Clinical relevance of different HER2 diagnostic tests to select patients for trastuzumab treatment

Hajar El Aouati

Student number: 7326254

Supervisors: Dr. J.A.A.G. (Anneke) Damen & L.F. (Linde) Huis in ‘t Veld

Second Examiner: Dr. A.E. (Anouk) Hiensch

Contents

Plain language summary

Human Epidermal growth factor Receptor 2 (HER2) is a protein on the surface of (cancer) cells that can promote growth of cancer. In some breast cancers, there are multiple copies of HER2 gene, leading to an overproduction of the HER2 protein. Identifying HER2-positive breast cancer is crucial because it determines eligibility for treatments that are designed to attack the cells with too much HER2 protein such as trastuzumab (Herceptin), which improves outcomes for these patients. This review examined different diagnostic tests used to determine HER2 status and aimed to identify the most effective test for selecting patients that would respond to trastuzumab treatment. Relevant articles were gathered through a search of two databases that contain medical articles. One predetermined objective was that the articles to be included had to be of a certain study design that was appropriate to answer our objective. Furthermore, the included articles were assessed for their risk of bias, which covers several study domains that might influence the findings. This was done using and adapting existing tools from Cochrane which is a leader in the field of reviews. The search resulted in 163 articles of which 6 remained after screening. The included articles only discussed the diagnostic tests immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH). One article investigated AQUA-TMA too, which is a certain type of IHC. Some studies found IHC to be a better predictor of survival and response after trastuzumab and others found FISH to be better. The response varied greatly between studies, which may be attributed to the difference in additional treatment the patients received and how skilled the assessors of the different diagnostic tests were. In the end we concluded that FISH may be better than IHC in selecting trastuzumab responders, although this difference is less obvious when the tests are performed in similar settings.

Abstract

Background: As breast cancer is one of the main causes of death for women worldwide, it is pivotal to accurately select patients that will benefit from treatment based on tumour characteristics. One of such characteristics is amplification/overexpression of the Human Epidermal growth factor Receptor 2 (HER2). Trastuzumab is an effective monoclonal antibody that has been approved as an anti-HER2 drug. However, it is unclear which diagnostic test is best to select patients for this treatment.

Methods: This review will assess which of the following test is best suited to do so: immunohistochemistry (IHC), in situ hybridization (ISH), enzyme linked immunosorbent assay (ELISA), polymerase chain reaction (PCR) and western blotting. The PubMed and CENTRAL databases were searched for relevant articles using keywords and MeSH terms. Eligibility criteria were set for the study population, diagnostic tests, treatment regimen and the outcome. Specific study designs that were deemed suitable to identify the best test were predefined. Cochrane's risk of bias tool and adjusted QUADAS-C and ROBINS-I were used to assess possible bias in the retrieved studies and how this may influence the findings.

Results: The search resulted in 163 articles of which 6 were eligible after screening. These 6 studies only discussed IHC, fluorescence ISH (FISH) and AQUA-TMA (which is a form of IHC). Treatment response rates were found to be similar for both fluorescence FISH positive and IHC positive patients, with a slight preference for FISH based on additional factors. The main reason of concern that could have biased the results is the lack of correcting for confounders in the single arm trials, which led to an oversimplified view on the effect of trastuzumab. Another point of concern was that none of the included studies were of the preferred study design which was hypothesized to introduce the least amount of bias.

Conclusion: We concluded that FISH and IHC are similar in selecting trastuzumab responders. FISH can be better than IHC in some cases, as it is more robust to influencing factors such as

patient population, presence of additional treatment and whether the diagnostic testing is done in a local or central laboratory.

Introduction

Breast cancer is currently the most common form of cancer among women worldwide (1), causing over 600,000 deaths per year (2). This is also the case for women in the Netherlands where one in seven women are diagnosed with breast cancer within their life time (3). Breast cancer can be divided into subtypes based on overexpression of hormone receptors on the cancer cells: human epidermal growth factor receptor 2 (HER2), progesterone or estrogen receptor (4). Each breast cancer subtype is associated with a certain prognosis and the type of hormone receptor(s) present determines what treatment will be administered (4). In HER2-positive breast cancers, an overexpression of the HER2 receptors leads to dysregulated cell proliferation (5). Due to this overexpression HER2-postive breast cancer is associated with more aggressive progression and has an increased metastatic potential (5). However, this characteristic is also the reason why HER2-targeted treatment in the form of monoclonal antibodies (MAB) is more successful in HER2-positive patients (6). The first introduced and most commonly used MAB is trastuzumab (5,7,8)

To determine whether patients are affected by HER2-positive breast cancer (and thus might be suitable for trastuzumab therapy), several diagnostic tests are available with the most commonly used being immunohistochemistry (IHC) and/or in situ hybridization (ISH) (6,9). However, several other tests can also be used for HER2 diagnosis such as enzyme linked immunosorbent assay (ELISA), polymerase chain reaction (PCR) and western blotting (10,11). Several studies have already proven that IHC and ISH are amongst the best of the diagnostic tests, have high concordance (12) and are thus taken up in diagnostic guidelines (9,12). On the other hand, little is known about whether this greater diagnostic power also translates to a better response to trastuzumab treatment. Several studies have shown that there may be a difference in

trastuzumab response based on which diagnostic test was used to determine the HER2-positive status (5).

This scoping review will assess the current evidence on whether the diagnostic test used to determine HER2 status is associated with the survival and response to trastuzumab therapy and thus has the most clinical utility.

Methods

In- and Exclusion Criteria

The intended study population were breast cancer patients with either newly or invasive breast cancer (as trastuzumab is approved for these breast cancers (13)), eligible for trastuzumab therapy. Exclusion criteria for the study population were all other cancer patients for which trastuzumab has been approved or was studied in including a mixed population that consisted partly of breast cancer patients. The HER2 diagnostic tests considered in this review are immunohistochemistry (IHC), (fluorescence/ silver-enhanced/ chromogenic) in situ hybridization (ISH), enzyme linked immunosorbent assay (ELISA), polymerase chain reaction (PCR) and western blotting (WB). The treatment of interest was specified as trastuzumab use with or without additional medication/ procedures (such as chemotherapy or surgery). A combination of trastuzumab with other monoclonal antibodies (MABs) targeting HER2 was not accepted due to the possible synergistic effect of multiple MABs that could distort the findings. The included studies had to have at least one treatment arm that included trastuzumab and at least two HER2 diagnostic tests had to be performed. Overall survival (OS), progression free survival (PFS), clinical impact and objective response rate were all included as measures of outcome after trastuzumab therapy of which at least one had to have been recorded for a study to be included.

Relevant study designs were randomized controlled trials, cohorts/observational studies (both prospective and retrospective), and single arm studies. Additional requirements were set for the eligibility of the study designs. The ideal study design to be included was as follows: patients are randomized to one of multiple diagnostic tests. Based on the result of this test, trastuzumab is administered accordingly and response is compared between individuals positive for test A and treated with trastuzumab, versus individuals positive for test B and treated with trastuzumab (see

figure 1A). An alternative to this design was when the researchers analysed the outcome after trastuzumab stratified by test that was used. This means that the treatment effect in patients who received test A is compared to those who received test B (see figure 1B). However, as it is expected that these study designs are rare, a study design in which the test received by the patient was not randomized was also permitted. The following three designs were also allowed:

1) patients received both tests and received trastuzumab (or an alternative treatment) either through randomisation or based on the test result (figure 2A)
2) patients received either test A or test B tests and received trastuzumab (or an alternative treatment) either through randomisation or based on test result (figure 2B)
3) patients received either test A or test B tests after which they were included in a trial due to a positive test result by either test (figure 2C)

In all three of these alternative designs the individuals with a positive result for test A were compared with the individuals with a positive result for test B. For the designs in figure 2A and 2B, studies were also included if the outcome (especially survival) was recorded for the alternative treatment and the analysis compared the patients who received test A compared to those who received test B. Articles were excluded from this review if the design used was a case-control or cross- sectional study. Furthermore, reviews, study protocols and opinion/ commentary articles were excluded.

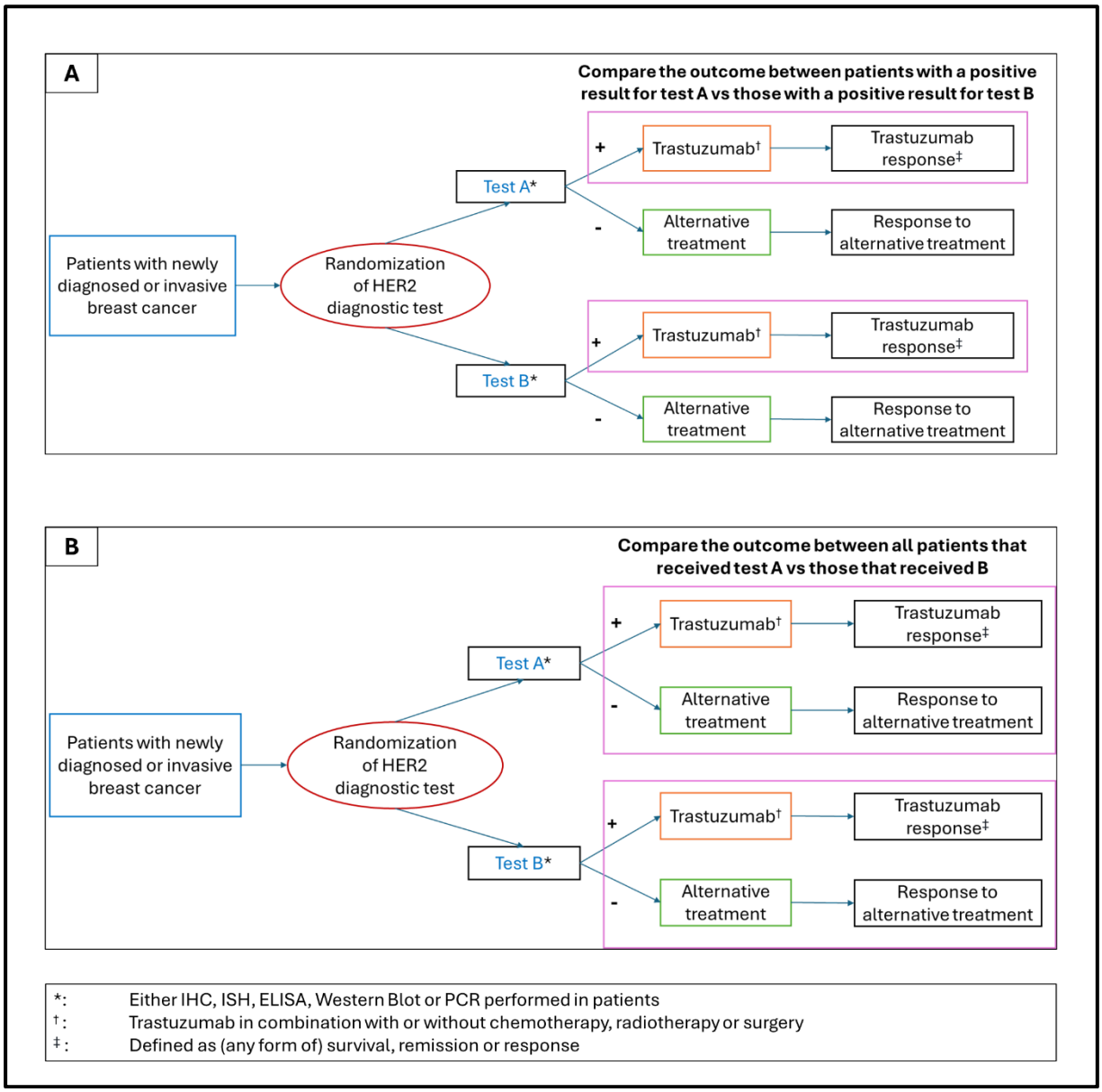


Figure 1: Preferred study design of the articles to be included. A) Patients are randomized to either test A or test B. Either trastuzumab or an alternative treatment is given based on test result. Amongst those that received trastuzumab, the patients with a positive result for test A are compared to those with a positive result for test B. B) Patients are randomized to either test A or test B. Either trastuzumab or an alternative treatment is given based on test result. Patients that received test A are compared to patients that received test B regardless of treatment.

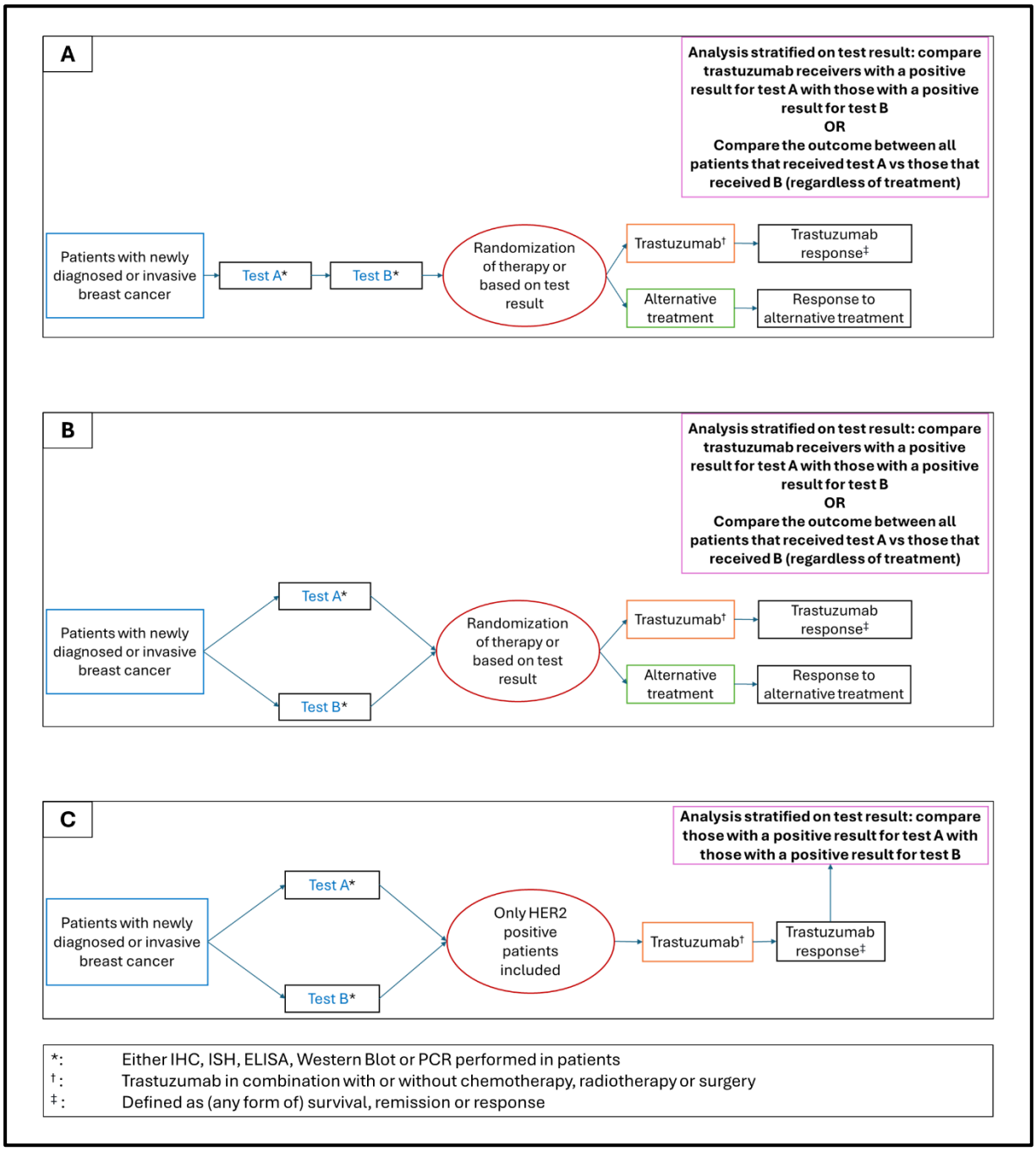


Figure 2: Alternative study design of the articles to be included. A) Patients receive two tests and trastuzumab or an alternative treatment either through randomisation or based on the test result. Amongst those that received trastuzumab patients with a positive result for test A are compared to patients with a positive result for test B or patients with a positive result for test A are compared to patients with a positive result for test B regardless of the treatment they received. B) Patients received either test A or test B tests and receive trastuzumab (or an alternative treatment) either through randomisation or based on test result. Amongst those that received trastuzumab patients with a positive result for test A are compared to patients with a positive result for test B or patients with a positive result for test A are compared to patients with a positive result for test B regardless of the treatment they received. C) Patients received either test A or test B tests after which they were included in a trial due to a positive test result by either test. The analysis/outcome is stratified on test.

Search Strategy and Data Extraction

Searches were performed in the electronic databases of PubMed and CENTRAL. Medical Subject Headings (MeSH) and specified word and spelling variations were used to search for relevant articles. The search also included filters on publication type. No filter on publication year was applied, all studies from inception were eligible for this review. The retrieved articles were screened based on title and abstract. Next full texts were retrieved, these were then analyzed for eligibility in this review based on the in- and exclusion criteria. The searches, screening and full text analysis were done by one person aided by additional researchers to reach a consensus on relevant topics/search variations on the 20th of June 2024. See Appendix 1 and 2 for an overview of the search strategies and results per database. The study selection was done in Zotero version 6.0.36 (14). The following information was extracted from the included articles: sample size, study design, location of the study, the diagnostic tests used, the study population, the treatment (combination), the outcomes that were measured and the findings/ final conclusions.

Critical Appraisal

Due to the unique focus of this review, which includes diagnostic, interventional and prognostic aspects, there is currently no single tool available to assess the risk of bias or critically appraise the included studies. Three existing tools have been used and partially adapted to fit the research question at hand.

- For RCTs the Cochrane tool for assessing Risk of Bias in randomized trials (RoB) was used (15)
- For the non-randomized studies, Cochrane's tool for assessing the risk of bias in Risk Of Bias In Non-randomized Studies - of Interventions (ROBINS-I) (16) was adapted (see Appendix 3)

- For the domain ‘Confounding’ only one simplified question that encompasses multiple questions was kept: ‘If present, did the study control for confounding that could affect the outcome within the study?’
- The ‘Patient Selection’ domain was removed as this domain is also discussed in the adapted QUADAS-C tool.
- Domain ‘Classification of Interventions’ was unaltered (however some questions are not applicable for single arm studies)
- For the domain ‘Deviations from the Intended Interventions’ only questions 1 and 2 were kept as the aim of assessing the effect of the intervention(s) was more relevant than to assess the effect of starting and adhering to intervention. The question was slightly adjusted: a study would still be considered at a low risk of bias if deviation of the intended intervention(s) were corrected for in the analysis.
- The questions in the domain ‘Missing Data’ were simplified to whether missing data was present and if this was appropriately handled in the analysis or unlikely to cause bias.
- The domain ‘Measurement of Outcomes was also simplified to whether the outcome assessor was aware of the administered intervention(s) and whether this could have influenced the measurement of the outcome.
- Lastly, the domain ‘Selection of the Reported Result’ was summarized into one question ‘Was the reported result selected based on the pre-specified analysis plan within the cohort’.

- To assess the bias in the diagnostic tests that were used, the Cochrane tool for assessing the Quality Assessment of Diagnostic Accuracy Studies-Comparative (QUADAS-C) (17) was adapted (see Appendix 4)
  - In the domain of ‘Patient Selection’ the questions from the single test section were not altered. An additional question was added to further assess applicability: ‘Was

the patient spectrum representative of the patients who will receive the test in practice'. In the comparative accuracy section only the first question was omitted as comparison of diagnostic ability was not the main interest of this review.

- The 'Index Test' domain was changed to 'Tests' and was adapted to examine the conduct of the tests that were used. The first question from the single test accuracy section was omitted as there was no reference standard of interest in this review. The second question was adapted to 'Were the tests clearly described'. The question 'Were the tests conducted as per the protocol' was added. The first question from the comparative accuracy section was also removed (again because comparison of diagnostic ability was not the main interest of this review).
- The domain 'Reference Standard' was completely omitted as there is no reference standard for these tests.
- Lastly in the domain 'Flow and Timing' all questions from the single test accuracy section were not used as these all mention a reference. Question 1 and 3 from the comparative accuracy section were removed for the same reason. A last question was added: Did all patients receive the same tests regardless of the result of the other test/ did all patients receive both tests? This question is not applicable for studies in which the diagnostic test was randomized.

Randomized controlled trials were assessed using the RoB tool and the adapted QUADAS-C. Cohort studies and single arm trials were assessed using the adapted ROBINS-I and the adapted QUADAS-C.

Results

Search Result

The PubMed and CENTRAL searches led to 121 and 69 articles respectively. Within the CENTRAL search 3 duplicates were present and subsequently removed. Between the PubMed and CENTRAL search results, an additional 24 duplicates were identified and removed, leading to 163 unique articles. Based on Title and Abstract screening, 110 articles were excluded. For 45 out of the remaining 53 articles the full text was retrieved and analyzed for eligibility. Reason for exclusion based on full text were:

- Incorrect study design/ reviews/ commentary articles (n= 7)
- Incorrect study population (n= 9)
    - Either a mix of cancer patients or selection bias on which patients received the diagnostic tests
- Excluded based on trastuzumab treatment/outcome (n= 7)
    - No trastuzumab administered/ no mention of administration
    - Trastuzumab specific response not recorded
    - Multiple MABs administered
- Excluded based on diagnostic HER2 test (n= 2)
    - No mention of tests or only one test was used for HER2
- No stratification of response to trastuzumab by diagnostic test (n= 13)
- Incorrect classification of positive test result with regard to outcome (n=1)
    - I.e. when the analysis combines IHC2+ and IHC3+ patients and the outcome is grouped for these test result

In the end 6 out of the 45 eligible articles were included in the study. See Figure 3 for an overview of the study selection process.

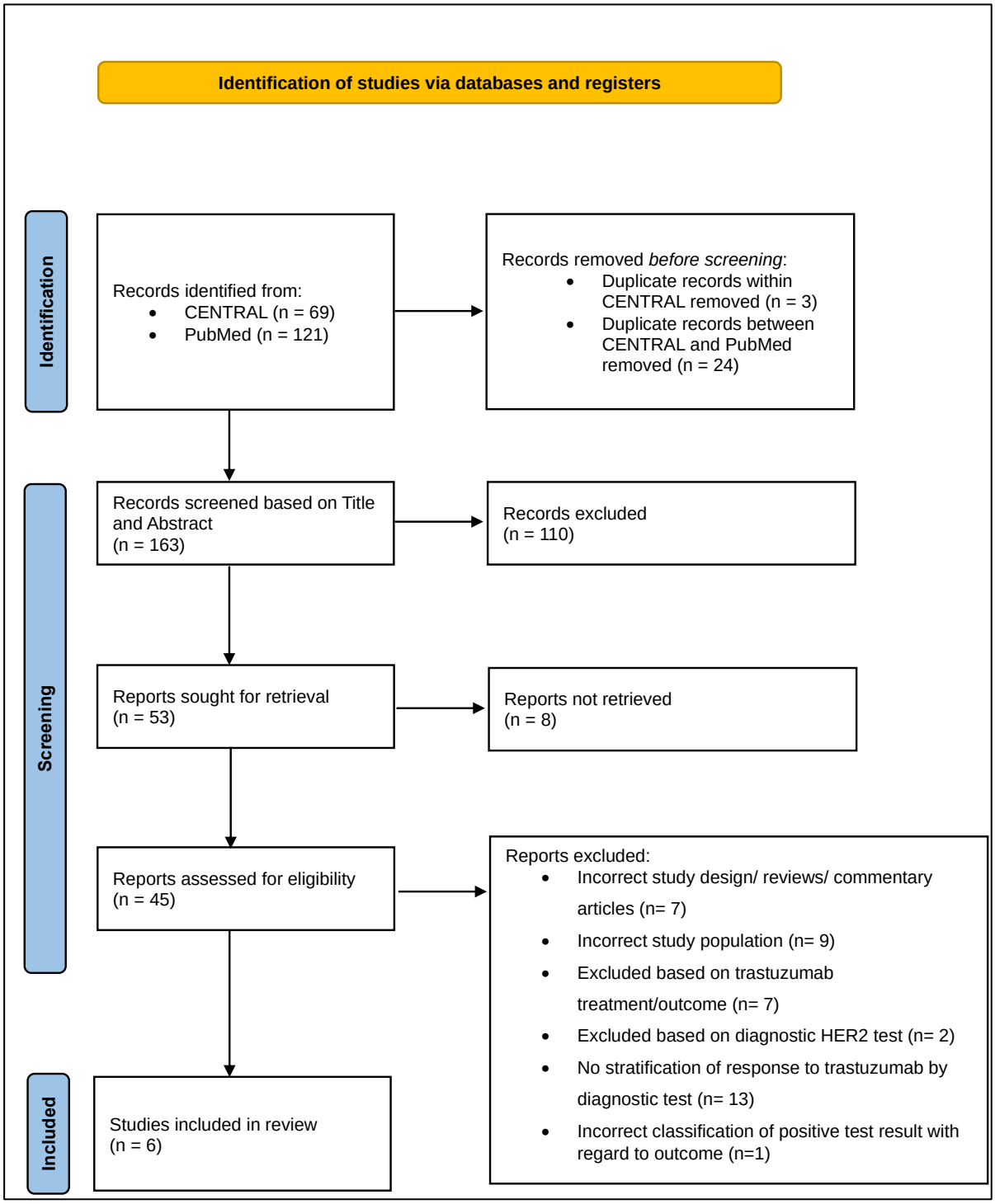


Figure 3: PRISMA 2020 flow diagram of the selected studies

Study Characteristics

The paper by Baselga (2001) used data from two clinical trials to answer several research questions in an individual participant data analysis (18). For this reason, the extracted information from this article is separated based on the original trials (from now on formulated as Baselga (2001) 1 and Baselga (2001) 2) leading to seven studies that were analyzed. Out of the seven included studies, two were randomized controlled trials in which at least one study arm was treated with trastuzumab (18,19). Of the remaining five studies, two were prospective single arm clinical trials (20,21) and three were retrospective analysis of single arm trials (18,22,23). Referring back to the above-mentioned accepted study designs for this review (see section *In- and exclusion criteria* under Methods or figures 1 and 2) none of the included papers had the 'ideal' study design in which randomization of diagnostic test occurs first after which trastuzumab is assigned based on test result or randomized. All studies performed a variation of the study designs described by figure 2A or 2C in which the included participants received both tests before inclusion or they could enter the study based on a positive test result on either one of two tests and were retested with the other test.

Initially 2 out of these 6 (19,21) articles were also excluded due to the fact that FISH was done only in IHC2+ and IHC3+ patients (which is a form of selection bias). These articles were kept as previous findings show that a FISH+ status is hardly found in IHC0+ and IHC1+ patients (24). The design of these studies still resembles that of the other included studies, except for the fact that the second test (which was FISH) was performed on a group of participants that had a higher probability of having a positive result (or the number of FISH+ patients was higher than what would have been expected if the test was performed in a representative sample of participants eligible for FISH).

All studies compared the same two diagnostic tests: immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH). The patients in the study by Giltnane et al. (2008) were retrospectively selected from a pharmacy database and had to have received trastuzumab therapy for at least 6 months. This study compared several versions of IHC and FISH: IHC and FISH whole slide (WS) analysis and IHC and FISH tissue micro array (TMA) analysis (22). Lastly AQUA-TMA was performed, a method that resembles IHC (22). See Table 1 for a complete overview of the study characteristics.

Table 1: Overview of the characteristics per study.

Author (year of publication)	*Sample size*	*Study Design*	*Location*	*Diagnostic tests performed*	*Population*
Baselga (2001) *	1. 213 (of which 60% received both diagnostic test) 2. 469 (of which 235 received trastuzumab along with chemotherapy)	1. Retrospective cohort analysis (single arm) 2. RCT	1. United States, Canada, Belgium, France, Germany, the United Kingdom, and Australia 2. N.R.	1. IHC and FISH 2. IHC and FISH	1. Patients with progressive breast cancer after one or two cytotoxic chemotherapy regimens for metastatic disease 2. Patients with HER2+ metastatic breast cancer who have not yet received therapy for metastatic disease
Burstein et al. (2003)	54	Prospective cohort analysis (single arm)	United States	IHC and FISH	Patients with stage IV breast cancer
Giltnane et al. (2008)	152 (of which 122 had a recorded trastuzumab response)	Retrospective cohort analysis (single arm)	Canada	IHC-WS, IHC-TMA, FISH-WS, FISH-TMA and AQUA-TMA	Patients with breast cancer that received trastuzumab for at least 6 weeks (patients with stage IV breast cancer and no evidence of disease were excluded)
Hofmann et al. (2008)	95	Retrospective cohort analysis (single arm)	N.R.	IHC and FISH	Patients with metastatic breast cancer
Tedesco et al. (2004)	26	Prospective cohort analysis (single arm)	United states	IHC and FISH	Patients (only women) with measurable metastatic breast cancer
Vogel et al. (2002)	114 (of which 108 were assessable)	RCT	United States and Canada	IHC and FISH	Patients (only women) with progressive HER2+ metastatic breast cancer

*: Baselga presented the results from two studies

RCT: randomized controlled trial; HER2: human epidermal growth factor receptor 2; N.R.: not reported; IHC: immunohistochemistry; FISH: fluorescence in situ hybridization; WS: whole slide; TMA: tissue micro-array; AQUA: automated quantitative analysis

Table 1 (continued)

Author (year of publication)	*Treatment*	*Outcome measured*
Baselga (2001) *	1. **Trastuzumab (a dosage of 4 mg/kg followed by weekly 2 mg/kg intravenous infusions).** Additional therapy was allowed in the case of disease progression. 2. Random allocation of either a chemotherapy alone or a chemotherapy combined with t**rastuzumab (given in the dosage of 4 mg/kg followed by weekly 2 mg/kg in until disease progression).**	1. Response rate defined as complete or partial response and other (not defined) 2. Response rate, time to disease progression and survival
Burstein et al. (2003)	**Trastuzumab (initially 4 mg/kg, thereafter 2 mg/kg),** additional vinorelbine was permitted as well as adjuvant chemotherapy and adjuvant therapy (CMF, anthracycline based, taxane based or anthracycline and taxane based).	Disease progression defined as progressive disease, stable disease, partial response or complete response
Giltnane et al. (2008)	**Trastuzumab for at least six weeks** (no additional information on dose reported). Concurrent chemotherapy was permitted (either none, vinorelbine, paclitaxel or other non-specified).	Response rate/categorical response with 2 levels: complete/partial response and stable/progressive disease
Hofmann et al. (2008)	**Trastuzumab monotherapy (initially a dose of 8 mg/kg followed by 6 mg/kg 3-weekly)**	Disease progression defined as progressive disease, stable disease, partial response or complete response
Tedesco et al. (2004)	**Trastuzumab (initially 4 mg/kg, then 2 mg/kg weekly until disease progression or toxicity)** was given alongside docetaxel.	Overall response rate defined as complete response, partial response, stable disease and progressive disease
Vogel et al. (2002)	**Random allocation to either standard dosage trastuzumab (initial dose of 4 mg/kg followed by 2 mg/kg weekly) or high dosage (initial dose of 8 mg/kg followed by 4 mg/kg weekly)**	• Objective tumor response defined as complete response, partial response, minor response, stable disease and disease progression • Clinical benefit which was defined as either complete response, partial response, minor response or stable disease for at least 6 months

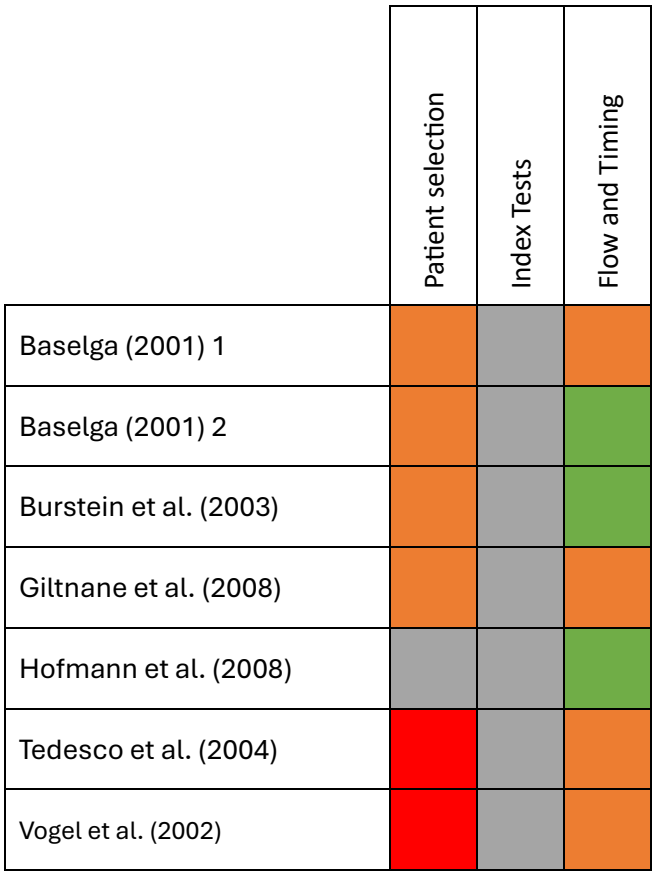
*: Baselga presented results from two clinical trials
mg/kg: milligrams per kilogram; CMF: cyclophosphamide, methotrexate, and fluorouracil

Risk of Bias Assessment & Critical appraisal

Adapted QUADAS-C

Five studies scored moderate risk of bias on the adapted QUADAS-C (see Table 2). The domain patient selection was mostly at a moderate or high risk of bias (respectively 4 and 2 studies). The main point of concern in the first four studies that led to the classification of a moderate risk of bias in this domain was that neither a fully paired design was used nor did all patients receive both diagnostic tests. Furthermore, there were some concerns about the patient selection as most studies retrospectively selected patient data on diagnostic test and response to trastuzumab. These variables all had to be available for the patients to be included in the analysis. This may be less representative of patients in routine clinical care in which both tests are not always performed. For the study by Hofmann et al. (2008), the risk of bias was scored unclear since it was not reported how the patients were selected and why patients were excluded. Lastly, due to the selection bias in the inclusion of participants that received the second test (FISH), Tedesco et al. (2008) and Vogel et al. (2002) were judged to be at a high risk of bias. All studies but one received an unclear risk of bias for the domain Index Tests as none of the studies reported whether the interpretation of one test was done without knowledge of the result of the other test. Vogel et al. (2002) did report on the blinding of the analysts. Furthermore, Baselga (2001) 1 and 2, Tedesco et al. (2008) and Vogel et al. (2002) did not report who interpreted the test results. Lastly for the domain Flow and Timing, the remaining studies scored a moderate risk of bias as not all patients received both tests and there was missing data that was not corrected for in the analysis.

Table 2. The adapted QUADAS for assessing risk of bias in comparative diagnostic accuracy studies.



Risk of bias by color; red: high; orange: moderate; green: low risk; grey: unclear.

RoB and adapted ROBINS-I

Two of the included studies were randomized clinical trials: Baselga (2001) 2 and Vogel et al. (2002) and were assessed with the Cochrane tool for assessing risk of bias in randomized trials (see Table 3). Both studies provided no or vague descriptions on how the random allocation of treatment was done and if/how the allocation was concealed. Vogel et al. (2002) reported that the researchers themselves assessed the outcome after trastuzumab. This combined with the lack of reporting on allocation concealment was reason to appoint a moderate risk of bias. Additionally, Vogel et al. (2002) did not report whether missing data was present and if this appropriately handled.

The other five studies were assessed with the adapted ROBINS-I for single arm trials (see Table 4). All single arm trials scored a high risk of bias for the domain Confounding as none of these trials corrected their results for additional variables that could explain the difference in response to trastuzumab and where applicable additional treatment. The additional therapy itself is not expected to be biasing in nature as this reflected clinical practice. Tedesco et al. (2004) was at a moderate risk of bias for Deviation from the Intended Intervention as the reason for these reported deviations was not mentioned. The domain Missing Data was also not free of bias as in the studies by Giltnane et al. (2008), Hofmann et al. (2008) and Tedesco et al. (2004) the missing data was nonignorable and not corrected for in the analysis.

Table 3. Cochrane tool for assessing risk of bias in randomized trials.

	Sequence generation	Allocation concealment	Blinding of participants, personnel and outcome assessors	Incomplete outcome data	Selective outcome reporting	Other sources of bias
Baselga (2001) 2						
Vogel et al. (2002)						

Risk of bias by color; red: high; orange: moderate; green: low risk; grey: unclear.

Table 4. The adapted ROBINS-I to assess the risk of bias assessment in single arm trials.

	Confounding	Classification of the Intervention	Deviations from the Intended Intervention	Missing Data	Measurement of Outcomes	Selection of the Reported Result
Baselga (2001) 1						
Burstein et al. (2003)						
Giltnane et al. (2008)						
Hofmann et al. (2008)						
Tedesco et al. (2004)						

Risk of bias by color; red: high; orange: moderate; green: low risk; grey: unclear.

Findings

All studies measured the response to trastuzumab therapy by pathological complete response (pCR) (also called overall response rate (ORR)). Response rate is defined as the percentage of participants with a complete or partial response. See Table 5 for an overview of the findings per study.

In the single arm trial analyzed by Baselga (2001), a FISH+ test result was found to be slightly more predictive of trastuzumab response than an IHC3+ test result: 20% vs 18%. Baselga suggested that FISH may be better in selecting the HER2 overexpressing patients that will actually have a clinical benefit from trastuzumab as it may also select certain IHC2+ patients that could benefit from trastuzumab therapy (but would have been denied trastuzumab based on their IHC status).

In the second trial that was analyzed (the RCT) one arm received chemotherapy and the other arm received trastuzumab and chemotherapy, again additional therapy was allowed. The clinical outcomes following trastuzumab were found to be rather similar in terms of time to disease progression and survival. When grouping all chemotherapies together, in the IHC3+ patients the response rate for the chemotherapy alone group vs chemotherapy with trastuzumab the response rates were 31% vs 56%. For the FISH+ patients this was 27% vs 56%. This signals that IHC3+ patients had a slightly greater response regardless of the treatment they received. When considering only the patients that received trastuzumab and paclitaxel, the time to treatment progression was similar: 7.1 and 7.0 months for IHC3+ and FISH+ respectively. The response rate was identical 49% for both groups and the median survival was also the same for both groups: 25 months. Some specific information on the median survival time and time to progression was specifically reported for IHC3+ patients receiving chemotherapy alone and chemotherapy with trastuzumab: 20 vs 29 months and 4.6 vs 7.8 months. This information was not reported for FISH

positive patients. Again, Baselga insinuates that FISH could detect potential trastuzumab responders that IHC testing may have missed: the group of patients with an IHC score of 2+ (that would be classified as HER2 ambiguous) and a FISH+ result.

Another single arm trial by Burstein et al. (2004) reported that 4 and 33 out of the 54 included patients achieved complete and partial response respectively after trastuzumab therapy, leading to an ORR of 68.5% (20). Out of these 54 patients, 44 were IHC3+ and 10 were FISH+. In the IHC3+ group 30 patients responded (68%) and in the FISH+ group 7 patients responded (70%). Fisher's exact test comparing the method of diagnostic test in relation to ORR resulted in a p-value of 0.99 indicating that there is no difference in the ability of these two tests to predict ORR to trastuzumab.

How the diagnostic tests were performed in Giltnane et al. (2008) was quite different from the other studies. Initially 152 patient samples were available for (re-) assessment of the HER2 status. However, due to differences in sample availability per test, the number of patient samples used per test differs (but is overlapping). The number of patients per test was as follows: IHC-WS: 77, IHC-TMA : 60, FISH-WS : 81, FISH-TMA: 113 and AQUA-TMA: 107. The response rate was calculated as the number of patients that received a partial or complete response divided by the total number of patients that received the test of interest. AQUA-TMA and IHC-TMA were considered to have the best predictive ability in terms of selecting trastuzumab responders by the researchers, 50.0% and 62.5% respectively. However, when calculating the response rate, FISH-WS and FISH-TMA both outperformed AQUA-TMA with response rates of 50.7% and 52.9% respectively.

Hofmann et al. (2008) described a monotherapy trial to test the efficacy of trastuzumab in patients with metastatic breast cancer. IHC and FISH test results were available in 95 of the 103 patients whose response to treatment was measured. Out of these 95 patients, 75 were IHC3+ and 74 were

FISH+. In the 75 IHC3+ patients 17 and 2 reached a partial and complete response respectively which led to an ORR of 25.3%. The response in the FISH+ patients was a bit lower as 14 and 2 patients had a partial and complete response respectively leading to an ORR of 21.6%. When only considering complete response, the response rate was 2.7% for both IHC3+ and FISH+ patients. Additionally, the researchers also underline that IHC was better than FISH in both sensitivity and specificity in terms of response to trastuzumab (100% vs 84.2% and 26.3% vs 23.7%).

The last single arm trial by Tedesco et al. (2004) aimed to determine the efficacy (and toxicity) of combined trastuzumab and docetaxel (a chemotherapy drug) in patients (specifically women) with measurable metastatic breast cancer. One of the inclusion criteria was for the participants to either have an IHC2+ or IHC3+. This led to the FISH tests (Abbott and Ventana) to have been performed on a less representative sample. In total 26 patients were included that completed 94% of the predetermined trastuzumab regimen and 91% of the docetaxel. The response rates (defined as both complete and partial response) were 63%, 63% and 65% for IHC3+, Abbott FISH+ and Ventana FISH+ respectively. The median time to progression was also found to be similar: 12.3 months for IHC3+ patients and 12.4 months for both FISH tests. It was concluded that there was no difference in ability of the tests to predict response to this treatment combination.

Finally, the second RCT by Vogel et al. (2002) investigated the response to different trastuzumab dosages (see table 1 under *Study Characteristics*) in progressive metastasized HER2 positive breast cancer patients. This was the second study that was subject to selection as patients were only included based on an IHC score of 2+ or 3+. This led to the inclusion of 114 patients of which 59 received the low dosage of trastuzumab and 55 the high dosage. FISH could not be performed in 2 and 1 patients respectively as the breast cancer sample was not available for retesting. The combination of complete and partial responses was used to describe the objective response rate. In the FISH+ participants this was 34% (95%CI 23.9 – 45.7%) and for the IHC3+ patients this was

35% (95%CI not reported). The clinical benefit which included complete, partial and minor responses and stable disease for at least six months, was also found to be identical for both groups: 48%. Lastly the median time to progression was 4.9 months (95%CI 3.4 – 8.0) for FISH+ patients. This was not reported for the IHC3+ group (nor for the combined IHC2+ and IHC3+ group). Despite the similar outcomes in both tests, the researchers note that FISH may be 'superior' in detecting the HER2-positive patients that will benefit from trastuzumab as it may be able to detect a subgroup of patients that have an IHC score of 2+ that will actually benefit from trastuzumab.

Table 5. Findings and conclusions per article

Author(s)	*Response rate (%) **	*Survival (median)*	*Time to progression (median)*	*Clinical Benefit*	*Conclusions*
Baselga (2001) *	1. FISH+ : **20%** IHC3+ : **18%**	1. N.R.	1. N.R.	1. N.R.	1. FISH is indicated to be a better selector of trastuzumab responders
	2. FISH+ that only received chemotherapy: **27%** FISH+ that received both trastuzumab and chemotherapy: **54%** IHC3+ that only received chemotherapy: **31%** IHC3+ that received both trastuzumab and chemotherapy: **56%** FISH+ that received trastuzumab and paclitaxel (a form of chemotherapy): **49%** IHC3+ that received trastuzumab and paclitaxel: **49%**	2. FISH+ that only received chemotherapy: **N.R.** FISH+ that received both trastuzumab and chemotherapy: **N.R.** IHC3+ that only received chemotherapy: **20 months** IHC3+ that received both trastuzumab and chemotherapy: **29 months** FISH+ that received trastuzumab and paclitaxel: **25 months** IHC3+ that received trastuzumab and paclitaxel: **25 months**	2. FISH+ that only received chemotherapy: **N.R.** FISH+ that received both trastuzumab and chemotherapy: **N.R.** IHC3+ that only received chemotherapy: **4.6 months** IHC3+ that received both trastuzumab and chemotherapy: **7.8 months** FISH+ that received trastuzumab and paclitaxel: **7.1 months** IHC3+ that received trastuzumab and paclitaxel: **7.0 months**	2. N.R.	2. Similar response rate, time to disease progression and survival were observed between FISH+ and IHC3+ patients. Again, FISH is indicated to be a better selector of trastuzumab responders and to select the proportion of IHC2+ (which would be considered a negative/equivocal IHC result) that may respond to trastuzumab treatment. A combination of trastuzumab with chemotherapy was found to be superior to chemotherapy alone
Burstein et al. (2003)	FISH+ : **70%** IHC3+ : **68%**	N.R.	N.R.	N.R.	Trastuzumab response was not dependent on which diagnostic test was used to determine the positive HER2 status.
Giltnane et al. (2008)	IHC-WS+ : **40.3%** IHC-TMA+ : **62.5%** FISH-WS+ : **50.7%** FISH-TMA+ : **52.9%** AQUA-TMA+ : **50.0%**	N.R.	N.R.	N.R.	AQUA-TMA and IHC-TMA are reported to have the best predictive ability in terms of the selection of trastuzumab responders by the researchers

*: Baselga presented results from two clinical trials

N.R.: not reported; IHC: immunohistochemistry; FISH: fluorescence in situ hybridization; WS: whole slide; TMA: tissue micro-array; AQUA: automated quantitative analysis

Table 5 (continued)

Author(s)	*Response rate(%) **	*Survival (median)*	*Time to progression (median)*	*Clinical Benefit*	*Conclusions*
Hofmann et al. (2008)	Complete response only FISH+ : **2.7%** IHC3+ : **2.7%** Complete and partial response FISH+ : **21.6%** IHC3+ : **25.3%** Complete and partial response reported as sensitivity FISH+ : **84.2%** IHC3+ : **100%** Complete and partial response reported as specificity FISH+ : **23.7%** IHC3+ : **26.3%**	N.R.	N.R.	N.R.	Better pathological response rate to trastuzumab (complete or partial) was seen in IHC3+ patients compared to the FISH+ patients. When only considering complete response after trastuzumab treatment neither FISH nor IHC was superior.
Tedesco et al. (2004)	FISH+ (Ventana) : **65%** FISH+ (Abbott) : **63%** IHC3+ : **63%**	N.R.	FISH+ (Ventana) : **12.4 months** FISH+ (Abbott) : **12.4 months** IHC3+ : **12.3 months**	N.R.	Pathological response rate after trastuzumab treatment was not dependent on diagnostic test. Furthermore, similar median time to progression were found for IHC and FISH (note two probes were used in the FISH assessment: Ventana and Abbott).
Vogel et al. (2002)	FISH+ : **34% (95%CI 23.9 - 45.7)** IHC3+ : **35% (95%CI N.R.)**	N.R.	FISH+ : **4.9 months (95%CI 3.4 - 8.0)** IHC3+ : **N.R.**	FISH+ : **48%** IHC3+ : **48%**	FISH positive patients were found to have similar responses and clinical benefit compared to IHC3+ patients. However, the researchers note that FISH may be 'superior' in detecting the HER2-positive patients that will benefit from trastuzumab.

N.R.: not reported; IHC: immunohistochemistry; FISH: fluorescence in situ hybridization; WS: whole slide; TMA: tissue micro-array; AQUA: automated quantitative analysis

Discussion

This review summarized and critically appraised the findings of 6 articles (that included 7 studies) with the aim to determine which HER2 diagnostic test is superior in selecting patients that will benefit/respond to trastuzumab therapy. Almost exclusively (IHC) and FISH were found to have been used as diagnostic tests and the most reported outcome following treatment was response rate. Overall, the response rates were very similar between FISH+ and IHC3+ patients, FISH performing slightly better in three studies and IHC being slightly better in four studies. The studies varied considerably in terms of treatment regimen and risk of bias, leading to a heterogeneous result. No studies were found that investigated the prognostic ability of ELISA, PCR or Western Blotting in terms of trastuzumab response.

What can be noticed from the findings (Table 5) is that the time to progression and response rates following treatment that contained trastuzumab varied considerably between studies, e.g. the response rates ranged from 18/20% to Baselga (2001) 1 to 68/70% in Burstein et al. (2003). The most likely cause for this discrepancy is the type of treatment that patients received. The dose of trastuzumab itself differed between some studies, but the additional treatment(s) that were either predetermined (in the randomized clinical trials) or permitted (in the single arm trial) were quite different and at times even tailored to specific tumor characteristics, such as targeted treatment for hormone positive breast cancer. Another reason for this difference is the different inclusion criteria for patients across the studies, as patients with previous adjuvant therapy were allowed in one trial (suggesting a population with an already worse prognosis) while another trial included HER2 positive patients with metastasis that had no prior treatment yet.

Previously the difference in diagnostic ability between IHC and FISH and their discordance has been highlighted by Memon et al. (2002) as they showed that the discordance percentage was 4%

over 43.468 breast cancer patients pooled from 46 studies (24). This difference also seems to translate in terms of prognostic ability as retrospective analysis by Stocker et al. (2020) showed that FISH had greater prognostic value in terms of recurrence, metastasis and overall survival compared to IHC. However, the overall prognosis in patients with newly diagnosed or metastasized breast cancer was assessed and not specifically for patients that received trastuzumab therapy. This review found evidence that FISH is able to select certain IHC negative patients (and would have been deemed HER2 negative and thus ineligible for trastuzumab) that would benefit from trastuzumab which is in line with a study by Gibbons-Fideler et al. (2019) that identified IHC-/ISH+ breast cancer patients that did respond to trastuzumab therapy (25).

This review presents that some studies found IHC to be superior while others found FISH to better predict trastuzumab benefit. A reason for this may be due to the difference in the setting/ quality of testing (which was not reported in most studies). IHC and FISH test results are known to differ based on whether the tests were analyzed in a central or local laboratory (26). Testing at central laboratories overall shows more concordance between the two tests (26,27). In the included studies, central determination of HER2 status was done by Hofmann et al. (2008), Tedesco et al. (2004) and Vogel et al. (2003), however the concordance between IHC and FISH was not particularly greater than in other studies (of which it is unknown whether testing was done locally or centrally).

A major concern in this review is that all single arm trials scored a high risk of bias for the domain Confounding in the adapted ROBINS-I as none of these trials corrected their results for additional variables that could explain the difference in response to trastuzumab (and where applicable for additional treatment). This was expected as most of these studies were phase II clinical trials in which the efficacy of the experimental drug(s) was tested. This could have distorted the observed efficacy of trastuzumab to be greater than it is in reality simply because other variables that could

explain this high response were not measured. Again this can also be seen in the wide range of response rates across the studies. Additionally, another influencing factor on the results is that none of the included studies abided by the preferred study design in which patients were randomized to one of two diagnostic tests. The difference in test allocation and how many patients were compared with each other is likely to have an impact on which test *seems* best.

One of the notable strengths of this review was the focus on a unique objective that has not been extensively explored in empirical research yet. This was facilitated by a comprehensive search of two key databases, PubMed and CENTRAL. Additionally, another positive aspect was the extensive critical appraisal of the studies. However, it must be noted that the adapted QUADAS-C and ROBINS-I can be subject to error as they have neither been validated nor were these tools designed for such a complex topic. Another limitation was that all of the included articles studied patients in the metastatic setting, even though this review aimed to make statements about both newly diagnosed patients and those with metastatic breast cancer as trastuzumab is approved for both of these groups (13). This does limit the generalizability of the findings.

We recognize that the proposed ideal study design in which breast cancer patients are randomized to one of two tests may not be completely ethical as several HER2 diagnostic tests are already known to perform less in terms of diagnosing and in predicting survival or treatment response. Given the lack of equipoise, we propose a fully paired study design in which eligible patients receive both tests. Then allocation based on the test results of which a positive test result takes precedence, i.e. if at least one test indicates a positive HER2 status then trastuzumab should be given. The response to treatment (both trastuzumab and the alternative treatment) should be stratified on the concordant and discordant test results (-/- , -/+, +/- and +/+).

Conclusion

The ability of IHC and FISH to predict response to trastuzumab therapy, and thus select the patients that are most likely to benefit, is very similar. The deciding factors are the population (newly diagnosed or patients with invasive breast cancer) that is tested, which concurrent or even synergistic treatment is given, and the quality of the laboratory where this testing is done. Overall, FISH seems most robust to these factors in accurately predicting trastuzumab treatment response.

References

1. The Global Breast Cancer Initiative [Internet]. [cited 2024 Jun 21]. Available from: https://www.who.int/initiatives/global-breast-cancer-initiative

2. Breast cancer [Internet]. [cited 2024 Jun 21]. Available from: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

3. Kans op borstkanker | RIVM [Internet]. [cited 2024 Jun 21]. Available from: https://www.rivm.nl/bevolkingsonderzoek-borstkanker/over-borstkanker/kans-op-borstkanker

4. Orrantia-Borunda E, Anchondo-Nuñez P, Acuña-Aguilar LE, Gómez-Valles FO, Ramírez-Valdespino CA. Subtypes of Breast Cancer. In: Breast Cancer [Internet] [Internet]. Exon Publications; 2022 [cited 2024 Jun 21]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK583808/

5. Stocker A, Trojan A, Elfgen C, Hilbers ML, Moskovszky L, Varga Z. Differential prognostic value of positive HER2 status determined by immunohistochemistry or fluorescence in situ hybridization in breast cancer. Breast Cancer Res Treat. 2020 Sep 1;183(2):311–9.

6. Breast Cancer HER2 Status | What is HER2 Status? [Internet]. [cited 2024 Jun 22]. Available from: https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-her2-status.html

7. Zhu K, Yang X, Tai H, Zhong X, Luo T, Zheng H. HER2-targeted therapies in cancer: a systematic review. Biomarker Research. 2024 Feb 2;12(1):16.

8. Harris C, Ward R, Dobbins T, Drew A, Pearson S. The efficacy of HER2-targeted agents in metastatic breast cancer: a meta-analysis. Annals of oncology : official journal of the european society for medical oncology. 2011;22(6):1308-1317.

9. Borstkanker - Receptorbepaling - Richtlijn - Richtlijnendatabase [Internet]. [cited 2024 Jun 19]. Available from: https://richtlijnendatabase.nl/richtlijn/borstkanker/pathologie/receptorbepaling.html

10. Zoppoli G, Garuti A, Cirmena G, di Cantogno LV, Botta C, Gallo M, et al. Her2 assessment using quantitative reverse transcriptase polymerase chain reaction reliably identifies Her2 overexpression without amplification in breast cancer cases. Journal of Translational Medicine. 2017 May 1;15(1):91.

11. Tan LD, Xu YY, Yu Y, Li XQ, Chen Y, Feng YM. Serum HER2 Level Measured by Dot Blot: A Valid and Inexpensive Assay for Monitoring Breast Cancer Progression. PLoS One. 2011 Apr 13;6(4):e18764.

12. Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, et al. Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. JCO. 2018 Jul 10;36(20):2105–22.

13. herceptin [Internet]. [cited 2024 Jul 2]. What Does Herceptin® (trastuzumab) Treat? | Herceptin® (trastuzumab) for HER2+ Metastatic Breast Cancer. Available from: https://www.herceptin.com/patient/metastatic-breast-cancer/about-herceptin/what-is-it.html

14. Zotero | Your personal research assistant [Internet]. [cited 2024 Jul 2]. Available from: https://www.zotero.org/

15. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011 Oct 18;343:d5928.

16. ROBINS-I tool | Cochrane Methods [Internet]. [cited 2024 Jul 8]. Available from: https://methods.cochrane.org/robins-i

17. Yang B, Mallett S, Takwoingi Y, Davenport CF, Hyde CJ, Whiting PF, et al. QUADAS-C: A Tool for Assessing Risk of Bias in Comparative Diagnostic Accuracy Studies. Ann Intern Med. 2021 Nov;174(11):1592–9.

18. Baselga J. Herceptin alone or in combination with chemotherapy in the treatment of HER2-positive metastatic breast cancer: pivotal trials. Oncology. 2001;61 Suppl 2:14–21.

19. Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L, et al. Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. J Clin Oncol. 2002 Feb 1;20(3):719–26.

20. Burstein HJ, Harris LN, Marcom PK, Lambert-Falls R, Havlin K, Overmoyer B, et al. Trastuzumab and vinorelbine as first-line therapy for HER2-overexpressing metastatic breast cancer: multicenter phase II trial with clinical outcomes, analysis of serum tumor markers as predictive factors, and cardiac surveillance algorithm. J Clin Oncol. 2003 Aug 1;21(15):2889–95.

21. Tedesco KL, Thor AD, Johnson DH, Shyr Y, Blum KA, Goldstein LJ, et al. Docetaxel combined with trastuzumab is an active regimen in HER-2 3+ overexpressing and fluorescent in situ hybridization-positive metastatic breast cancer: a multi-institutional phase II trial. J Clin Oncol. 2004 Mar 15;22(6):1071–7.

22. Giltnane JM, Molinaro A, Cheng H, Robinson A, Turbin D, Gelmon K, et al. Comparison of quantitative immunofluorescence with conventional methods for HER2/neu testing with respect to response to trastuzumab therapy in metastatic breast cancer. Arch Pathol Lab Med. 2008 Oct;132(10):1635–47.

23. Hofmann M, Stoss O, Gaiser T, Kneitz H, Heinmöller P, Gutjahr T, et al. Central HER2 IHC and FISH analysis in a trastuzumab (Herceptin) phase II monotherapy study: assessment of test sensitivity and impact of chromosome 17 polysomy. J Clin Pathol. 2008 Jan;61(1):89–94.

24. Memon R, Prieto Granada CN, Harada S, Winokur T, Reddy V, Kahn AG, et al. Discordance Between Immunohistochemistry and In Situ Hybridization to Detect HER2 Overexpression/Gene Amplification in Breast Cancer in the Modern Age: A Single Institution Experience and Pooled Literature Review Study. Clinical Breast Cancer. 2022 Jan 1;22(1):e123–33.

25. Gibbons-Fideler I sanna, Nitta H, Murillo A, Tozbikian G, Banks P, Parwani AV, et al. Identification of HER2 Immunohistochemistry-Negative, FISH-Amplified Breast Cancers and Their Response to Anti-HER2 Neoadjuvant Chemotherapy. American Journal of Clinical Pathology. 2019 Jan 7;151(2):176–84.

26. Reddy J, Reimann J, Anderson S, Klein P. Concordance between central and local laboratory HER2 testing from a community-based clinical study. Clinical breast cancer. 2006;7(2):153-157.

27. Press MF, Sauter G, Bernstein L, Villalobos IE, Mirlacher M, Zhou JY, et al. Diagnostic evaluation of HER-2 as a molecular target: an assessment of accuracy and reproducibility of laboratory testing in large, prospective, randomized clinical trials. Clin Cancer Res. 2005 Sep 15;11(18):6598–607.