør få tilbud om cellegift. Pasienter med lav risiko for tilbakefall vil ha vanligvis ha mindre nytte av cellegift, og dermed kan faren for bivirkninger veie tyngre enn forventet nytte av behandlingen. Vurderingen av faren for tilbakefall baseres i dag på kliniske funn, blant annet svulstens størrelse, spredning til lymfekjertler og tilstedeværelse av visse egenskaper på kreftcellene.

I denne metodevurderingen har vi vurdert om Prosigna, en genprofiltest, kan bidra til mer nøyaktig prediksjon av risiko for tilbakefall blant kvinner som har gjennomgått kirurgisk behandling for brystkreft. Vurderingen er basert på dokumentasjon innsendt av Prosigna-produsenten, Nanostring.

## Formål

Formålet har vært å undersøke den prognostiske nøyaktigheten, kliniske effekten, samt kostnadseffektiviteten av Prosigna for pasienter diagnostisert med brystkreft. I Norge er det brystkreftopererte i kategorien hormone receptor-positive/human epidermal growth factor receptor 2-negative (HR+/HER2-) uten spredning til lymfeknuter som regnes som aktuelle for Prosigna-testing.

## Metode

### Prognostisk nøyaktighet og klinisk nytte

For å etterprøve den innsendte dokumentasjonen hentet vi ut data fra nøkkelpublikasjoner og vurderte resultatene med tanke på risiko for systematiske feilkilder.

### Helseøkonomi

Vi vurderte innsendte estimater for Prosignas kostnadseffektivitet sammenlignet med gjeldende praksis for vurdering av prognose i den aktuelle pasientgruppen, og tilsvarende for en definert undergruppa av pasienter med økt risiko for tilbakefall ("Luminal B like pT1c-pT2 pN0").

I dokumentasjonspakken ble det benyttet en hybridmodell, med beslutningstre kombinert med en Markov-modell. Prosigna-produsenten brukte denne modellen til å beregne kostnadseffektiviteten av Prosigna sammenlignet med gjeldende praksis over en 50 års tidshorisont, for 58 år gamle pasienter. Vi utførte separate analyser der vi justerte inputdata i den innsendte modellen i tråd med reviderte forutsetninger. Vi utførte også tre *de novo* scenarioanalyser der vi vurderte ulike andeler av kjemoterapibruk med og uten Prosigna.

## Resultat

### Prognostisk nøyaktighet

Det er overbevisende dokumentasjon for at det er sammenheng mellom Prosignas risikoprediksjon og faktisk risiko for tilbakefall av brystkreftsykdom. I gruppen som klassifiseres som lavrisikopasienter av Prosigna er faren for tilbakefall omkring fire prosent over ti år. For dem som kategoriseres som middels eller høy risiko er faren for tilbakefall anslått til henholdsvis 10 prosent og 21 prosent.

Testens egenskaper kan også uttrykkes i form av prognostisk sensitivitet og spesifisitet. Da vi slo sammen lav- og middels-risiko gruppene og regnet disse som «negativ test» ble sensitiviteten 52 % og spesifisiteten 77 %. Hvis vi isteden slo sammen gruppene med middels og høy risiko («positiv test») ble sensitivitet og spesifisitet beregnet til henholdsvis 83 % og 42 %. Estimatene som presenteres over reflekterer testens egenskaper når den brukes alene. I praksis kan man forvente at testen ikke vil benyttes alene, men som supplement til andre tilgjengelige risikostratifiseringsmetoder.

### Klinisk nytteverdi

Vi ble ikke forelagt sammenliknende studier der pasienter ble fulgt opp over tid etter å ha blitt vurdert med eller uten bruk av Prosigna. Uten slike sammenliknende studier er det vanskelig å beregne den kliniske nytteverdien av testen. Den prognostiske verdien av å bruke Prosigna som supplement til annen prognostisk informasjon er i noen studier vurdert ved hjelp av multivariate regresjonsmodeller. Slike regresjonsanalyser viser at Prosigna bidrar med prognostisk tilleggsinformasjon som kan være

nyttig når man skal ta beslutninger om bruk av kjemoterapi. I hvilken grad tilleggsinformasjonen man får fra Prosigna bidrar til reduksjon i antall tilbakefall eller unødig kjemoterapibruk forblir imidlertid usikkert.

### Helseøkonomi

Den inkrementelle kostnadseffektivitetsratio (ICER) som var basert på den reviderte økonomiske modellen for HR+/ HER2-, lymfeknutenegative pasienter, er 897,923 norske kroner per vunnet QALY i base-case analysen. Den totale årlige merkostnaden av å implementere Prosigna for denne gruppen i Norge beregnes til 13,5 millioner norske kroner.

For subgruppen «Luminal B like pT1c-pT2 pN0»-pasienter (38 % av alle lymfeknutenegative pasienter) viste vår reviderte økonomiske modell en inkrementell kostandseffektivitetsratio (ICER) på 98,188 norske kroner per vunnet QALY. Videre utførte vi også en budsjettkonsekvensanalyse for subgruppen «Luminal B like pT1c-pT2 pN0»-pasienter. Den totale årlige kostnadsbesparingen for å implementere Prosigna til denne subgruppen i Norge beregnes til 9,9 millioner norske kroner.

## Diskusjon

### Klinisk effekt og sikkerhet

Flere studier viser at Prosigna kan bidra til å predikere hvilke pasienter som har lav, middels eller høy risiko for tilbakefall, men helsegevinsten som følger av denne informasjonen er uviss. Usikkerheten skyldes til dels at vi mangler informasjon om hvor godt dagens praksis predikerer nytte av kjemoterapi, men vi mangler også informasjon om hvor stor vekt Prosigna vil bli tillagt når det skal tas beslutninger om kjemoterapi. Det må tas høyde for at Prosigna-testen gir et betydelig antall falske negative prediksjoner, noe som innebærer en risiko for at pasienter ikke tilbys kjemoterapi selv om de kunne hatt nytte av det.

Data fra regresjonsanalyser tyder på at Prosigna har prognostisk verdi utover de tester og undersøkelser som brukes i dag. Produsenten har ikke lagt resultater fra disse analysene til grunn i sine beregninger av testens nytteverdi, og vi har heller ikke klart å bruke disse tallene til å anslå forventet gevinst av å innføre Prosigna i klinisk praksis.

### Helseøkonomi

Når det gjelder modellens inputdata mangler vi empiriske data på kjemoterapibruken i Norge. I produsentens modell er andelene av kjemoterapibruk etter Prosignatesting basert på vurderingene til 11 britiske onkologier. Andelen kjemoterapibruk i dagens praksis (ingen test) ble antatt å være lik på tvers av de ulike risikogruppene –

en tvilsom antakelse. I våre analyser brukte vi data fra en norsk studie og anbefalinger fra norske kliniske retningslinjer. Vi mener dette gir mer pålitelige estimater enn dataene som ble brukt i den innsendte analysen.

Videre er vi usikre på nytteverdien av «risikoprofileringskunnskap til pasienter», som er en nytteverdi antatt av produsenten. Vi vurderer ikke en preferanse om å få vite sitt testresultatet som «helsemessig», og anser derfor denne parameteren som irrelevant i denne sammenhengen.

Den økonomiske modellen som NanoString har benyttet inkorporerer ikke sensitivitet og spesifisitet, og vi er usikre på hvilken innvirkning dette har for beregningene av kostnadseffektivitet. Hvis vi skulle ha utarbeidet en helseøkonomisk modell for en prognostisk test som Prosigna, ville vi sannsynligvis ha valgt en annen tilnærming, og inkludert testens prognostiske sensitivitet og spesifisitet i modellen.

## Konklusjon

Det er sannsynligvis nær sammenheng mellom Prosignas risikoprediksjon og faktisk risiko for tilbakefall av brystkreftsykdom. Det er usikkert om Prosigna bidrar med prognostisk informasjon som lar seg omsette til bedre kliniske resultater i form av lavere tilbakefallsrater og samtidig reduksjon i bruk av kjemoterapi.

Vi kan ikke trekke konklusjoner om kost-nytte-forholdet for Prosigna ettersom vi verken har gode sammenliknende data for kjemoterapibruk eller kliniske resultater for pasienter som har, eller ikke har, blitt vurdert med bruk av testen.

# Table of contents

# Preface

A single-technology assessment is one type of health technology assessment (HTA) that can be mandated in "The National System for Managed Introduction of New Health Technologies" within the Specialist Health Service in Norway (https://nyemetoder.no/).

Within this system, the Ordering Forum RHA ("Bestillerforum RHF"), where the four Regional Health Authorities are represented, decides on which technologies should be assessed and the type of assessment needed. In a single-technology assessment, the technology (a pharmaceutical or a device) is assessed based on documentation submitted by the company owning the technology, or their representatives ("the submitter").

The HTA unit of the Norwegian Institute of Public Health (NIPH) receives and evaluates the submitted documentation, but is not the decision-making authority. Single-technology assessments conducted at NIPH are published on our website (www.fhi.no) and on https://nyemetoder.no/

The following were involved in the process of making this single-technology assessment:

| Role | Name |
|---|---|
| Project coordinator | Kishan Kumar Chudasama |
| Health economists | Beate C. Fagerlund |
| | Bjarne Robberstad |
| Effectiveness and safety evaluators | Kishan Kumar Chudasama |
| | Kjetil G. Brurberg |
| | Atle Fretheim |
| Statisticians | Chris Rose |
| | Kjetil G. Brurberg |
| Research librarian | Elisabeth Hafstad |
| Department director | Lene K. Juvet |
| Submitter | Nanostring |

We thank the following for commenting on draft versions of this report, or parts of it: Marius Stensland, Håkan Olsson, Bjørn Naume, Ivar Sønbø Kristiansen.

The aim of this report is to support well-informed decisions in health care that lead to improved quality of services. The evidence should be considered together with other relevant issues, such as clinical experience and patient preferences.

| Kåre Birger Hagen | Lene K. Juvet | Kishan K. Chudasama |
|---|---|---|
| *Scientific director* | *Department director* | *Project coordinator* |

## Progress log

| Date | Correspondence |
|---|---|
| June 25, 2017 | Proposal on this device |
| Sept 25, 2017 | The commissioning forum commissioned a single technology assessment |
| Nov 22, 2017 | Dialogue with technology manufacturer |
| Dec 13, 2017 | Meeting with technology manufacturer |
| Jan 9, 2018 | Manufacturer confirms intent to submit documentation |
| Jan-March, 2018 | Clinical experts asked, and PICO made and sent to manufacturer |
| July 16, 2018 | Submission received from manufacturer |
| August 18, 2018 | Valid submission acknowledged |
| May-June, 2019 | Norwegian Institute of Public Health external review process |
| May-June, 2019 | Norwegian Institute of Public Health internal review process |
| June 19, 2019 | Feedback from technology manufactory on the report |
| June 28, 2019 | Report Submitted |
| Aug 15, 2019 | Minor changes in wording of abstract and numbering of tables/figures. |

# Objective

The objective was to investigate the prognostic accuracy, clinical effectiveness and cost effectiveness of the Prosigna™ Breast Cancer Prognostic Gene Signature Assay in patients with early-stage breast cancer.

# Background

Breast cancer is the most common type of cancer in women and constitutes 22% of all newly diagnosed cancers among females (1). It primarily affects women above the age of 50 (1). In 2017, 3905 women were diagnosed with breast cancer in Norway (2). Compared to people of the same age and sex without breast cancer, patients with breast cancer have a five-year survival rate of 90% (relative survival rate) (2).

Early-stage breast cancer patients undergo surgery (mastectomy or breast conserving surgery) to remove the primary tumor. Subsequently, some patients are treated with hormone therapy, chemotherapy, radiation or a combination of these to prevent future breast cancer recurrence. The choice of treatment strategy is based on prognostic and predictive parameters such as size of the tumor, spread to lymph nodes, and tumor characteristics including expression of certain receptors/biomarkers on the cancer cells (1).

The identification of certain tumor biomarkers expressed by the tumor cells are important in determining the best treatment for the individual patient (1). For this purpose, immunohistochemical investigations of the estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and the proliferation marker Ki67 are performed routinely to classify patients, e.g. as “ER+ HER2-“ (1). National clinical practice guidelines are available that provide guidance for the management of patients with breast cancer, including the use of chemotherapy (1). To what extent treatment guidelines are adhered to is not well known, but it has been reported that the histological grading of tumors varies across hospital departments (2). Thus, it is not clear how accurate the current approach to prognosis assessment is, in practice.

A multitude of new biomarkers have been detected and validated in recent years, a development which has led to the emergence of genomic profiling technologies and selective molecular targeted therapies. These new technologies and therapies are adding onto the concept of personalized medicine or precision medicine. A number of tumor profiling tests have been developed to provide prognostic decision support in clinical practice. Currently, the breast cancer tumor profiling tests EndoPredict,

Oncotype DX, MammaPrint, IHC4 and Prosigna have gained popularity and are under investigation in several studies.

Recommendations on the use of tumor profiling or multigene tests varies across countries. In Sweden tumor profiling tests are not recommended in clinical practice due to lack of evidence (3) while in Denmark Prosigna (4) is recommended for use among women with breast cancers that are ER+/HER2- with less than three affected lymph nodes.

In the United Kingdom, the National Institute for Health and Care Excellence (NICE) recently published an assessment of five different tumor profiling tests: EndoPredict, MammaPrint, IHC4, Oncotype DX and Prosigna. In their guidance, NICE recommends EndoPredict, Oncotype DX and Prosigna to aid adjuvant chemotherapy decisions (5;6). The NICE-recommendation restricts the use of the tests to patients who have been assessed as being at intermediate risk of recurrence by means of the PREDICT tool or the Nottingham Prognostic Index.

In the present report we have only assessed the Prosigna Breast Cancer Prognostic Gene Signature Assay (Nanostring). The Prosigna assay is performed on tissue that has been removed during the original biopsy or surgery, and it yields a risk of recurrence score from 0 to 100. The risk of recurrence in node-negative cancers is classified as either low (score from 0 to 40), intermediate (41 to 60), or high (61 to 100).

According to Norwegian clinicians, the main target population for the Prosigna test is breast cancer patients who have had their tumor surgically removed, have no spread to lymph nodes (node-negative), and with a tumor classified as HR+/HER2-. The Prosigna manufacturer estimates this group to constitute around 1/3 of all patients diagnosed with breast cancer (i.e. approximately 1400 patients per year in Norway). These breast cancer patients are routinely treated with hormone therapy. In addition, around 35–40% also receive chemotherapy (Bjørn Naume, personal communication). The current use of chemotherapy is perceived as too high, and the Prosigna test is mainly seen as a prognostic tool that will identify more patients at low risk of recurrence, leading to less use of chemotherapy and thereby a reduced burden from side effects (7). The test is not foreseen to be used as a standalone test which guides treatment choices directly, but rather as an add-on tool to the current approach for assessing the patients' prognosis (1).

# Methods – Clinical evaluation

In line with the routine for single technology assessments for The National System for Managed Introduction of New Health Technologies within the Specialist Health Service, this report is based on a documentation package submitted by the supplier of the Prosigna Breast Cancer Prognostic Gene Signature Assay, Nanostring.

## Literature search and selection

### Literature search and identification relevant literature

The National Institute of Health and Care Excellence (NICE) recently reviewed Prosigna (5;6), and as a part of their review process NICE conducted systematic searches for literature. The selection of studies that Nanostring included in their submission was largely based on the search conducted for the NICE-review.

## Data extraction and analyses

In order to validate the data provided in the submission from Nanostring, we extracted the following variables from the articles that were identified by Nanostring as providing the most relevant evidence:

- Information about the study (authors, year of publication, setting, study design, clinical trial identification number and funding source)
- Participant characteristics (number of participants in the trial, age)
- Intervention and control characteristics
- Outcome data, in the case of Lænkholm et al., data were extracted from graphs using WebPlotDigitizer (available at https://apps.automeris.io/wpd/, accessed in November 2018).

### Risk of bias assessments

We assessed methodological quality using the QUADAS-2 checklist (8). This checklist is primarily used for assessing studies of diagnostic tests, and we took this into account in our application of the checklist, and in the interpretation of our assessments.

### Certainty of the evidence

We evaluated certainty of the evidence using the GRADE-tool developed by the GRADE working group (Grading of Recommendations Assessment: GRADE Working Group: Group GW. Available from: http://www.gradeworkinggroup.org/). According to this system, we categorized the certainty of the evidence for each outcome into one of four levels: high, moderate, low and very low certainty.

### Analysis

Prosigna classifies women as being at low, intermediate or high risk of recurrence. We analysed the results in two steps. First, we summarised the evidence for estimating risk of distant disease recurrence for patients in each of the three Prosigna risk groups. To assess risk across included studies we performed meta-analysis based on Kaplan-Maier survival estimates. Study results were pooled using a random effect model and generic invers variance methods offered by the *meta* package in R (9).

As a second analytic step, we assessed the prognostic performance of the Prosigna test in terms of sensitivity and specificity. To achieve this, we dichotomised the test results into "positive" or "negative". The choice of cut-off will impact on a test's performance. It is not obvious whether the intermediate risk group should be considered "positive" or "negative", i.e. whether it should be combined with the high or the low risk group. We therefore calculated the sensitivity and specificity both ways.

We used data from the available studies to estimate a «Summary ROC curve» based on a bivariate model developed by Reitsma and co-workers (10). Based on para-estimates from the model we calculated a «Summary operating point» (estimates for sensitivity and specificity) with 95% confidence intervals. Analyses were performed using the package *mada* (11) in R (9). A random effects model was assumed due to heterogeneity in the populations studied, and study endpoint and censoring event definitions.

## Stakeholder involvement

Initially, the project leader contacted external clinical experts, designated by the Regional Health Authorities, and provided information about the project.

In line with how the National System for Managed Introduction of New Health Technologies is meant to operate, the submission from Nanostring served as the main evidence base. Internal experts and external clinical experts have commented on drafts of this report. Likewise, internal and external health economists have commented on the health economic analyses.

# Results – Clinical evaluation

## Study selection

The search for literature was based on a search made by NICE (5), and the selection process was described in the manufacturer's submission. Briefly, 2336 references were identified of which 539 references were retrieved in full text. A total of 504 references were excluded because they did not include data on Prosigna or were irrelevant to the question of interest. A total of 34 studies were listed as partly relevant in the submission from Nanostring (Figure 1). Four studies were highlighted as the documentation basis for Prosigna's prognostic performance (12-15).

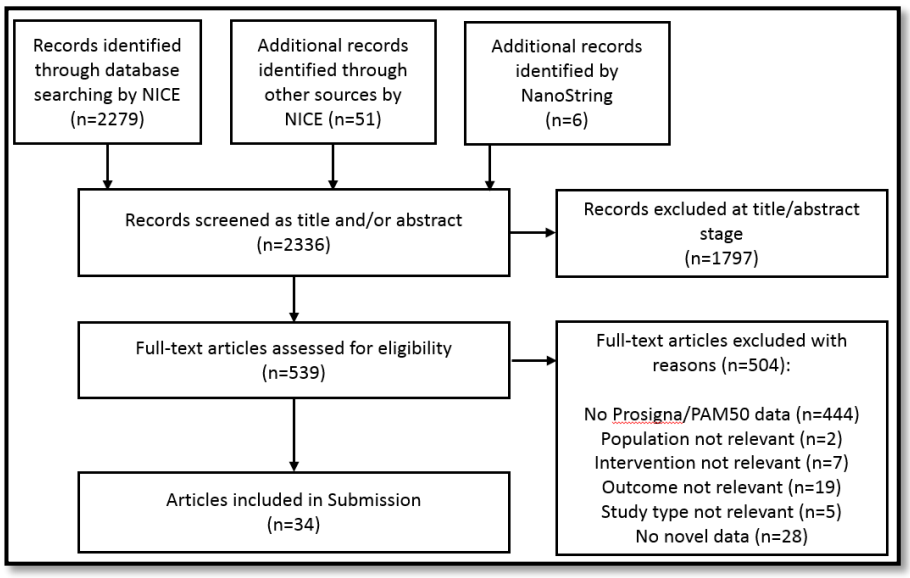



**Figure 1.** *Flow chart of the literature search. The illustration is taken from the manufacturer's submission.*

## Description of included studies

Table 1 lists the four studies that form the evidence base for prognostic performance, in the manufacturer's submission. Each of the four studies are described more thoroughly in the text.

**Table 1. Included references, sorted under the corresponding source of data.**

| References | Participants[1] | Outcome (Follow up time) |
|---|---|---|
| ***Translational Sub Study of Arimidex, Tamoxifen, Alone or in Combination (TransATAC Study)*** | | |
| Sestak 2018 (12) | 774 participants<br>591 with nodal status N0<br>Mean age was 64 years | Time to Distant Recurrence<br>(Median 10 years) |
| ***Adjuvant Treatment in Patients With Hormone Receptor-positive Breast Cancer With Good to Moderate Differentiation (ABCSG8 Study)*** | | |
| Gnant 2014 (14) | N=1478;<br>1047 with nodal status N0<br>Median age was 63 years | Time to Distant Recurrence<br>(Median 11 years) |
| ***Danish Breast Cancer Cooperative Group (DBCG Study)*** | | |
| Lænkholm 2018 (13) | N=2558<br>1163 with nodal status N0<br>939 between 50-59 years<br>1082 between 60-69 years<br>537 ≥ 70 years | Time to Distant Recurrence<br>(Median 9.2 years) |
| ***Oslo1 Study*** | | |
| Ohnstad 2017 (15) | 653 participants<br>419 with nodal status N0<br>382 < 55 years<br>271 ≥ 55 years | Distant Disease Free Survival<br>(Median 7.1 years) |

[1]The number of participants refers to the total number of participants included in each study, and may deviate from the subset of participants who are included in the analysis due to nodal, HR, HER2 and treatment status.

### *TransATAC*

The TransATAC (Translational Sub Study of Arimidex, Tamoxifen, Alone or in Combination) trial (12) randomized 9,366 patients to receive either 1) anastrozole (1 mg) plus tamoxifen placebo, 2) tamoxifen (20 mg) plus anastrozole placebo or 3) a combination of tamoxifen/anastrozole.

Sestak et al. (12) used tumor blocks from the TransATAC study and included blocks from patients with hormone receptor-positive early-stage breast cancer treated for five years with either tamoxifen or anastrozole. The main endpoint was distant recurrence-free survival defined as time from diagnosis until distant recurrence or death due to breast cancer. Contralateral breast cancers and deaths due to causes other than breast cancer were treated as censoring events. All analyses were performed on 10-year follow-up data.

*ABCSG8 Study*

The ABCSG8 study was initiated to test adjuvant treatment in patients with hormone receptor-positive breast cancer with good to moderate differentiation (14). A total of 3,901 women with hormone-positive breast cancer were randomized to two years of adjuvant tamoxifen followed by three years of anastrozole or five years of tamoxifen. Patient were recruited between 1996 and 2004.

The testing of Prosigna by Gnant et al. (14) was based on a retrospective analysis of samples from 1,478 patients from the ABCSG8 study that passed the quality assurance for the Prosigna assay. The main endpoint was distant recurrence-free survival defined as the interval from diagnosis until distant recurrence or death due to breast cancer. Contralateral breast cancers and deaths due to other causes than breast cancer were treated as censoring events. All analyses were performed on 10-year follow-up data.

*DBCG Study*

The Danish Breast Cancer Cooperative Group (DBCG)-cohort consists of breast tumor tissue samples collected and archived from postmenopausal women with hormone receptor-positive primary breast cancer diagnosed between 2000 and 2003. The cohort includes all women in the population-based DBCG-database who received the required loco-regional treatment and were allocated to five years of tamoxifen. None of the patients received chemotherapy, and radiotherapy was administered according to DBCG-guidelines.

Lænkholm et al. (13) is based on retrospective analysis of biopsies from 2558 patients. The primary endpoint was time to distant recurrence or death due to breast cancer. All secondary carcinomas (including contralateral breast cancer) and deaths due to other causes than breast cancer were treated as competing risk events.

*Oslo1 Study*

The Oslo1 study population consists of consecutive patients with early-stage breast cancer from the observational Oslo Micrometastasis Project. Patients were treated according to national recommendations for surgery, radiation therapy, adjuvant chemotherapy, and adjuvant endocrine therapy at the time of enrollment (1995-1998) based on clinicopathologic characteristics.

The aim of Ohnstad et al. (15) was to evaluate the long-term prognostic value of the Prosigna-defined subtypes and risk of recurrence scores in patients with HR+/HER2– early-stage breast cancer. The study end-points were distant disease-free survival and breast cancer-specific survival.

## Results derived from manufacturer's submission

### *Prognostic estimates*

We used the manufacturer's submission to extract data on distant recurrence free survival for our target population, i.e. HR positive, HER2 negative, node negative women who had not received chemotherapy.

Data from the four studies suggest that Prosigna risk of recurrence scores clearly differentiate between risk groups (Table 2). Two studies (12;14) report Kaplan-Meier survival estimates for ten year distant recurrence free survival between 96% (95% CI 94 to 98) and 97% (95% CI 94 to 98) in the low risk group, between 86% (95% CI 81 to 92) and 90% (95% CI 86 to 93) in the intermediate risk groups, and between 67% (95% CI 59 to 76) and 85% (95% CI 78 to 89) in the high risk groups.

For two of the studies, the manufacturer's submission did not report survival estimates for our target population (Table 2). We were able to extract the needed data from the publication (13) or by contacting study authors (15).

**Table 2. Number of participants and events (number of distant recurrences) across risk of recurrence based risk groups. Data limited to HR+/ HER2- and node negative patients.**

| *Study* | Risk of recurrence-score | No. of patients | No. of recurrences | % without recurrence after 10 years |
|---|---|---|---|---|
| *TransATAC* *Sestak et al.*[1] | Low (LN0) | 431 | 17 | 96 |
| | Inter (LN0) | 180 | 22 | 88 |
| | High (LN0) | 128 | 38 | 70 |
| *ABCSG8 Study* *Gnant et al.*[1] | Low (LN0) | 474 | 15 | 97 |
| | Inter (LN0) | 311 | 27 | 91 |
| | High (LN0) | 199 | 27 | 86 |
| *DBCG Study* *Lænkholm et al.*[2] | Low (LN0) | 361 | 18 | 95 |
| | Inter (LN0) | 375 | 27 | 93 |
| | High (LN0) | 427 | 76 | 82 |
| *Oslo1 Study* *Ohnstad et al.*[3] | Low (LN0) | 16 | 2 | 88 |
| | Inter (LN0) | 27 | 2 | 93 |
| | High (LN0) | 22 | 5 | 77 |

[1]Data derived from manufacturer's submission (Nanostring). [2]Numbers extracted from Figure 2 in publication and converted into percentages. [3]Additional data received from study authors.

### *Hazard ratios*

Three studies reported results from multivariate analyses in which risk of recurrence-score from Prosigna-testing was added as an explanatory factor alongside other well-known prognostic factors (e.g. tumor staging) to model the risk of distant recurrence. These analyses show that Prosigna adds prognostic value also when several other prognostic factors are controlled for.

Gnant and coworkers (14) showed that the continuous risk of recurrence score added prognostic information when it was added to a model consisting of other clinicopathological factors (e.g. age, tumor grade and tumor size). Briefly, a ten-point increase in risk of recurrence corresponded to a 37.5% increase in the risk of recurrence. When risk of recurrence-score was merged into three risk groups, the risk of recurrence seemed to be higher in the intermediate (HR 2.15; 95% CI 1.21 to 3.81) and in high risk groups (HR 4.26; 95% CI 2.44 to 7.43), compared to the low risk groups. In comparison, Ohnstad and coworkers (15) reported a hazard ratio of 2.25 (95% CI 0.77 to 6.60) for intermediate versus low, and 6.82 (95% CI 2.62 to 17.81) for high versus low risk groups.

Lænkholm and coworkers (13) reported hazard ratios for patient with zero to three positive nodes, and they showed lower risk of recurrence in the low risk group compared with the intermediate risk group (HR 0.53; 95% CI 0.33 to 0.85). The risk of recurrence in the high-risk group was higher than in the intermediate risk group (HR 1.81; 95% CI 1.33 to 2.44).

Taken together, the multivariate analyses show that the risk of recurrence-score adds prognostic information to what can be obtained from other clinicopathological factors. However, estimates of what these findings are likely to mean in terms of decision making about chemotherapy, or improved clinical outcomes, are not included in the manufacturer's submission.

### *Meta-analysis*

The manufacturer's submission conveys results from a meta-analysis in which the authors have pooled risk of recurrence estimates from two studies (12;14). Estimated percentages without distant recurrence at ten years were reported to be 96.2 (95% CI 94.7 to 97.3) in the low risk group, 89.2 (95% CI 86.1 to 91.7) in the intermediate risk group and 77.7 (95% CI 72.8 to 81.9) in the high-risk group.

## Additional analysis done by NIPH

We decided to perform some additional analyses to draw a more complete picture of the available documentation.

### Risk of bias and applicability concerns

The four included studies were in general well conducted, but patient selection may represent a potential risk of bias: Some of the included studies were sub-studies of randomised controlled trials, and the inclusion criteria in the trials may have led to a biased selection of patients. Further, the exclusion of patients who received chemotherapy is likely to imply that the spectrum of patients in the included studies is not fully representative of the patient group of interest in our report. We have not identified other reasons to suspect serious risks of bias affecting the study findings.

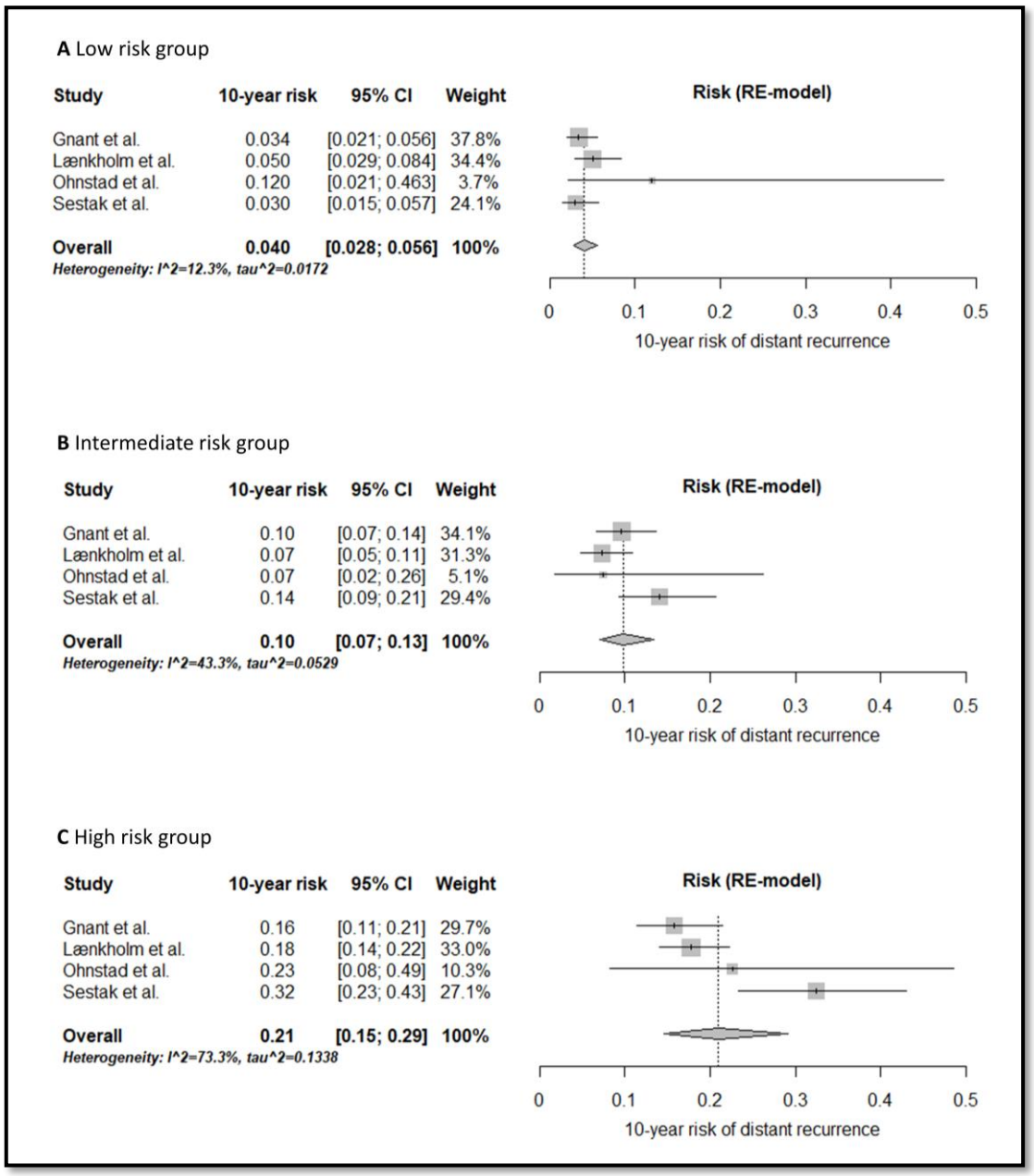
Regarding applicability concerns, our assessment of prognostic performance is based on the use of Prosigna as a standalone test. This assumption is a simplification, as decisions about adjunctive chemotherapy are not likely to be based on Prosigna scores alone. Rather, Prosigna scores are expected to supplement other clinicopathological information that currently informs decisions about adjunctive treatment. Thus, the question of interest is not if Prosigna offers utility when used as a standalone test, but to what extent Prosigna improves current risk stratification procedures. We lack data to provide clear answers to the latter question.

### Risk of recurrence

We estimated the 10-year risk of recurrence for each of the three risk groups (low, intermediate and high) by combining all available studies in meta-analysis (Figure 2). The analyses were limited to the subgroup of node negative women receiving hormone therapy.

The risk of recurrence was 4% (95% CI 3 to 6) in the low risk group, 10% (95% CI 7 to 13) in the intermediate risk group, and 21% (95% CI 15 to 29) in the high risk group. Hence, the differences in recurrence risk was statistically different in the three risk stratification groups. The observed differences in prognosis across the risk groups suggest that Prosigna may play a role in risk stratification. To further elucidate how the implementation of the test in clinical practice might affect patient flow, we also analysed the data using a diagnostic framework.

***Figure 2.*** *10-year risk of distant disease recurrence in each risk stratification group*




## Prognostic accuracy

*Sensitivity and specificity*

When analysing data within a diagnostic framework it is important to analyse data and interpret test results in accordance with suggested clinical practice. The choice of cut off for positive/negative test is important for a test's performance. Based on expert opinion, we decided to merge the low and intermediate risk groups, i.e. we considered these test results "negative". Hence, we classified the patients who had been stratified to the low or intermediate risk groups as true negatives if they did not experience distant recurrence during the observation period, and as false negatives if they experienced distant recurrence. Similarly, we classified patients stratified to the

high-risk group as true positives if they experienced distant recurrence, and false positive if they did not experience distant recurrence.

Sensitivity and specificity-pairs were calculated for each study and included in a bi-variate meta-analysis (Figure 3). The resulting sensitivity was 0.52 (95% CI 0.41 to

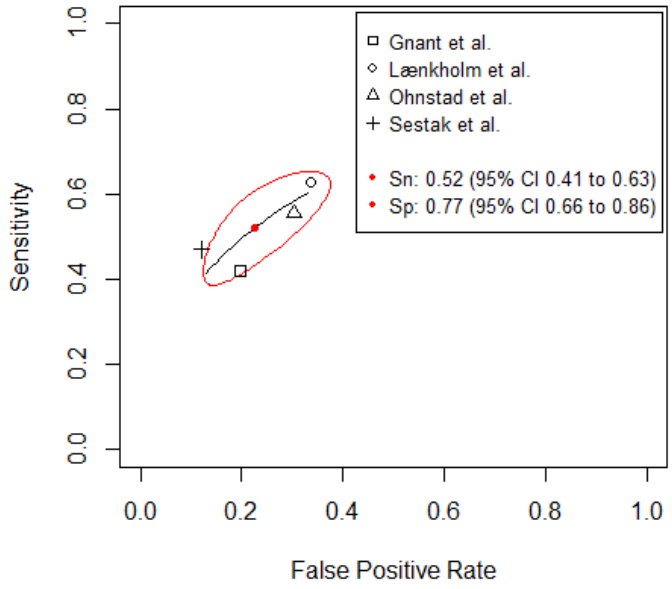



***Figure 3*** *Summary ROC resulting from the pooling of sensitivity and false positive ratios (fpr=1-specificity) from each of the four available studies. The red point is the summary sensitivity/false positive rate-point, and the red line delineates the 95% confidence region.*

0.63), and the resulting specificity was 0.77 (95% CI 0.66 to 0.86). These results suggest that Prosigna, when used as outlined above, can correctly detect a relatively large proportion of the patients who are not likely to experience distant recurrence. At the same time, a sensitivity around 50% entails that the test only detects 1 of 2 patients who will experience distant recurrence.

Since it is not obvious whether grouping the intermediate risk group with the low or high-risk group is most appropriate, we repeated the calculations with the intermediate risk group merged with the high-risk group. The resulting sensitivity was 0.83 (95% CI 0.78 to 0.87) and the specificity 0.42 (0.18 to 0.43). As expected, this implies that the prognostic accuracy depends strongly on how an "intermediate risk of recurrence" test results is interpreted, and how patients in this group are managed.

*Prognostic accuracy in absolute numbers*

This section uses the estimates from the preceding section to calculate numbers of women, out of 1000, whose risk levels would be anticipated to be incorrectly assessed. The following anticipated numbers of women are provided for illustrative

purposes, and for this reason we do not consider uncertainty (e.g., due to sampling error).

We start with a sample of 1000 women who have been diagnosed with HR+/ HER2- node negative breast cancer (Figure 4). Based on the results presented in this report (table 2) we can anticipate that about ten percent of these patients will experience a distant recurrence within ten years. We have estimated the risk of experiencing distant recurrence to around 21% for patients in the high-risk group (risk of recurrence-score > 60), and 6% for patient in the combined low/intermediate risk group (risk of recurrence-score ≤ 60). Prosigna can be expected to predict the ten-year risk of distant recurrence correctly for 745 of 1000 patients, of whom 691 patients will not experience distant recurrence, and 54 will. Among the 103 patients who will, in average, experience a distant recurrence, Prosigna will detect 54 as high-risk patients, whereas 49 will erroneously be stratified as low risk patients (Figure 4). Among the 897 who will not experience distant recurrence, Prosigna will correctly classify 691 as low risk, and 206 erroneously as high risk.

In the alternative scenario, where the intermediate risk score is grouped with the high-risk group, Prosigna predicts the ten-year risk of distant recurrence correctly for 461 of 1000 patients, of whom 378 will not experience distant recurrence, and 83 will. For the 103 patients experiencing a distant recurrence, Prosigna will detect 83 as high-risk patients, and 17 are erroneously classified as low risk. Among the 897 who will not experience recurrence, 378 will be correctly classified as low risk, but 522 will be erroneously classified as high risk.

These findings demonstrate Prosigna's performance as a standalone test and serve as an indication of the test's usefulness in clinical decision making. However, they do not tell us whether, or to what extent, adding Prosigna to the current decision-making process will lead to improved treatment decisions.

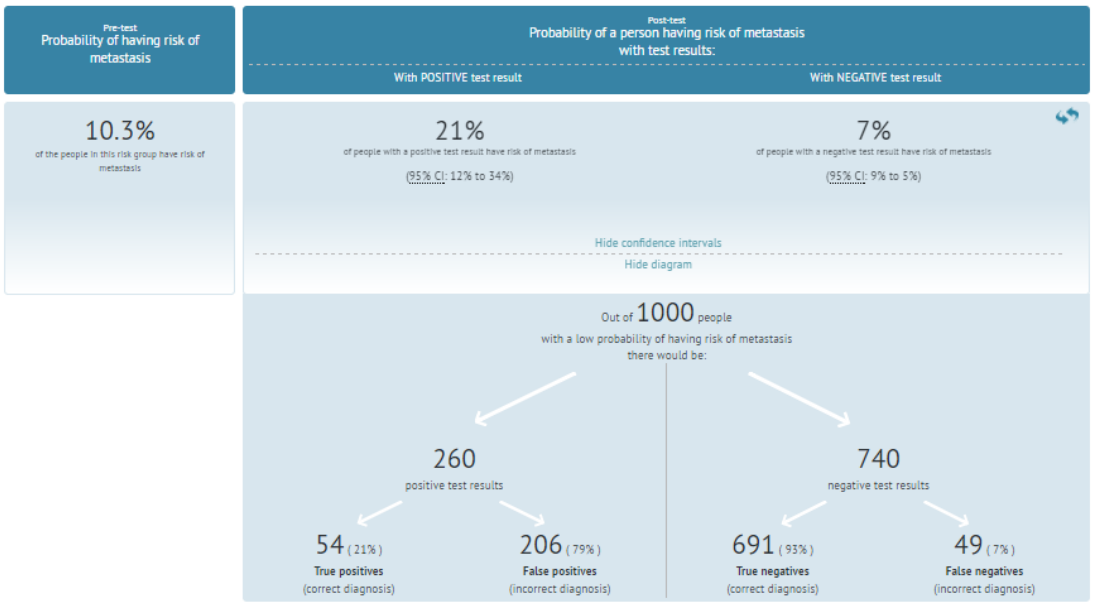



***Figure 4*** *Summary of findings table demonstrating the difference in pre-test and post-test probability of distant recurrence when Prosigna is used for risk stratification.*

### Certainty of evidence

The 10-year risk of distant recurrence in the intermediate and high-risk groups varies considerably between the four available studies, a heterogeneity that can also be recognized in Figure 3. Overall, we judge the certainty of the evidence as moderate, i.e. we believe the prognostic estimates are probably close to the true prognosis for each of these groups.

### "Decision impact"-studies

In the manufacturer's submission, three decision impact studies are presented as "evidence for decision changes" that result from making Prosigna test results available to clinicians and patients. We do not view these as providing evidence for the test's clinical utility, since the influence a test result has on clinical decision-making may, or may not, translate into improved patient care or outcomes.

However, these studies are of some relevance to the health economic assessment, since the assumed shift in chemotherapy is a key variable in the health economic model.

### Ongoing studies

Two ongoing randomised trials are mentioned in the manufacturer's submission: The UK OPTIMA-trial (Optimal Personalised Treatment of early breast cancer using Multi-parameter Analysis) with an expected primary completion date of September 2019 (16) and the EXPERT-trial (17).

These studies will probably provide valuable information for assessing the usefulness of Prosigna for clinical treatment decisions, but it is not clear if the findings will be directly applicable to Norwegian practice. In the UK OPTIMA-trial, 4,500 patients will be randomised to standard care (endocrine therapy plus chemotherapy) or test-directed therapy (endocrine therapy with the addition of chemotherapy for patients with Prosigna risk of recurrence-scores > 60), including patients in Norway. In the EXPERT-trial, 1,167 early-stage breast cancer patients identified as "Low Risk" (Prosigna risk of recurrence-score <41) will be randomised to breast irradiation or not, following breast conserving surgery.

# Method - Cost-effectiveness analysis

## Methods for evaluating submitted cost-effectiveness models

### Cost-effectiveness analysis

The primary objectives of health economic modelling are to provide a mechanism to determine the relative cost-effectiveness of the specified health intervention(s) compared to standard treatment using the best available evidence, and to assess the most important sources of uncertainty surrounding the results. In order to make comparisons across different types of treatments and multiple potential health outcomes, economic models typically measure health outcomes in terms of quality-adjusted life years (QALYs), a variable designed to capture both life extension and health improvement. QALYs, by definition, take on a value of 1 for perfect health and 0 at death. The output of a cost-effectiveness model is expressed as an incremental cost-effectiveness ratio (ICER), which can be thought of as the extra cost of obtaining an extra life-year in perfect health. The ICER is defined as

$$Cost_{Intervention} - Cost_{Comparator} \Big/ QALY_{Intervention} - QALY_{Comparator}$$

### Evaluating cost-effectiveness models

There is no single correct way to build an economic model to estimate the cost-effectiveness of a specific health intervention. Modelling requires consulting with clinical experts to gain an understanding of normal disease progression, and to determine, based on the research question, the relevant treatment population, relevant comparator; and important health outcomes and adverse events connected to treatment. This information informs the basic model structure, and also determines which clinical effect data is most important to retrieve in the systematic literature search. Once the model structure is in place, systematic searches and evidence grading are used to provide the most reliable risk information for the model, but also to collect all of the relevant cost and quality of life data that is needed for cost-effectiveness calculations.

A model is rarely meant to capture every potential detail of the treatment landscape; rather the goal is to include enough detail to provide a realistic view of the most significant pathways in disease progression, given the research question(s) one is trying to answer. Evaluating any given health economic model is primarily about determining whether the choices made by the submitter regarding model structure and treatment comparator are reasonable given the research question; whether baseline epidemiological data reflect the population in which the analysis is being performed; whether the clinical effect data used in the model are of adequate quality; whether resource use and costs reflect the conditions of the healthcare system in question; whether there has been sufficient sensitivity and scenario analyses to determine the degree and sources of uncertainty in the model results; and whether the model displays external and internal validity. Checklists are available to help researchers systematically examine these issues.

We proceed by first describing the health economic model used in the manufacture's submission and the results generated by the model. We then provide our evaluation of the model, focusing on the following issues: model structure, choice of model parameters, use of appropriate sensitivity and/or scenario analyses to examine the extent of uncertainty in model results, and relevance of the model for the Norwegian context.

## Previously published cost-effectiveness analysis

The submitter identified and provided one published cost-effectiveness evaluation of Prosigna compared with no use of the Prosigna test. The economic analysis was prepared for NICE, and examined British women with early-stage breast cancer (0-3 positive lymph nodes and ER+, HER2- tumors) who were tested for risk of recurrence with the Prosigna test (5). We searched for other published economic evaluations of Prosigna, withouth finding any relevant ones.

The health economic analysis was undertaken from the perspective of the National Health Service (NHS) and Personal Social Services (PSS) and was largely based on the model developed to inform NICE DG10 (18). They adopted a hybrid decision tree–Markov structure. The model parameters were informed by an analysis of the TransATAC trial (12), a survey disseminated by the UK Breast Cancer Group (5), the NHS England Access Scheme Database, standard costing sources, and other literature. The estimated ICERs for Prosigna test versus current practice varied from £26,058 to £91, 028 per QALY gained, across different patient subgroups (5).

## Population, intervention and comparator in the cost-effectiveness model submitted by the manufacturer

In the submitted model, patients are assumed to start at age 58, the mean age in the Oslo 1 study, as reported by Ohnstad et al. 2017 (15).

The submitted economic analysis categorizes risk of recurrence according to the results of the Prosigna test: 1) low risk of recurrence, 2) intermediate risk of recurrence and 3) high risk of recurrence.

Two groups of women with breast cancer are analyzed in the submitted cost-effectiveness model:

1) Patients classified as "HR+/ HER2-, node negative" (or "All node negative")
2) Patients classified as "Luminal B like pT1c-pT2 pN0", a subgroup constituting around 38% of the "All node negative"-group

The latter subgroup is thought to have a higher risk of recurrence than the remaining "All node negative"-group.

The Nottingham Prognostic Index is a scoring system used to determine prognosis following surgery for breast cancer. The scores are based on three criteria: tumor size; number of involved lymph nodes; and histological grade. A Nottingham Prognostic Index-score under 3.4 is considered "good", a score between 3.4 and 5.4 is considered "intermediate", and a score above 5.4 is considered "poor" (5).

## Cost-effectiveness model provided by the submitter

The submitter used the decision analytic model prepared and presented in the NICE report (5) to assess the cost-effectiveness of Prosigna compared with current practice (no test) in Norway. The model is a hybrid decision tree combined with a Markov model, built in Microsoft Excel.

The submitted decision tree includes two treatment arms: (1) Prosigna test and (2) no test. From the treatment arms, patients are classified into their risk of recurrence level based on their test result: high risk, intermediate risk or low risk.

There are two options for each risk of recurrence-level: (1) chemotherapy or (2) no chemotherapy. These branches are linked to a Markov model (models 1-6). The Markov model predicts lifetime QALYs and costs according to the patient's risk of distant recurrence and whether or not they receive chemotherapy (see Figure 5).

The submitted model captures the impact of Prosigna testing by changing the probability that patients within each risk category will receive adjuvant chemotherapy.

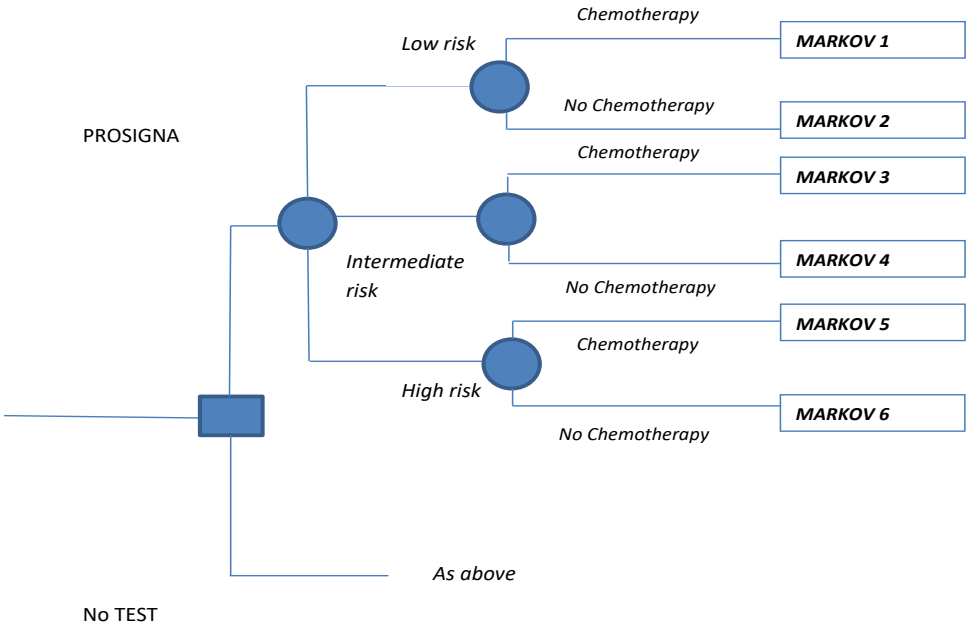



*Figure 5: Decision tree part of the submitted model (provided by the submitter)*

The submitted Markov model cohort (models 1-6) uses 6-month cycles to track patients over a lifetime, until the cohort reaches 100 years of age.

The Markov model includes four health states: (1) recurrence-free; (2) distant recurrence/metastases; (3) long-term adverse events (specified by the submitter as acute myelogenous leukaemia), and (4) dead (see figure 6). The submitter assumes that a proportion of patients who experience distant recurrence will have developed local recurrence previously, and therefore they do not explicitly model 'local recurrence' as a separate health state. Local recurrence is assumed by the submitter to be associated with additional costs and a onetime QALY loss.

Each of the six Markov models differs with respect to the patient's risk of recurrence, which is determined by their risk classification (high risk, intermediate risk or low risk) and whether or not they receive adjuvant chemotherapy.

Patients enter the model in the recurrence-free health state and during any 6-month cycle, patients who are recurrence-free can remain in their current health state, be transferred to the long-term adverse events state, develop distant metastases, or die. Patients in the distant metastases state can remain in their current health state, be transferred to the long-term adverse events (acute myelogenous leukaemia) state, or die. Patients in the long-term adverse events (acute myelogenous leukaemia) state are assumed to remain in this state until death (if free from breast cancer recurrence), and further they are assumed to not subsequently be able to develop distant

metastases due to their breast cancer. The 'death' health state incorporates both patients that die due to breast cancer, acute myelogenous leukaemia or other causes.

The submitted model applies the standard Markovian assumption that the prognosis of patients with acute myelogenous leukaemia and the costs and QALYs accrued within the acute myelogenous leukaemia-state are independent of whether the patient has previously developed distant metastases due to their breast cancer. Once a patient develops acute myelogenous leukaemia, this model assumes that this alone determines their survival prognosis.

Congestive heart failure is also a potentially relevant long-term adverse event associated with chemotherapy, however, this was excluded from the submitted model due to lack of evidence on the joint survival impact of congestive heart failure and metastatic breast cancer.

The submitter modelled the benefit of adjuvant chemotherapy by using a relative risk of distant recurrence within each risk classification group. They assumed that the relative benefit of adjuvant chemotherapy was the same across all risk score categories for all tests (the same relative risk reduction is applied to all patients, irrespective of test risk score). They also assumed disutility values associated with adverse events related to chemotherapy, and local recurrence as mentioned above.

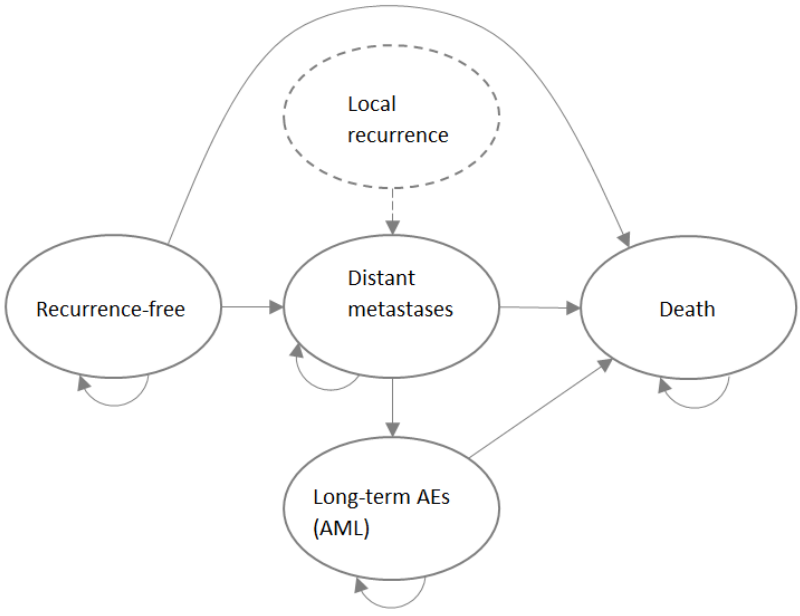



*Figure 6: Markov model provided by the submitter*

Although breast cancer is not considered a chronic disease, the submitter used a lifetime time horizon in their model, because of the risk of recurrence, which remains

over a patient's recurrence-free lifetime (except patients with acute myelogenous leukaemia). The model assumes no risk of recurrence for patients with long-term adverse events. The submitted costs and QALYs in the model were both discounted at 4% per annum, in accordance with recommendations from the Norwegian Directorate of Health (19).

## NIPH comments to the structure of the submitted model

We had access to the model built in Microsoft Excel as well as to the underlying assumptions and parameters.

In this part of the report, we consider whether the applicant by the use of a decision analytical model is able to provide convincing arguments that the test's performance translates into cost-effective resource use for Norwegian health services.

As mentioned, the health economic model consists of a decision tree that captures the short term implication of test results on treatment strategies, and a Markov model to capture life time implications of these. The Markov model appears as a standard cancer model, and seems largely well considered, uncontroversial and well performed. The driving factors in the model lie in the decision tree and how probabilities of whether or not to treat with chemotherapy are influenced by the availability of Prosigna test-results.

A challenge regarding the structure of the decision tree is that it does not model costs- and consequences based on individual test results, and therefore the model does not attempt to reflect practice related to individual clinical decisions. Prosigna is a prognostic test, and the model attempts to capture its prognostic properties compared to the current standard of care. However, in Norwegian clinical practice, the results of this prognostic test are intended to be used for individual level decision about whether or not to use chemotherapy. The decision analytical model does not convincingly capture cost- and health implications of false negative and false positive test results. Test accuracy and the implications of false test results are common in health economic analyses of medical tests. This shortcoming limits our ability to run sensitivity analyses to consider the implications of the prognostic accuracy of the test, and more generally, to consider the validity of the model projections. The submitter does not even provide a scenario analysis with alternative assumptions regarding improvement of precision in treatment decisions.

The submitted model simply divides patients into low-, intermediate- and high risk groups, and then proceeds by making assumptions at patient population level about

how treatment practice currently is without the test and how it might change after introduction of Prosigna.

While this approach represents a major simplification of clinical practice, it is also the case that most decision analytical models are simplifications of reality, and they may still produce valid estimates that are relevant for clinical decision making. The important question is therefore not how simplified this decision model is compared to reality, but rather whether it is simplified in a way that enables it to produce valid and useful results.

## Efficacy input in the economic model

The submitted model calculates distant recurrence rates conditional on information about how women are stratified into risk classifications with Prosigna, and the probabilities of receiving chemotherapy with and without the test, respectively. These three factors are discussed in more detail below.

### Risk classification probabilities

The risk classification probabilities in the submitted model are obtained directly from an analysis of the TransATAC trial provided as a part of the work undertaken by the independent assessment group for the NICE appraisal (5). The TransATAC trial included only post-menopausal women. The submitter assumes that these risk classifications can be translated to a pre-menopausal population. The population assessed in the health economic model has a start age of 58 years, which may indicate that most of the women are post-menopausal (20).

A dirichlet distribution was specified to the conditional probabilities in the submitted model to ensure that these sum to 1 (21).

Also in the no test-arm of the decision tree, patients were classified according to Prosigna test results, i.e. the test result the patients would have had if a test had been provided. For each group (high, intermediate and low risk of recurrence) estimates for the proportion of patients who would receive chemotherapy were applied.

NIPH adjusted the proportions related to each risk profile in the population group by using data from the Oslo 1 study (provided by Bjørn Naume). We considered these to be more valid for a Norwegian context. The received and adjusted input data are presented in table 3.

**Table 3: Distribution of patients across risk of recurrence groups, as defined by Prosigna**

| Risk group | All node negative | | Luminal B like pT1c-pT2 pN0 | |
|---|---|---|---|---|
| | Submitted input (N=663) (12) | NIPH[2] (N=301) (expert opinion[3]) | Submitted input (N=253) (12) | NIPH[2] (N=124) |
| Low ROR[1] | 365/663=0.55 | 132/201=0.44 | 69/253=0.27 | 0.12 |
| Intermediate ROR[1] | 196/663=0.30 | 99/301=0.33 | 96/253=0.38 | 0.50 |
| High ROR[1] | 102/663=0.15 | 69/301=0.23 | 88/253=0.35 | 0.30 |

[1]Risk of recurrence, [2]Norwegian Institute of Public Health, [3]Bjørn Naume (personal communication)

### Baseline probability of receiving adjuvant chemotherapy (standard care)

The model used by the submitter requires data on the probability of a patient receiving chemotherapy across the three Prosigna defined risk groups, for both the Prosigna test-arm, and the no test-arm.

The submitter made some assumptions regarding the probability that a patient receives adjuvant chemotherapy today based on their known tumor stage, nodal status and grade (see table 4). The submitter assumed a flat rate of chemotherapy use regardless of Prosigna risk group in both the "all node negative" group and the "Luminal B like pT1c-pT2 pN0"-subgroup. They assumed a higher rate for the Luminal B like pT1c-pT2 pN0 subgroup and a somewhat lower rate for the "all node negative" group (see tables 4 and 5).

Applying a flat rate for the probability of receiving chemotherapy across the three risk groups in the no test arm (i.e. current practice) implies an assumption that there is no association between the risk assessments that are currently undertaken without Prosigna, and the risk stratification that Prosigna yields. Consequently, the submitted model seems to compare Prosigna to a lottery for treatment decisions. We do not find this assumption believable.

Based on suggestions from our clinical experts we changed the baseline chemotherapy probabilities in the no test-arm to more trustworthy probabilities (see table 4) for both the "all node negative"-group and the subgroup of patients classified as "Luminal B like pT1-pT2" (table 5). The estimates were based on data from the Oslo-

study and recommendations in Norwegian clinical practice guidelines for breast cancer management (Bjørn Naume, personal communication – see Appendix 1).

We also performed two scenario analyses based on chemotherapy probabilities with or without Prosigna testing found in two decision impact studies: Wuerstlein et al. 2016 (22) and Martin et al. 2015 (23). In these studies, clinicians assessed whether a patient should receive chemotherapy without knowing the result of the Prosigna-test, and reassessed their decision after learning the result of the test.

The probabilities we assumed in the revised analysis are shown in tables 4 and 5, together with the submitted assumed probabilities.

**Table 4: Estimated proportion of patients receiving chemotherapy per Prosigna risk group, when the Prosigna result is not known (all node negative)**

| Risk group | Submitted estimate | NIPH[1] estimate | | |
|---|---|---|---|---|
| | Assumption | Based on data from Oslo-study[2] (NIPH[1] base case) | Wuerstlein et al. 2016 (22) (Scenario analysis 1) | Martin et al. 2015 (23) (Scenario analysis 2) |
| Low ROR[3] | 0.25 | 0.06 | 0.09 | 0.16 |
| Intermediate ROR[3] | 0.25 | 0.53 | 0.17 | 0.38 |
| High ROR[3] | 0.25 | 0.71 | 0.58 | 0.58 |

[1]Norwegian Institute of Public Health, [2]Data from Bjørn Naume (personal communication), [3]Risk of recurrence

**Table 5: Estimated proportion of patients receiving chemotherapy per Prosigna risk group, when the Prosigna result is not known (subgroup Luminal B like pT1-pT2)**

| Risk group | Submitted probabilities | NIPH[1] estimate |
|---|---|---|
| | Assumption | Based on data from Oslo-study[2] |
| Low ROR[3] | 0.55 | 0.77 |
| Intermediate ROR[3] | 0.55 | 0.68 |
| High ROR[3] | 0.55 | 0.71 |

[1]Norwegian Institute of Public Health, [2]Data from Bjørn Naume (personal communication), [3]Risk of recurrence

### Probability of receiving chemotherapy conditional on Prosigna result

The submitter claimed that there is a paucity of data for Norway on the probability that patients receive chemotherapy conditional on test results, i.e. when patients are categorized as low risk, intermediate risk or high risk based on Prosigna-results. In the absence of Norwegian data, data collected for the UK NICE appraisal, the UKBCG survey (5), were used in the submitted model.

The input data for the probability of receiving chemotherapy conditional on results of the test stem from a UK Breast Cancer Group-survey (5). In the survey, members of the UK Breast Cancer Group were asked to estimate the proportion of patients they believed would receive adjuvant chemotherapy based on the results of various risk stratification tests. Only eleven oncologists completed the questionnaire, and their averaged responses were utilised in the model.

Probabilities were collected for three populations of women with ER-positive, HER2-negative early-stage breast cancer: (1) Node-negative Nottingham Prognostic Index <3.4, (2) Node-negative Nottingham Prognostic Index >3.4 and (3) Node-positive (not presented here).

The mean probabilities concerning Prosigna that were obtained from this small and simple survey are presented in table 6. For the "All node negative" group, a weighted average of the chemotherapy use was calculated based on 38.2% of patients being in the Nottingham Prognostic Index >3.4 group evaluable by Prosigna in transATAC (12). The submitter considered the Luminal B-like pT1c-pT2 pN0 patients as equivalent to the Nottingham Prognostic Index >3.4 subgroup.

Our clinical expert assisted with risk score estimates for the Norwegian setting. As for the estimate for the no test arm (table 6), the estimates for the Prosigna arm were based on data from the Oslo-study and recommendations in Norwegian clinical practice guidelines for breast cancer management (Bjørn Naume, personal communication). Norwegian clinical practice guidelines which include recommendations for treatment decisions when Prosigna results are available (1). We also included scenarios using alternative chemotherapy probabilities based on findings from the previously mentioned decision impact studies (22;23) (see table 6).

**Table 6: Estimated proportion of patients receiving chemotherapy per Prosigna risk group, when the Prosigna result is known (all node negative)**

| Risk group | Submitted probabilities | NIPH[1] estimates | | |
|---|---|---|---|---|
| | UKBCG[2] survery (5) | Oslo-study<br>Based on data from Oslo-study[3]<br>(NIPH[1] base case) | Wuerstlein et al. 2016 (22)<br>(Scenario analysis 1) | Martin et al. 2015 (23)<br>(Scenario analysis 2) |
| Low ROR[4] | 0.017 (0.00-0.40) | 0.01 | 0.06 | 0.03 |
| Intermediate ROR[4] | 0.28 (0.04-0.60) | 0.05 | 0.24 | 0.38 |
| High ROR[4] | 0.82 (0.44-1.00) | 0.91 | 0.93 | 0.85 |

[1]Norwegian Institute of Public Health, [2]UK Breast Cancer Group, [3]Bjørn Naume (personal communication), [4]Risk of recurrence

The estimates we used for chemotherapy use across the risk of recurrence groups for the "Luminal like pT1-pT2 pN0"-patients, when the Prosigna test result was known, were based on data from the Oslo-study and recommendations in Norwegian clinical practice guidelines for breast cancer management (Bjørn Naume, personal communication). The risk scores are presented in table 7.

**Table 7: Estimated proportion of patients receiving chemotherapy per Prosigna risk group, when the Prosigna result is known (Luminal B like pT1-pT2 pN0)**

| Risk group | Node-negative NPI>3.4[1] | NIPH[2] estimate |
|---|---|---|
| | Assumption | Based on data from Oslo-study[3] |
| Low ROR[4] | 0.04 (0.00-0.15) | 0.01 |
| Intermediate ROR[4] | 0.41 (0.10-0.75) | 0.05 |
| High ROR[4] | 0.91 (0.70-1.00) | 0.80 |

[1]Nottingham Prognostic Index > 3.4 (corresponding to Luminal B like pT1-pT2 pN0) [2] Norwegian Institute of Public Health [3]Norwegian Institute of Public Health [4]ROR: Risk of recurrence

The input data used by the submitter to estimate chemotherapy probabilities based on risk score for Prosigna, appear questionable. For "All node negative" patients, the eleven British oncologists' answers ranged from 0% to 4% for low risk patients, 4% to 60% for intermediate risk patients and 44% to 100% for high risk patients. For "Luminal B like pT1-pT2 pN0" patients, the oncologists' answers ranged from 0% to 15% for low risk, 10% to 75% for intermediate risk patients and 70% to 100% for high risk patients.

**Distant recurrence rates (10 years) per Prosigna risk group**

The submitter based the risk of recurrence on 10-year distant metastasis free interval outcomes for each test risk classification. In the manufacturer's base case model, these probabilities were derived from the data analysis of the TransATAC study. They converted the 10-year DMFI probabilities to a cumulative probability of recurrence within each risk classification category (1-DMFI) and a 6-month probability of distant recurrence assuming a constant rate (see table 8).

NIPH adjusted the input data for distant recurrence probabilities to the distant recurrence probabilities from our own meta-analysis (see earlier in this report and table 8).

**Table 8: 6-months distant recurrence probabilities by risk classification for Prosigna**

| Risk group | All node-negative | | Node-negative NPI>3.4[1] | |
|---|---|---|---|---|
| | Submitter TransATAC (12) | NIPH Meta-analysis | Submitter TransATAC (12) | NIPH TransATAC (12) |
| Low ROR[2] | 0.0007 | 0.0002 | 0.0040 | 0.0040 |
| Intermediate ROR[2] | 0.0035 | 0.0052 | 0.0113 | 0.0113 |
| High ROR[2] | 0.0223 | 0.0117 | 0.0178 | 0.0178 |

[1]Nottingham Prognostic Index > 3.4 (corresponding to Luminal B like pT1-pT2 pN0), [2]Risk of recurrence

The submitter pointed out that there is uncertainty with respect to the long-term risk of distant recurrence. Their model assumes that the risk of distant metastases between 10 and 15 years is half the risk during the preceding period (0-10 years); beyond 15-years, the risk of distant recurrence is assumed to be zero. This assumption was also applied in the UK NICE model and in the model reported by Ward et al. (18).

### 10-year relative risk of recurrence with chemotherapy versus without chemotherapy

The submitter applied identical relative risk reductions for patients receiving chemotherapy to both the "All node negative" group and the Luminal B like pT1c-pT2 pN0-subgroup because the relative benefit of chemotherapy is assumed to be the same across all risk of recurrence-groups.

The submitter derived the relative risk of recurrence for chemotherapy versus no chemotherapy from a meta-analysis (24). The relative risk for chemotherapy versus no chemotherapy was calculated based on the difference between the projected 10-year recurrence free probabilities for the two groups, a relative risk of 0.76. Six month probabilities of recurrence for patients receiving chemotherapy were then calculated assuming a constant rate.

### Risk of death following distant recurrence

The submitter based the survival prognosis of patients with distant metastases on an analysis of complete hospital and community records for 77 women randomly selected from 232 women who had breast cancer recurrence between 2000 and 2005 (25). Median survival was reported to be 40.1 months following distant recurrence. The 6-month probability of death was estimated by fitting an exponential distribution with a median of 40.1 months. Based on this approach, the 6-month probability of death following distant recurrence was estimated to be 0.098, assuming a constant rate. The submitted model assumes that the rate of death due to distant metastases is constant across the different model subgroups and across each test risk classification group due to a lack of population or risk group specific data. We adopted the same estimate in our revised model.

### Probability of local recurrence

The submitted model assumes that 10.5% of patients entering the distant recurrence health state have previously experienced a local recurrence. The submitter based this estimate on a study by de Bock et al. (26).

Our clinical expert considered this assumption as plausible also for the Norwegian clinical context.

### Probability of acute myeloid leukaemia

The submitter obtained the probability of developing acute myelogenous leukaemia following chemotherapy from an analysis of 20,063 patients with Stage I-III breast cancer treated at US academic centres between 1998 and 2007 (27). The estimated 10-year risk of developing acute myelogenous leukaemia was reported to be 0.0049

which translates into a 6-month probability of developing acute myelogenous leukaemia of 0.00025, assuming a constant event rate.

### Probability of death following onset of acute myelogenous leukaemia

The submitter estimated the mean survival following the onset of acute myelogenous leukaemia from the UK NICE STA of azacitidine for myelodysplastic syndromes (MDS) (28) to approximately 8 months. With this input, the 6-month probability of death following acute myelogenous leukaemia was estimated to be 0.53 when a constant event rate was assumed.

### Other-cause mortality (life tables)

The submitter included all-cause mortality in the model as well as mortality due to recurrence and acute myelogenous leukaemia. This was taken from Statistics Norway, 2017 (29). Probabilities of death are age specific and the probabilities for females are applied in the model.

**Table 9: Summary of 6 months' probabilities used in the revised model**

| Variable | Probability | Source (Reference) |
|---|---|---|
| Death following distant recurrence | 0.0980 | Thomas et al. (25) |
| AML[1] | 0.0003 | Wolff et al. (27) |
| Death following onset of AML[1] | 0.5300 | NICE et al. (28) |
| Base-line all-cause mortality | 0.0014 | Statistics Norway (29) |

[1]Acute myelogenous leukaemia

## Cost and resource input in the submitted economic model

### Costs of Prosigna test

The cost of the Prosigna test was provided by the manufacturer, Nanostring, and is assumed to be NOK 16,254. This assumption takes into account that the test will be conducted in a publicly funded laboratory, and includes costs associated with the test kit, instrument rental, and labour (pathologist and technician).

### Costs of adjuvant chemotherapy acquisition and administration

The submitted costs associated with adjuvant chemotherapy were obtained from Norwegian clinical opinion. The Norwegian clinical expert advised that 50% of women would receive 4x EC90+taxane and 50% 4x EC90 alone as the preferred chemotherapy regimen.

EC90 alone constitutes: Epirubicin 90 mg/m2 + cyclophosphamide 600 mg/m2 i.v. every 3 weeks x 4, Neutral support 1-3 days after each course.

EC90 + taxane constitutes: Epirubicin 90 mg / m2 + cyclophosphamide 600 mg / m2 i.v. every 3 weeks x 4, followed by paclitaxel 80 mg / m2 weekly x 12, Neutral support 1-3 days after each course.)

Societal costs related to nausea and wigs, which are caused by adjuvant chemotherapy, were also included in the submitted model.

Our clinical expert confirmed the submitted information. The unit costs are listed in the Appendix 2, table 1 and 2.

**Costs of endocrine therapy**

The submitted model assumes that 50% of surviving patients receive endocrine therapy for a period of between 5 and 8 years. This is consistent with the proportion of women who received endocrine therapy in the Norwegian study (15), and our clinical expert agrees that this is a reasonable assumption.

The costs associated with endocrine therapy were based on the assumptions employed within an economic analysis reported by Ward et al. (18). The model assumes that patients receive one of four possible endocrine therapy regimens: (1) tamoxifen for 5 years; (2) anastrozole for 5 years; (3) letrozole for 5 years or (4) tamoxifen for 2 years followed by exemestane for 3 years. The proportion of patients receiving each regimen was also taken from Ward et al. (18) (tamoxifen – 40%; anastrozole – 20%; letrozole – 20%; tamoxifen then exemestane 20%). Ten per cent of patients are also assumed to receive extended letrozole for 3 further years (years 6-8).

The submitted unit costs were estimated to be NOK 1,256 for year 0-2, NOK 1,207 for year 2-5 and NOK 201 for year 5-8. The submitted total cost was estimated to be NOK 13,468 for endocrine therapy.

We made some changes in line with opinions of our clinical expert: We decreased the proportion of patients receiving tamoxifen from 40% to 10%, because this regime is primarily given to premenopausal women. We also adjusted the number of years related to the tamoxifen regime from 5 years to 10 years. Further, we increased the proportion of women receiving letrozole from 20% to 50%, since this is the most used regime for endocrine therapy in Norway (see Appendix 2, table 3 for unit cost data).

Our unit costs were estimated to be NOK 2,239 for year 0-2, NOK 3,866 for year 2-5, NOK 325 for year 5-6 and NOK 62 for year 6-10. The submitted total cost was estimated to be NOK 33,302 for endocrine therapy.

### Costs of additional treatments

The submitted model assumes that 30% of women with early-stage breast cancer will receive 4 mg bisphosphonates (zoledronic acid) every 6 months by i.v. infusion for up to 3 years. In addition 14% of women (30) receiving chemotherapy are assumed to receive three cycles of G-CSF (1 vial per day).

We also adjusted the numbers of years related to the zoledronic acid from 3 to 10 years, based on advice from our clinical expert. The unit costs are listed in Appendix 2, table 4.

### Follow-up costs

The submitted model assumes that all patients receive two routine follow-up visits during the first year following surgery, with annual visits thereafter for a period of 5 years. Patients are also assumed to undergo a routine annual mammogram for up to 5 years.

### Costs of treatments for local and distant recurrence

The submitted costs associated with treating local recurrence are applied as a once-only cost. The submitted costs were taken from a UK-based patient-level costing analysis of breast cancer recurrence (31). The mean cost was calculated to be about NOK 153,213.

The submitted costs associated with treating distant recurrence included those associated with visits, drugs, pharmacy, hospital admission and intervention, imaging, radiotherapy, pathology and transport. These were taken from a study reported by Thomas et al. (25). The 6-monthly mean cost of treating metastatic breast cancer was assumed to be NOK 50,006.

The submitted costs associated with acute myelogenous leukaemia are also assigned as a one-off cost. These are taken from Ward, et al. (18) and converted to a mean cost of NOK 140,041.

**Table 10. Calculated cost in the model**

| Cost variable | Quantity | Submitted total cost | NIPH total cost |
|---|---|---|---|
| Prosigna test | 1 | NOK 16,254 | NOK 16,254 (Nanostring) |
| Adjuvant chemotherapy acquisition and administration | 0.5<br>0.5 | NOK 37,028[1]<br>NOK 53,211[2] | NOK 47,582* (Appendix 2)<br>NOK 77,482 ** |
| Endocrine therapy | Per cycle (6 months) | NOK 1,256 | NOK 2,238 (Appendix 2) |
| **Additional treatment costs** | | | |
| Course of 4mg biphosphonates | Per cycle (6 months) | NOK 2,258 | NOK 2,595 (Appendix 2) |
| G-CSF (for patients receiving chemotherapy only) | 1 | NOK 16,603 | NOK 11,961 (Appendix 2) |
| **Follow up costs** | | | |
| Routine visits | 1 | NOK 1,761 | 1,761 (18) |
| Mammogram | 1 | NOK 548 | NOK 1,530 (32) |
| **Treatments of local and distant recurrence** | | | |
| Local recurrence | 1 | NOK 153,213 | NOK 153,213 (31) |
| Distant recurrence | Per cycle (6 months) | NOK 50,006 | NOK 50,006 (25) |
| AML | 1 | NOK 140,041 | NOK 140,041 (18) |

[1]EC90, [2]EC90 + Paclataxel

### Investment cost related to Prosigna test

The submitter did not include the investment cost related to nCounter, which is the analysis system required for using the Prosigna test (33). Our clinical expert estimated the price of the system to NOK 2,600,000 plus a service fee at 10% of the system price per year. Each of the four Norwegian regional health authorities are offered to buy the nCounter analysis system. We included this cost in our budget impact analysis.

## Utility input in the economic model

The submitted utility data were conducted in 2017 to identify the most appropriate utility values for the NICE model (5). Health-related quality of life (HRQL) utility values, based on the instrument EQ-5D, were available for all health states.

The submitter applied different health utilities to each of the modelled health states: recurrence free, distant recurrence and acute myelogenous leukaemia (see table 11).

A decrement in a patient's utility was applied during the first model cycle (6-months) for patients receiving adjuvant chemotherapy, to account for health losses associated with short-term chemotherapy-related adverse events (34).

A decrement in a patient's utility associated with local recurrence was taken from a published model of first, second, and third generation adjuvant chemotherapy regimens for breast cancer reported by Campbell et al. (34).

An additional one-time HRQoL gain of 0.0177 is added to account for the additional value of risk profiling knowledge to patients (35).

The submitted QALYs were calculated by multiplying the time spent in a particular health state by the utility associated with that health state; area under the curve method.

**Table 11: Summary of utility values**

| Health situation | Type of utility | HRQoL[1]-weighting | SE[2] | Time horizon | Reference |
|---|---|---|---|---|---|
| Recurrence free | Health state utility | 0.824 | 0.018 | Indefinite | (36) |
| Distant recurrence | Health state utility | 0.685 | 0.019 | Indefinite | (36) |
| Acute myeloid leukaemia | Health state utility | 0.265 | 0.04* | Indefinite | (37) |
| Chemotherapy AE[3] | Disutility | 0.038 | 0.04* | 6-months | (34) |
| Local recurrence | Disutility | 0.108 | 0.04* | One-time QALY loss applied on transition to distant recurrence state | (34) |
| Benefit risk classification (Prosigna test) | Utility | 0.0177 | | Indefinite | (35) |

[1]Health related quality of life, [2]Standard error, [3]Adverse events;

## NIPH comments to input data provided by the submitter

There are uncertainties associated with the efficacy data, especially related to chemotherapy use in the two arms (no test and Prosigna).

The expert opinions regarding treatment practices are crude and not based on any form of empirical evidence. This lack of stringency is particularly problematic, since assumptions about treatment practice are by far the most important driver of the results of the model. Moreover, the model is not well suited to test the implications of these assumptions in sensitivity analyses.

Another big uncertainty concerns the assumed utility value for the additional value of risk profiling knowledge to patients. The model attaches a utility gain of 0.0177 for this, and this assumption is influential for the size of the incremental QALYs and the incremental cost-effectiveness of Prosigna.

There are at least two technical and two principal problems associated with this assumption. The first technical problem is that the submission states that the time horizon of this value is "Indefinite", meaning that patients who learn the results of their prognostic gene expression test are assumed to experience an improvement in their health related quality of life of 0.0177 per year for the rest of their lives. In reality, this effect was counted only once per patient receiving the test. Second, the calculation is based on a contingent valuation method (35) were patients were asked about their willingness to pay out of pocket to receive the gene expression test. This contingent of $997 (NOK 8 081) was then divided with an assumed willingness to pay threshold of NOK 750 000, resulting in the estimated annual QALY-improvement of 0.01077.

The first principal problem with the utility of knowledge parameter is related to the White paper on Norwegian health priorities, which defines health utilities to include direct and sometimes indirect health utilities. The white paper also states that utilities at group level are to be quantified using "good years of life", or QALYs, in order to capture both life extension and improvements in health related quality of life. We do not consider a preference for knowing the test result "health related" as such, and therefore consider this utility irrelevant for priority decisions. Secondly, to the extent that the submitter would present a fair argument that this utility value captures real health benefits in addition to knowledge, the inclusion would represent double counting. The reason is that health benefits would be captured first through direct health benefit estimation, and then again through the contingent valuation of the same perceived health benefits. Its inclusion would therefore be methodologically flawed, irrespective of white paper interpretations. When the women consider their willingness to pay, they obviously consider the perceived future health benefits related to more precise treatment – not just the value of learning the test result per se. The direct benefit estimation includes improved recurrence free survival, reduced distant metastases, local recurrence and reduced adverse events from redundant chemotherapy. It is inappropriate to double count the same effects a second time through contingent valuation.

# Results - Cost-effectiveness

## Base-case cost-effectiveness results

### Base-case cost-effectiveness results by submitter

The submitter provides cost-effectiveness results comparing Prosigna to current practice (no test) for the populations HR+/ HER2-, node negative ("node negative all") and the subgroup "Luminal B like pT1c-pT2 pN0".

The submitter's results show that the use of Prosigna for women in "All node negative"- group gives somewhat higher QALY and higher cost, resulting in a cost-effective ICER (see table 12).

**Table 12: Cost-effectiveness results (All node negative) of Prosigna versus usual care according to submitter's model**

| Measure: | Total costs (NOK[1]) | Incremental costs | Total number of QALYs[2] | Incremental effectiveness* | ICER[3] |
|---|---|---|---|---|---|
| Prosigna | 186,648 | 19,445 | 12.83 | 0.06 | 295,012 |
| Current care | 167,203 | | 12.76 | | - |

[1]Norwegian Kroner, [2]Quality adjusted life year, [3]Incremental cost effectiveness ratio

The submitter's result shows that the use of Prosigna for women in "Luminal B like pT1c-pT2 pN0"-subgroup gives somewhat lower absolute and incremental QALYs and a little lower absolute and incremental costs, resulting in a more cost-effective ICER (see table 13).

**Table 13: Cost-effectiveness results (Luminal B like pT1c-pT2 pN0) of Prosigna versus usual care according to submitter's model**

| Measure: | Total costs (NOK[1]) | Incremental costs | Total number of QALYs[2] | Incremental effectiveness* | ICER[3] |
|---|---|---|---|---|---|
| Prosigna | 139,369 | 11,505 | 12.09 | 0.05 | 224,161 |
| Current care | 127,864 | | 12.04 | | - |

[1]Norwegian Kroner, [2]Quality adjusted life year, [3]Incremental cost effectiveness ratio

### Base-case cost-effectiveness and scenario results by NIPH

We made some changes to the base case analysis to investigate the effect of parameters of interest in line with our clinical expert opinions. These include: The probability of death following distant recurrence, the resource use and drug prices of adjuvant chemotherapy and administration, the proportions of patients receiving different endocrine therapies and resource use affecting the endocrine therapies, and the resource use related to additional treatments and following costs. We also adjusted the cost related to distant metastases and modified input data related to the patients' risk classifications.

Further, we used the model with revised input data to assess the cost-effectiveness of Prosigna compared with current practice (no test).

We preformed two additional scenario analyses based on different use of chemotherapy in current practice and after introducing Prosigna for the "all node negative"-group, as shown in table 14.

Our base-case anaysis shows that, compared to the submitter's model, the use of Prosigna for women in the "All node negative"-group gives somewhat higher QALY, but at a much higher cost, resulting in an ICER of NOK 897,923 per QALY gain. Our alternative scenario analyses for "All node negative"-patients yielded higher QALY-gains at a somewhat higher cost, and resulted in substantially more cost-effective results (see table 14).

These cost-effectiveness estimates demonstrate that assumptions about shifts in the use of chemotherapy are crucial when comparing the Prosigna test with current practice (no test).

**Table 14: NIPH's base-case analysis and scenario analyses for "all node negative"-patients with different chemotherapy probabilities**

| Chemotherapy use | Prosigna | | Current practise | | Incr. cost (NOK[1]) | Incr. eff. (QALY[2] gained) | ICER[3] |
|---|---|---|---|---|---|---|---|
| | Total cost (NOK) | Total QALYs | Total cost (NOK) | Total QALYs | | | |
| Base-case: Clinical expert, Oslo 1 study | 237,517 | 12.641 | 214,171 | 12.615 | 23,346 | 0.026 | 897,923 |
| Wuestlein et al. 2016 (22) | 226,410 | 12.663 | 216,694 | 12.614 | 9,717 | 0.056 | 173,518 |
| Martin et al. 2015 (23) | 225,067 | 12.659 | 208,196 | 12,612 | 16,871 | 0,047 | 358,957 |

[1]Norwegian Kroner, [2]Quality adjusted life year, [3]Incremental cost effectiveness ratio

We performed one base-case analysis for the "Luminal B like pT1c-pT2 pNo"-sub-group based on chemotherapy data from the Oslo-study and recommendations in Norwegian clinical practice guidelines for breast cancer management (see table 15).

Our result show that the use of Prosigna for women in "Luminal B like pT1c-pT2 pNo"- subgroup gives an ICER of NOK 98,188 per QALY gain (see table 15).

**Table 15: NIPH[1] base-case analysis for "Luminal B like pT1c-pT2 pN0"-subgroup**

| Chemotherapy (Current practice) | Prosigna | | Current practise | | Incr. cost (NOK[2]) | Incr. eff. (QALY[3] gained) | ICER[4] |
|---|---|---|---|---|---|---|---|
| | Total cost (NOK) | Total QALYs | Total cost (NOK) | Total QALYs | | | |
| Base-case: | 168,578 | 11.87 | 175,343 | 11.94 | -6,775 | -0,069 | 98,188 |

[1]Norwegian Institute of Public Health, [2]Norwegian kroner, [3]Quality adjusted life year, [4]Incremental cost effectiveness ratio

## Severity considerations – Absolute shortfall

The calculation of absolute shortfall (AS) is based on the submission guideline of the Norwegian Medicines Agency (38) which is based on the white paper on priority setting (39), a Norwegian life table (29) and health related quality of life information from a Swedish population (40). Absolute shortfall is defined as the difference in quality adjusted life expectancies at age (A) without the disease ($QALY_{sA}$) and prognosis with the disease ($P_A$):

$$AS = QALY_{sA} - P_A$$

In accordance with the economic model, we first assume that patients are 58 years of age when entering the model. At this age, the expected quality adjusted life expectancy is 20.7. The prognosis with disease expected to be 12.76 QALYs for the usual care (no test) is based on simulations of our base case analysis. The absolute shortfall with these assumptions is:

$$AS = 20.7 - 12.34 = \underline{8.36 \text{ QALYs}}$$

According to the white paper (39), the cost-effectiveness threshold should be weighted according to severity classes suggested by the Norheim and Magnussen commissions. It was suggested that AS falling below 4 QALYs belong to the least severe group, and AS being above 20 QALYs are to be considered among the highest severity diseases. With AS of 8.36, the argument for giving special priority to Prosigna based on severity appears moderate.

## Budget impact analysis

### Budget impact analyses by submitter

The submitter calculated the budget impact, from a Norwegian health care perspective, of applying Prosigna to women with HR+/ HER2- early-stage breast cancer with 0-3 nodes. The budget impact is estimated as the net cost difference between a scenario in which the Prosigna is adopted for a full cohort of eligible individuals relative to a scenario in which the test is not adopted. The submitted budget impact was estimated over a 5-year time horizon. The model is based on the Norwegian breast cancer incidence rates and general population data (41). The submitter assumed an invasive breast cancer incidence of 125.1 per 100,000 women to be stage I/II (42;43), 92% HR+ (44) and 82% HER2- (44). 82% of these patients are assumed to be node-negative (45) leading to a population of 1,378 patients in 2015.

Further, the submitter considered the Luminal B-like pT1c-pT2 pN0 patients as equivalent to the Nottingham Prognostic Index >3.4 subgroup. Hence, 38.2% of the node negative patients were estimated to be Luminal B-like pT1c-pT2 pN0 based on the proportion of patients in the Nottingham Prognostic Index >3.4 subgroup from transATAC, leading to an estimated 526 patients in this subgroup. The submitter assumed these numbers to be similar for 2018.

**Table 16: Submitter's total eligible population (women with HR+/ HER2- early-stage breast cancer with 0-3 nodes)**

| | Year | | | | |
|---|---|---|---|---|---|
| | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 |
| Growth | 1.14% | 1.14% | 1.14% | 1.14% | 1.14% |
| All NO[1] | 1,378 | 1,391 | 1,409 | 1,425 | 1,445 |
| Luminal B…[2] | 526 | 532 | 538 | 544 | 551 |

[1]All node negative-group, [2]Subgroup Luminal B like pT1c-pT2 pNO

Further, the submitter assumed that the test uptake among eligible patients would be 11% in year 1, 31% in year 2, 51% in year 3, 54% in year 4 and 59% in year 5. The number of eligible patients tested are presented in table 17.

The submitter estimated that a total of 146 and 56 patients would be tested in year 1 in all node negative and Luminal B like pT1c-pT2 pN0 subgroup respectively. By Year 5, a total of 851 and 325 patients are expected to be tested in respective groups (see table 17).

**Table 17: Submitted number of eligible patients tested**

| | Year | | | | |
|---|---|---|---|---|---|
| | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 |
| All NO[1] | 146 | 426 | 718 | 772 | 851 |
| Luminal B…[2] | 56 | 163 | 274 | 295 | 325 |

[1]All node negative-group, [2]Subgroup Luminal B like pT1c-pT2 pNO

The submitter created a budget impact scenario analyses in which Prosigna is compared to current practise (no test), as described in the cost-effectiveness analysis.

In the budget calculation the costs related to Prosigna testing, chemotherapy use and local/distant recurrence were considered.

The submitted net budget associated with introducing Prosigna in all node negative patients was estimated at NOK 2.2 million in Year 1, and NOK 11.7 million in year 5, with a cumulative budget impact of NOK 41.7 million over 5 years.

**Table 18. The submitted total budget impact (NOK[1]) over a 5-year time horizon for the "Node negative all" group (based on the submitted input data).**

| | Year | | | | | Total |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 5-year total |
| All NO[2] | 146 | 425 | 718 | 772 | 851 | 2,912 |
| Cost of testing | 2,378,669 | 6,919,765 | 11,677,103 | 12,542,529 | 13,839,529 | 47,357,139 |
| Cost of chemo[3] | -224,218 | -652,269 | -1,100,705 | -1,182,238 | -1,304,539 | -4,463,969 |
| Cost or rec[4] | -3,123 | -23,080 | -101,502 | -261,808 | -836,897 | -1,226,410 |
| Total budget Impact | 2,151,329 | 6,244,415 | 10,474,896 | 11,098,027 | 11,698,093 | 41,666,759 |

[1]Norwegian Kroner; [2]All node negative patients, [3]Chemotherapy, [4]Recurrence

The submitted net budget associated with introducing Prosigna to patients in the Luminal B like pT1c-pT2 pN0-subgroup were estimated at NOK 732,591 in Year 1, and would give cost savings of NOK 3.2 million in year 5, with a cumulative net saving of NOK 15 million over 5 years.

**Table 19. The submitted total budget impact (NOK[1]) over a 5-year time horizon for the Luminal B like pT1c-pT2 pN0 subgroup (based on the submitted input data).**

| | Year | | | | | Total |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 5-year total |
| Luminal B…[2] | 56 | 163 | 274 | 296 | 325 | 1,114 |
| Cost of testing | 908,652 | 2,643,350 | 4,460,653 | 4,791,072 | 5,286,700 | 18,090,427 |
| Cost of chemo[3] | -175,152 | -7,913,400 | -8,172,074 | -8,291,589 | -8,427,683 | -32,979,898 |
| Cost or rec[4] | -908 | -6,608 | -22,122 | -49,681 | -86,901 | -166,220 |
| Total budget Impact | 732,591 | -5,276,657 | -3,733,544 | -3,550,197 | -3,227,884 | -15,055,691 |

[1]Norwegian Kroner; [2]Luminal B like pT1c-pT2 pNO, [3]Chemotherapy, [4]Recurrence

### Budget impact analyses by NIPH

The submitted budget impact models were based on the same input data as the cost-effectiveness analyses. We made some changes in the input data in the submitted base-case cost-effectiveness analysis which affected the results of the budget impact analysis.

In addition to considering the costs related to Prosigna testing, chemotherapy use and local/distant recurrence in the budget calculation, we also included costs related to investing in the nCounter system.

Finally, with regard to our clinical expert opinions, we assumed that the test uptake among eligible patients would be a constant rate, were all of the patients in the "Node negative all" group or in the "Luminal B like pT1c-pT2 pN0 subgroup" would be tested during the five years (see tables 20 and 21).

NIPH's net budget associated with introducing Prosigna in all node negative patients was estimated at NOK 16.1 million in Year 1, and NOK 13.5 million in year 5, with a cumulative budget impact of NOK 75 million over 5 years.

**Table 20. NIPS's total budget impact (NOK[1]) over a 5-year time horizon for the "node negative all" group**

| | Year | | | | | Total |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 5-year total |
| Number of patients | 1,378 | 1,391 | 1,409 | 1,425 | 1,442 | 7,047 |
| Cost of testing | 22,392,397 | 22,647,670 | 22,905,854 | 23,166,980 | 23,431,084 | 114,543,986 |
| Cost of chemo[2] | -9,137,164 | -9,241,327 | -9,346,679 | -9,453,231 | -9,560,998 | -46,739,398 |
| Cost or rec[3] | -5,920 | -31,396 | -81,801 | -140,337 | -647,252 | -906,806 |
| Cost of nCounter | 2,860,000 | 260,000 | 260,000 | 260,000 | 260,000 | 3,900,000 |
| Total budget impact | 16,109,313 | 13,634,947 | 13,737,374 | 13,833,413 | 13,482,835 | 74,697,882 |

[1]Norwegian Kroner; [2]Chemotherapy, [3]Recurrence

Our net budget associated with introducing Prosigna to patients in the Luminal B like pT1c-pT2 pN0-subgroup was estimated at - NOK 3,558,474 in Year 1, and - NOK 9,949,531 in year 5, with a cumulative –NOK 40,060,066 over 5 years.

**Table 21. NIPH[1] total budget impact (NOK[2]) over a 5-year time horizon for the Luminal B like pT1c-pT2 pN0-subgroup**

| | Year | | | | | Total |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 5-year total |
| Number of patient | 526 | 532 | 538 | 544 | 551 | 2692 |
| Cost of testing | 8,553,896 | 8,651,410 | 8,750,036 | 8,849,787 | 8,950,674 | 43,755,803 |
| Cost of chemo[3]. | -14,993,389 | -19,240,742 | -19,460,087 | -19,681,932 | -19,906,306 | -93,282,457 |
| Cost or rec[¤] | 21,019 | 120,588 | 283,334 | 495,546 | 746,101 | 1,666,589 |
| Cost of nCounter | 2,860,000 | 260,000 | 260,000 | 260,000 | 260,000 | 3,640,000 |
| Total budget impact | -3,558,474 | -10,208,744 | -10,166,716 | -10,076,600 | -9,949,531 | -40,060,066 |

[1]Norwegian Institute of Public Health, [2]Norwegian Kroner, [3]Chemotherapy, [¤]Recurrence

# Discussion

## Summary of results

This is a summary of our results for the use of Prosigna™ Breast Cancer Prognostic Gene Signature Assay for patients with HR+/ HER2- node negative breast cancer.

### Prognostic accuracy and clinical effectiveness

The Prosigna test places a patient into one of three risk of recurrence-categories: low, intermediate and high. There is convincing evidence of an association between the risk of recurrence and the risk stratification generated by the Prosigna test. For patients classified as low risk, the risk of recurrence is around 4%, over 10 years. For the intermediate risk group, the risk is around 10%, and for the high-risk group around 21%.

The performance of the test can also be expressed in terms of prognostic sensitivity and specificity. When we merged the intermediate risk group with the low risk group, we estimated the test's sensitivity to 52%, and its specificity to 72%. Alternatively, the intermediate group can be merged with the high-risk group. This yields a sensitivity of 83% and a specificity of 42%. Both scenarios entail a substantial proportion false positive and false negative test results. A large proportion of false positive test implies that many patients will receive unnecessary chemotherapy. People who receive false negative test results are at risk of not receiving chemotherapy even if chemotherapy would have improved their prognosis. Depending on life situation, values and preferences, it seems likely that individual patients will put different emphasis on the risk of false positive (unneeded chemotherapy) and false negative (effective treatment is withheld) test results. Such expectations of differences in patient preferences invoke the need for shared decision making. It can therefore be contested whether more standardised treatment decisions, which is one argument for introducing the Prosigna test, is a reasonable goal.

These findings describe the performance of Prosigna as a standalone prognostic tool. However, Prosigna is more likely to be used as supplementary test alongside other prognostic tools. As we were not able to identify comparative studies where patients have been allocated to assessment with or without the Prosigna test, we were not able to estimate the clinical utility of Prosigna.

Some studies have attempted to estimate the prognostic value of adding Prosigna to other prognostic tools by means of multivariate regression analyses. Such studies have shown that the test adds prognostic information that may be useful for clinical decision making. The Prosigna manufacturer did not present analyses based on these data, and we were not able to use these results to estimate an expected health gain from introducing Prosigna testing in clinical practice.

### Cost effectiveness

We were not able to estimate a trustworthy cost-effectiveness estimate using the submitter's model, the main reason being that we do not have reliable data on chemotherapy use among patients who have or have not undergone Prosigna testing, for the different Prosigna risk categories. These data are key variables in the submitted model. Our best estimate using the submitter's model, indicates that for the cost effectiveness for the node-negative groups as a whole, is an incremental cost effectiveness ratio of NOK 897,923 per QALY gained.

We also have concerns about the model as such, as it does not incorporate shifts in the proportions of false/true negative and false/true positive prognostic assessments that may result from adding the Prosigna test to the current approach – see discussion below.

## Methodological challenges

There is no standard approach to assessing the clinical utility of prognostic tools (46). While a simple estimation of the risk of recurrence can demonstrate a convincing association between the risk group a patient is placed in by a test and the risk of recurrence, we cannot automatically assume that the categorization will lead to changes in clinical management. And even if the addition of a test to the assessment of a patient's prognosis leads to changes clinical management, we cannot automatically assume that the changes will lead to improved patient outcomes (47).

Consider a group of 100 patients where 53 would receive chemotherapy without the Prosigna test, and this proportion would be reduced to 5 of 100 if Prosigna-testing was included in the decision-making process. The risk of recurrence in the group as

a whole is, say, 10%, thus the goal of the prognostic process is to identify, to the extent possible, the 10 patients that will experience a recurrence, and the 90 who will be unnecessarily burdened if they receive chemotherapy. As pointed out by others, the problem with the approach used in the current report is that we are unable to estimate "how many of the changes in therapy are correct and how many are wrong" (47). How many of the 10 patients that will experience recurrence are among the 53 who received treatment without a Prosigna-result? And how many of the 10 patients are among the 5 who will receive treatment with the Prosigna-result? These questions are important but cannot be answered based on the current evidence base.

It can be argued that in the assessment of prognostic test performance we should employ similar standards as for diagnostic tests, e.g. that sensitivity and specificity should be estimated and included in the economic model: "one cannot automatically assume that all predictions are correct, but rather the decision analysis should use an enumeration of correct and erroneous predictions and their consequences in order to fully estimate the ICER associated with the test" (47). We are uncertain about the consequences of using an economic model that does not incorporate sensitivity and specificity when estimating Prosigna's cost-effectiveness. If we were to prepare a health economic model for a prognostic test such as Prosigna, we would probably have opted for a different approach, and incorporated the test's prognostic sensitivity and specificity into the model.

The key question is whether the test leads to improved patient outcomes at an acceptable cost, or to cost savings without poorer patient outcomes. In order to assess this reliably, randomised controlled trials are needed, where the use of Prosigna is compared to not using the test. Such trials are often considered costly and lengthy, and are rarely done for prognostic tests (48). Still, as others have pointed out: "tests can do as much harm and as much good as drugs or devices; thus, a rigorous appraisal of their clinical utility, including both the possible benefits and the possible harms, is necessary" (48).

**Assessments conducted by others**

We identified one HTA of a gene expression signature test, prepared by the European Network for Health Technology Assessment (49). Their initial aim was to compare the available tests on the market, but the only test they identified that had been evaluated in a randomized trial was Mammaprint (Agendia, Netherlands), they only assessed that test.

In the report commissioned by The National Institute for Health and Care Excellence (NICE) in England, several gene expression panel tests were assessed, including Prosigna (5). The authors found that for lymph node negative patients, "All tests

provided additional prognostic information over most commonly used clinicopathological factors and over clinical treatment score (CTS) and Nottingham Prognostic Index". They did not assess test performance in terms of sensitivity or specificity, or clinical effectiveness in terms of patient outcomes. The manufacturer's submission is largely based on this report. For their cost effectiveness analysis, the report authors used responses from 11 oncologists who responded to a survey circulated to members of the UK Breast Cancer Group. The oncologists were asked to subjectively estimate the probability that a woman with given risk assessment (including Prosigna) would go on to receive adjuvant chemotherapy. The averaged responses were used in the health economic model, to reflect the test's clinical utility. For the corresponding probabilities for patients not assessed with Prosigna, the authors opted for a flat rate across the three Prosigna-defined risk groups: "Within the base case analysis, the proportion of patients who receive chemotherapy under current practice (no test) is assumed to be the same for each test risk classification (low-, intermediate- and high-risk)" (Manufacturer's submission, 2018). The authors further reported the cost-effectiveness of Prosigna for different risk groups, as defined by the Nottingham Prognostic Index. They found that for lymph node negative patients with a low Nottingham Prognostic Index-score (less than 3,4) the ICER per QALY gained was £91,000, while for lymph node negative patients with a Nottingham Prognostic Index-score above 3,4 the ICER per QALY was £26,000.

The model submitted by the manufacturer is structurally the same as the one used in the NICE-commissioned report.

## Discussion of cost-effectiveness

There are some major uncertainties related to the cost-effectiveness analysis.

First of all, from a technical point of view the model utilised by the submitter is transparent, but the structure of the model represents a major simplification of clinical practice. The submitted model simply divides patients into low-, intermediate- and high-risk groups, and then proceeds by making assumptions at patient population level about how treatment practice currently is without the test and how it might change after introduction of Prosigna. However, most decision analytical models are simplifications of reality, and may still produce valid estimates that are relevant for clinical decision making. The important question is whether it is simplified in a way that enables it to produce valid and useful results. Since we had to rely on modifications of the submitter's model, our estimates are affected by the same

structural uncertainties as the cost-effectiveness estimates produced by the submitter.

Another important but highly uncertain estimation in the cost-effectiveness analyses, was the proportion of chemotherapy use in both the Prosigna arm and the current practice arm. We received estimates for the use of chemotherapy for the different Prosigna risk groups based on Norwegian data, which we find substantially more trustworthy than the input data in the submission, but also the data we used are also based on assumptions, not real-life observations.

To highlight the uncertainties related to the input data on chemotherapy use in current practice, we carried out two analyses in addition to our base case scenario. In the alternative scenarios we adjusted the proportion of patients receiving chemotherapy using data from two studies where clinicians had been asked how knowing the Prosigna result would change their opinion about chemotherapy use, for specific patients (22;23). The resulting cost effectiveness estimates were dramatically different from our base case scenario, highlighting how sensitive the model is to changes in assumptions about how the Prosigna test will affect chemotherapy use. As a further illustration of this: In our base case analysis we decided to assume 10-20% non-adherence to guideline recommendations for the use of chemotherapy (see Appendix 1). This decision was taken at a late stage, and led to a massive 50% reduction in the estimated ICER (from NOK 1.8 million to 0.9 million).

Finally, we find the way the utility value "risk profiling knowledge to patients" has been included in the submitted analysis, inappropriate.

## Need for further research

Comparative studies, preferably randomised controlled trials, are needed to establish the clinical benefit of including Prosigna when assessing the risk of recurrence in breast cancer patients (HR+/ HER2- node negative). Two trials are underway from which the findings are expected to strengthen the evidence base for the clinical usefulness of the Prosigna test. An economic evaluation is also needed that incorporates the prognostic performance of Prosigna for individual level clinical decisions.

# Conclusions

There is probably a close association between risk prediction provided by Prosigna and the observed risk of recurrence among patients with HR+/HER2-N0 type breast cancer. However, it remains uncertain whether Prosigna contributes with prognostic information that translates to better clinical results, i.e. reduction in the use of chemotherapy without an increase in the rate of distant recurrences.

We were not able to estimate trustworthy estimates of cost-effectiveness using the submitter's model, largely because we lack reliable data on chemotherapy use for different risk categories with and without Prosigna testing. There are also uncertainties regarding how a shift in the use of chemotherapy will translate into clinical outcomes.

# References


1. Helsedirektoratet. Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av pasienter med brystkreft. Oslo: 2019. IS-2760.
2. Kreftregisteret. Årsrapport 2017 med resultater og forbedringstiltak fra Nasjonalt kvalitetsregister for brystkreft. Oslo: 2018. ISBN: 978-82-473-0060-2.
3. Statens beredning för medicinsk och social utvärdering (SBU). Genexpressionsbaserade tester för vägledning av återfallsförebyggande behandling vid bröstcancer. 2018.
4. Kræftens Bekæmpelse. Medicinsk behandling efter operationen[cited].
5. Harnan S, Tappenden P, Cooper K, Stevens J, Bessey A, Rachid R, et al. Diagnostics Assessment Report commissioned by the NIHR HTA Programme on behalf of the National Institute for Health and Care Excellence. Tumour profiling tests to guide adjuvant chemotherapy decisions in people with breast cancer (update of DG10) 2017.
6. National institute for Health and Care Excellence. Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer. London: NICE [cited].
7. Nye Metoder. Referat fra møte i Bestillerforum RHF. 25.09.2017.
8. Whitning PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529-36.
9. Schwarzer G. meta: General Package for Meta-Analysis. R package version 4.3-2. ed2015.
10. Reitsma JB, Glas, A.S., Rutjes, A.W., Scholten, R. J., Bossuyt, P. M., Zwinderman, A.H. . Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. Clin Epidemiol 2005;58(10):982-90.
11. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria2017.
12. Sestak I, Buus R, Cuzick J, Dubsky P, Kronenwett R, Denkert C, et al. Comparison of the Performance of 6 Prognostic Signatures for Estrogen Receptor-Positive Breast Cancer: A Secondary Analysis of a Randomized Clinical Trial. JAMA Oncol 2018;4(4):545-53.
13. Laenkholm AV, Jensen MB, Eriksen JO, Rasmussen BB, Knoop AS, Buckingham W, et al. PAM50 Risk of Recurrence Score Predicts 10-Year Distant Recurrence in a Comprehensive Danish Cohort of Postmenopausal Women Allocated to 5 Years of Endocrine Therapy for Hormone Receptor-Positive Early Breast Cancer. J Clin Oncol 2018;36(8):735-40.
14. Gnant M, Filipits M, Greil R, Stoeger H, Rudas M, Bago-Horvath Z, et al. Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: using the PAM50 Risk of Recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. Ann Oncol 2014;25(2):339-45.
15. Ohnstad HO, Borgen E, Falk RS, Lien TG, Aaserud M, Sveli MAT, et al. Prognostic value of PAM50 and risk of recurrence score in patients with early-stage breast cancer with long-term follow-up. Breast Cancer Res 2017;19(1):120.

16. ISRCTN42400492: Optimal personalised treatment of early breast cancer using multiparameter analysis[cited]. Available from: https://doi.org/10.1186/ISRCTN42400492
17. NCT02889874. EXamining PErsonalised Radiation Therapy for Low-risk Early Breast Cancer (EXPERT)[cited]. Available from: https://clinicaltrials.gov/ct2/show/NCT02889874
18. Ward S, Scope A, Rafia R, Pandor A, Harnan S, Evans P, et al. Gene expression profiling and expanded immunohistochemistry tests to guide the use of adjuvant chemotherapy in breast cancer management: a systematic review and cost-effectiveness analysis. Health Technology Assessment (Winchester, England) 2013;17(44):1-302.
19. Helsedirektoratet. Økonomisk evaluering av helsetiltak – en veileder. Divisjon helseøkonomi og finansiering, Avdeling finansiering og DRG; 2012. Veileder. IS-1985.
20. Norsk gynekologisk forening, editor. Legeforeningen. Klimakteriet. Norsk gynekologisk forening; 2015. Veileder i gynekologi 2015. Available from: https://legeforeningen.no/Fagmed/Norsk-gynekologisk-forening/Veiledere/Veileder-i-gynekologi-2015/Klimakteriet/
21. Drummond MF, Sculpher, M. J., Calxton, K., Stoddart, G. L., Torrance, G. W. Methods for Economic Evaluation of Health Care Programmes. 4 ed.: Oxford University Press; 2015.
22. Wuerstlein R, Sotlar K, Gluz O, Hofmann D, Otremba B, Von Schumann R, et al. Prosigna results impact on adjuvant decision making in early breast cancer (EBC): Final analysis of the prospective WSG study. Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Publication: 2016;76(4 SUPPL. 1).
23. Martin M, Gonzalez-Rivera M, Morales S, de la Haba J, Gonzalez-Cortijo L, Manso L, et al. Prospective study of the impact of the Prosigna (TM) assay on adjuvant clinical decision-making in women with estrogen receptor-positive, HER2-negative, node-negative breast cancer: A GEICAM study. Cancer Res 2015;75:2.
24. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Comparisons between diff erent polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100 000 women in 123 randomised trials. Lancet Infect Dis 2012;379:432-44.
25. Thomas RJ, Marshall C, Glen J, Callam M. The total hospital and community UK costs of managing patients with relapsed breast cancer. Br J Cancer 2009;100:598-600.
26. de Bock GHP, Bonnema J, van der Hage JA, Bartelink H, van de Velde CJ. The impact of loco-regional recurrences on metastatic progression in early-stage breast cancer: a multistate model. Breast Cancer Res Treat 2009;117:401-8.
27. Wolff AC, Visvanathan K, Rugo HS, Moy B, Goldstein LJ. Risk of marrow neoplasms after adjuvant breast cancer therapy: the National Comprehensive Cancer Network experience. J Clin Oncol 2015;33:340-9.
28. National institute for Health and Care Excellence. Azacitidine for the treatment of myelodysplastic syndromes, chronic myelomonocytic leukaemia and acute myeloid leukaemia. London: 2011. Technology Appraisal Guidance 218.
29. Statistics Norway, Deaths per 100000 mean population 2017 [Internet]. 2017. Available from: https://www.ssb.no/en/dode/
30. Vettus E, Oselin K. Comparison of the usage of granulocyte colony-stimulating factors (G-CSF) in the Baltic and Nordic countries in 2011-2014. Annals of oncology : official journal of the european society for medical oncology 2916;27.
31. Karnon J, Jack W, Papo NL, Cameron DA. Health care costs for the treatment of breast cancer recurrent events: estimates from a UK-based patient-level analysis. Br J Cancer 2007;97:479-85.
32. Aleris. Prisliste røntgentjenester[cited 13.06.2019]. Available from: https://www.aleris.no/mer-om-oss/priser/prisliste-rontgentjenester/

33. Wallden B, Storhoff, J., Nielsen, T., Dowidar, N., Schaper, C., Ferree, S., Liu, S., Leung, S., Geiss, G., Snider, J., Vickery, T., Davies, S. R., Mardis, E. R., Gnant, M., Sestak, I., Ellis, M. J., Perou, C. M., Bernard, S. B., Parker, J. S. Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. BMC Med Genomics 2015;8(54).
34. Campbell HE, Bloomfield D, Griffin S, Manca A, Yarnold J. The cost-effectiveness of adjuvant chemotherapy for early breast cancer: A comparison of no chemotherapy and first, second, and third generation regimens for patients with differing prognoses. Eur J Cancer Clin Oncol 2011;47:2517-30.
35. O'neill SC, Lillie SE. Women's Interest in Gene Expression Analysis for Breast Cancer Recurrence Risk. . J Clin Oncol 2007;25(29):4628-34.
36. Lidgren M, Jonsson B, Rehnberg C. Health related quality of life in different states of breast cancer. Qual Life Res 2007;16:1073-81.
37. Younis T, Skedgel C. Adjuvant chemotherapy in breast cancer: is TC a cost-effective regimen compared to AC? CTRC-AACR San Antonio Breast Cancer Symposium 2008.
38. Legemiddelverket. Engelsk oversettelse av retningslinjer og mal - hurtig metodevurderinger[cited 13.06.2018]. Available from: https://legemiddelverket.no/nyheter/engelsk-oversettelse-av-retningslinjer-og-mal-hurtig-metodevurderinger
39. Meld. St. 34. (2015-2016). Verdier i pasientens helsetjeneste - melding om prioritering. 2016.
40. Burström K, Johannesson M, F. D. Swedish population health-related quality of life results using the EQ-5D. Quality of life research 2001;10(7):621-35.
41. Population by sex age and marital status. [Internet]. Statistics Norway.
42. Age-standardised (Norwegian standard) incidence rates per 100 000 person-years by primary site and year, 2016.: Cancer in Norway 2016 [cited]. Available from: https://www.kreftregisteret.no/globalassets/cancer-in-norway/2016/cin-2106.pdf
43. Age-standardised (Norwegian standard) incidence rates per 100 000 person-years for selected cancers by stage and period of diagnosis: Cancer in Norway 2016 [cited]. Available from: https://www.kreftregisteret.no/globalassets/cancer-in-norway/2016/cin-2106.pdf
44. Saini KS, Azim, H. A. Jr., Metzger-Filho, O., Loi, S., Soltiriou, C., de Azambuja, E., Piccart, M. Beyond trastuzumab: new treatment options for HER2-positive breast cancer. The Breast 2011;20(3):20-7.
45. Proportion screen-detected breast cancers with lymph node involvement by screening area and year, 1996-2016 [Internet]. Cancer in Norway 2016.
46. Buchanan J, Wordsworth. S., Schuh, A. . Issues surrounding the health economic evaluation of genomic technologies. Pharmacogenomics 2013;14(15):1833-47.
47. Mushlin AI. To (genetic) test or not to test, that is the question. J Comp Eff Res 2015;4(5):429-31.
48. Ioannidis JP, Khoury, M. J. . Are randomized trials obsolete or more important than ever in the genomic era? . Genome Med 2013;5(18):32.
49. National Health Care Institute (ZIN) BHCKCK. Added value of using the gene expression signature test MammaPrint® for adjuvant chemotherapy decision-making in early breast cancer. EUnetHTA; 2018. Rapid assessment of other technologies using the HTA Core Model® for Rapid Relative Effectiveness Assessment.
50. Legemiddelsøk [Internet]. 2018 [cited Aug-Dec 2018]. Available from: https://www.legemiddelsok.no/
51. Paclitaxel [Internet]. [cited 2018]. Available from: https://reference.medscape.com/drug/taxol-paclitaxel-342187
52. Afipran 10 mg tablett [Internet]. 2018. Available from: https://www.felleskatalogen.no/medisin/pasienter/pil-afipran-takeda-545797

# Appendix 1

## Estimates for proportions of patients receiving chemotherapy, per Prosigna risk group

Our base case estimates for proportions of patients that are likely to receive chemotherapy are provided by Bjørn Naume, one of the co-authors of the Oslo1-study:

"The estimates are based on the patients in the Oslo1 study where prosigna-test results and tumor characteristics are available. Those data were used together with recommendations from Norwegian clinical practice guidelines, to decide how patients would be selected for chemotherapy today.

For the Luminal B like group, the estimates were based on either grade 3 or grade 2 with Ki67 hot spot above 15% (which was the median value for the tested patients in Oslo1, mainly pT1 group).

The selection of chemotherapy candidates in the all "node negative" and in the Luminal B like-group used grade 3 or grade 2 and Ki67 above median as selection criteria for chemotherapy (except for pT1a-b: no routine chemotherapy for these patients).

After prosigna test, in general all Luminal A patients (ROR low or intermediate) were selected to no chemotherapy. For the ROR intermediate Luminal B, chemotherapy was selected only for the pT2 tumors (endocrine for the pT1). All ROR high with pT1c-pT2 tumors were selected to receive chemotherapy.

In addition, to provide estimates for the Oslo1-study indicating chemotherapy candidates, for patients without the prosigna test-result, 20% were subtracted to take into account that not all candidates for chemotherapy will end up receiving it due to factors such as age, comorbidity, clinical judgement, or patient preferences. This is based on an assumption, not on data. For patients with the prosigna test-result, only 10% were subtracted from the chemotherapy candidates as it was assumed that the result would increase the confidence in deciding for chemotherapy."

(Bjørn Naume, June 21st 2019)

# Appendix 2

## Unit costs

The unit costs related to the pharmaceuticals are mainly average prices taken from the Norwegian Medicines Agency (NOMA) database in 2018 (50). The assumed patient proportions are based on clinical expert suggestions.

**Table 1. Unit costs used to calculate the total chemotherapy costs (EC90 only)**

| Variables | Unit cost (ref) | Resource use | Total cost (4 cycles) | Sources |
|---|---|---|---|---|
| EC90 (AUP cost) | (see table 1) | (see table 1) | NOK 59,478 | (see table 1) |
| Paclitaxel | NOK 258 / 6 mg | 12 | (NOK 3,096 x 3 x 4)<br>NOK 37,152 | (51) |
| Cetirizine | | 30 packing | NOK 150 | (50) |
| Ranitidine | | 150 mg per day | NOK 72 | (50)<br>Assumption |
| TOTAL COST (AUP) | | | NOK 96,852 | |
| TOTAL COST (AUP exclusive VAT) | | | NOK 77,482 | |

**Table 2. Unit costs used to calculate the total chemotherapy costs (EC90 + Paclitaxel)**

| Variables | Unit cost (ref) | Resource use | Total cost (4 cycles) | References |
|---|---|---|---|---|
| Nursing time (45 min x 4) | NOK 348/hour | 3 hours | NOK 1,740 | Assumption by clinical expert suggestions |
| Physician time (30 min x 4) Including both the wrist and the infusion device doctor | NOK 750/hour | 2 hours | NOK 1,500 | Assumption by clinical expert suggestions |
| Epirubicin 90mg/m2 x 4 | NOK 1664 / 50 ml<br>NOK 850.10 / 25 ml<br>NOK 361.8 / 10 ml<br>NOK 199 / 5 ml | 90 mg/m2 | (NOK 3,074 x 4)<br>NOK 12,296 | NOMA (50) |
| Cyclophospha-mide 600 mg/m2 x 4 | NOK 204.8 / 1000 mg<br>NOK 70 / 200 mg | 600 mg/m2 | (NOK 274.8 x 2)<br>NOK 549,6 | NOMA (50) |
| Neulasta/ Lon-quex (single dose of 6mg) x 4 | NOK 8959,5 / 6 mg | 6 mg | (NOK 8959,5 x 4)<br>NOK 35,838 | NOMA (50) |
| Akynzeo (Nausea –society cost) | NOK 738 / 300 mg | 300 mg | (NOK 738 x 4)<br>NOK 2,952 | NOMA (50) |
| Dexamethasone (Nausea – society cost) | NOK 154 / 40 mg | 40 mg | (NOK 154 x 4)<br>NOK 616 | NOMA (50) |
| Afipran (Nausea – society cost) | NOK 21.6 /10 mg | 10 mg | (NOK 21.6 x 4)<br>NOK 86.4 | Felleskataolgen (52)<br>NOMA (50) |
| Wig (Society cost) | NOK 650 | 1.5 | (NOK 975 x 4)<br>NOK 3900 | Assumption |
| TOTAL COST (AUP) | | | NOK 59,478 | |
| TOTAL COST (AUP exclusive VAT) | | | NOK 47,582.40 | |

**Table 3. Unit costs used to calculate the total endocrine therapy costs**

| Variables | Dosage | Price per pack | Annual cost | Patient share | References price per pack/ patient share |
|---|---|---|---|---|---|
| Tamoxifen | 20mg/day | NOK 341 | NOK 1,2445 | 10% | (50) / clinical expert as-sumption |
| Anastrozole | 1mg/day | NOK 1,983 | NOK 7,382 | 20% | (50) / clinical expert as-sumption |
| Letrozole | 2.5mg/day | NOK 1,440 | NOK 5, 256 | 50% | (50) / clinical expert as-sumption |
| Exemestane | 25mg/day | NOK 1,440 | NOK 17, 520 | 20% | (50) / clinical expert as-sumption |

**Table 4. Unit costs used to calculate the total additional treatment costs**

| Other treatments | Dosage | Cost | Patient share |
|---|---|---|---|
| Bisphosphonates | 4mg vial | NOK 1440 (50)<br>(+ administration cost at first attendance NOK 2,152 (5)) | 30% (5) |
| G-CSF Neupogen Singleject 48million units | 0.5ml solution for injection pre-filled syringes | NOK 1,329 (1)<br>(18 days treated) | 14% (5) |
| District nurse visit | 30 min | NOK 116 (5)<br>(18 times) | All G-CSF patients |

NIPH

Published by the Norwegian Institute of Public Health
August 2019
P.O.B 4404 Nydalen
NO-0403 Oslo
Phone: + 47-21 07 70 00
The report can be downloaded as pdf
at www.fhi.no/en/publ/