Reviewer's report

Title: Validation of HER2 testing with core needle biopsy specimens from primary breast cancers in terms of interobserver reproducibility and concordance with surgically resected specimens

Version: 1 Date: 24 May 2010

Reviewer: Cathy Moelans

Reviewer's report:

Major compulsory revisions

1. Figures 1 and 2 need to be drastically improved! As 2+ scores are already difficult to assess, it is of utmost importance that these figures are of adequate quality.

2. Materials and Methods, FISH: since the authors clearly mention that they performed FISH testing according to ASCO/CAP guidelines, the cut-offs used in this study were not entirely according to the guideline. The guideline considers 2.2 as equivocal and not as amplified as the authors state. Amplification according to the guideline is defined as a >2.2 ratio. The authors should therefore see whether there were cases with an exact 2.2 ratio and if so, make the necessary adjustments. It is however quite strange that, in Table 3 legend, 2.2 was considered equivocal and not amplified (as in the guideline), but here an 1.8 ratio was considered negative while the guideline considers 1.8 as equivocal. In the M&M, this was the other way around and thus, as in the guideline. Please clarify.

3. Please explain why the author’s think that the HER2 FISH amplification percentage in CNB (33%) as well as in surgical specimens (48%, although in a smaller group) is so high? The IHC 3+ percentage in CNB (21%) and surgical specimens (25%) is also quite high but still in the “acceptable range”. Could there be some selection bias? In this point of view, the authors should add clinicopathological data (age/stage/histological type). Were the cases consecutive?

4. Please calculate confidence intervals for kappa scores. I trust kappa was calculated always as weighted kappa?

Minor essential revisions

1. Abstract, results section: change “…92% (kappa=0.88) for 3 x 3 categories” into “92% (kappa=0.88) for 2 x 2 categories”

2. English style should be improved, further some small textual errors:
   a. Introduction: “However, it would be expected that examination of CNB specimens alone might result in a proportion of false-positive AND/OR false-negative results, because CNB samples REPRESENT only part of the tumor, notwithstanding the possible … “
b. Introduction: “Therefore, we examined the levels of interobserver agreement for HER2 status DETERMINATION in both CNB specimens and corresponding …”

c. Introduction: “On the basis of theSE results, we EVALUATED the utility and challenges of HER2 testing withIN CNB specimens.”

d. Materials and Methods: tissue samples: “HER2 testing is performed very frequently for routine diagnosTICS and/or …”

e. Materials and Methods, interobserver agreement: “The percentage OF agreement was calculated as FOLLOWS: …”

3. Materials and Methods, tissue samples: Were IHC and FISH assays performed on the same CNB’s/surgical specimens performed on parallel slides?

4. Materials and methods, IHC: HercepTest II: what means II ???

5. Materials and methods, IHC: three observers independently evaluated IHC results: what was their experience (pathologists?, experienced technicians?) and were the observers blinded from FISH results?

6. Materials and Methods, FISH: again, blinded from IHC results and results of corresponding CNB or surgical resection specimen?

7. Materials and Methods, FISH: the authors state that after FISH re-testing of FISH equivocal cases, the cut-off of 2.0 was used to define amplification. How many FISH equivocal cases were found in this study and how many of these were defined as amplified after re-testing?

8. Results: title “Comparison of IHC AND FISH test results between CNB and surgically resected specimens”

9. p10: “Overall, consensus HER2 IHC …specimens in 13/100 cases (Table 5)”

10. p12, first line. Please state interobserver agreement for case 54, which is not at all of borderline nature (3+ and 0). Maybe the paper would benefit from adding the interobserver agreement or disagreement to Table 5.

11. p12, line 6-10: the first case was not traceable in tables 1 and 2, so I assume there was no IHC or FISH interobserver disagreement? The second case was however traceable as case 16? The fact that there was interobserver disagreement for this case should be mentioned here.

12. p12, correlation between IHC and FISH results: Since institute C has 82% amplification in the CNB IHC 2+ cases and 80% in the surgical specimen 2+ cases, this reviewer would be very interested in the comparison of processing protocols between all three institutes. Although the authors mention in the Discussion that there is an unusual and unexplainable immunoreaction in institute C specimens, this reviewer would be very interested in their annual percentages of IHC and FISH 2+/3+ positivity. Does this institution participate in external quality assessment and how was their performance? As a result of this percentage, I would even advise this institution not to perform primary IHC testing.

13. p12, last line: specimens from institution A, the rate of HER2 gene
amplification IN IHC 2+ CASES was 0% (0 of 3),

14. p15, first paragraph: “…. FISH, chromogenic in situ hybridization (CISH), or single hapten in situ hybridization (SISH) would …”: do the authors mean silver-enhanced in situ hybridization by SISH? This reviewer has not heard of single hapten in situ hybridization and knows a hapten as a small molecule such as digoxigenin that can elicit an immune response when attached to a large carrier. Do the authors refer to a specific single hapten method or do they mean silver in situ hybridization? Also, there are now dual-color dual-hapten methods (DDISH, R. Tubbs, Breast cancer symposium 2009, abstract 73) too.

Furthermore, the authors should give references that have examined these techniques in CNB vs surgical resection specimens (for example Shousha et al for SISH, Evaluation of automated silver-enhanced in situ hybridization (SISH) for detection of HER2 gene amplification in breast carcinoma excision and core biopsy specimens, Histopathology 2009 and Wolf et al for CISH, Journal of Clinical Oncology, 2009 ASCO Annual Meeting Proceedings (Post-Meeting Edition). Vol 27, No 15S (May 20 Supplement), 2009: e22141). The authors state in the Discussion and Conclusion that, to improve reliability of HER2 testing on CNB, the introduction of judgement by multiple observers could be one possible solution. Given the high amount of amplification by FISH in 2+/3+ IHC discrepancies in this study, would the authors then suggest to assume 3+ as the overall IHC score in case of such a disagreement?

16. There is one recent relevant reference lacking (examining FISH on CNB vs surgical specimens): D’Alfonso et al, AJSP 2010, Accurately assessing her-2/neu status in needle core biopsies of breast cancer patients in the era of neoadjuvant therapy: emerging questions and considerations addressed. Furthermore, the paper of Moelans et al, Cell Oncol 2009, HER-2/neu amplification testing in breast cancer by Multiplex Ligation-dependent Probe Amplification in comparison with immunohistochemistry and in situ hybridization, showed that, although some studies have suggested that the validity of IHC score 3+ in core biopsies is limited [Taucher et al 2004], reporting high rates of false positives (19.3%), there was only a slightly higher percentage of IHC 3+ positivity in biopsies compared to resections, but this did not reach statistical difference. This is in line with this study (25% vs 21%).

17. The authors state that quality assessment is very important in current HER2 diagnostics but references confirming that statement are lacking.

18. Conclusion: “In most of specimens with equivocal IHC results, accurate HER2 status was known DETERMINED by retesting WITH FISH.

19. Figure 1 legend: Cases with CONcordant (and not discordant) judgment ….

20. Table 1 legend:” … regarding the results of HER2 immunohistochemistry”, furthermore the legend lacks abbreviations such as IHC and FISH used in the table.

21. Table 2 legend: same comments as Table 1. Is there a specific reason why S17 FISH was not performed, especially since this table and paper focuses on interobserver and FISH/IHC agreement? Why does case S26 have an *, this is not mentioned in the legend, perhaps loss?
22. This reviewer is quite puzzled about case 54. As seen in Table 2, there is an IHC discrepancy in the surgically resected specimen between the three observers with 2 observers scoring 0 (no staining at all) and one observer scoring a 2+, which is quite a difference! Was this re-evaluated and if so, what was the reason for this relatively big difference? If not, please re-evaluate this case, at least for IHC. Furthermore, in Table 5, case 54 has an IHC 3+ result on the CNB, and as I could not trace this case in Table 1, I assume there was no interobserver discrepancy? Given the negative FISH result (this case was not mentioned in Table 3, so I assume there was no interobserver agreement), this would indicate a false positive IHC result, do the authors have an explanation for this? As already mentioned in comment 13, this table would benefit from adding interobserver agreement/disagreement.

23. Table 3 legend: “… results of HER2 fluorescence in situ hybridization”. Furthermore, why was the second FISH count for S67 not done? Was this because of fading? If so, then this would indicate an advantage of chromogenic techniques such as CISH and SISH over FISH.

24. Table 4 legend: “… core needle biopsy and CORRESPONDING surgically resected specimens”

25. Table 6 legend: “… HER2 FISH results of between core needle biopsy…”. Why do the authors use 2.0 as cut-off in this table? This is in contrast to the cut-off mentioned in Materials and Methods (2.2) and to their legend that says “according to the 2007 ASCO/CAP guideline”. Which cut-off was used in Table 7 (which says FISH “positive” or “negative”)?

Discretionary revisions
1. Change title into “HER2 testing on core needle biopsy specimens from primary breast cancers: interobserver reproducibility and concordance with surgically resected specimens
2. Materials and Methods, FISH: “and/or was discordant among the three observers…” Since the paper focuses on HER2 testing, “with regard to HER2 scoring” can be removed from this sentence.
3. Results: Interobserver agreement levels for IHC and FISH. Remove “levels”.
4. p13, line 4 and 7: “IHC scores of between 1+/2+ and 2+/3+, …”
5. p14, paragraph 3: “The disagreement in the results obtained with IHC BETWEEN CNB AND SURGICAL SPECIMENS appeared to be derived … “

Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

Statistical review: No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests:

I declare that I have no competing interests