Reviewer's report

Title: Meta-analysis confirms BCL2 is an independent prognostic marker in breast Cancer

Version: 2 Date: 21 December 2007

Reviewer: Marianne Paesmans

Reviewer's report:

After having read with interest your report, I have the following comments to submit for your consideration.

Comments leading to major revisions:

One essential aspect when conducting a meta-analysis is to identify all the results (ideally published or not) linked to the asked question and to try to include all the information collected as far as possible to avoid a selection bias. I have several concerns related to this issue:

- The reports search was mentioned to have been done only using a PubMed search. Is this enough? It is generally recommended to use other search tools (like consultation of another data base, consultation of the bibliography of the selected references, consultation of experts, ...)

- The authors discarded publications with less than 100 patients included in the analysis arguing about a publication bias. This is true (and in particular in the field of prognostic factors) that negative results are less likely to be published than positive ones (and negative ones have probably smaller sample sizes). However, I would have much preferred to see all publications included to reach exhaustivity with an assessment of the possible publication bias and a sensitivity analysis after exclusion of same papers than to see the approach adopted by the authors. At least, a descriptive analysis of these reports (number of patients, individual result) should be provided.

- Finally and this is the most important comment related to this issue: 36 publications were excluded because of absence of publication of an estimate of the HR although there are some methods to estimate HR from other published statistical results (Parmar et al) and only 18 studies were used for estimation of a combined HR. This appears really critical to me and jeopardizes to my opinion the validity of the results; publication or not of HR is a publication bias in itself. At least, a descriptive analysis of these reports (number of patients, individual result) should be provided.

- It might be very useful to have a more detailed flow chart of potentially eligible studies and excluded studies with reason of exclusion.

Aggregation of the results

Both unadjusted HR and adjusted HR have been aggregated. However, adjusted
HR in the individual publications have been adjusted for various covariates. There is therefore heterogeneity between the HRs that have been combined. Is it then really pertinent to carry out the aggregation? Can we recognize the heterogeneity by using only random-effects model? Also, one might guess that adjusted estimates of HR are more likely to be published when they are significantly different than 1. Another important publication bias issue. At least, more details should be provided. In view of these concerns, I don't think that we can accept the claim that the present meta-analysis report strongly supports the independent prognostic value of BCL2.

Populations of patients

The reports that have been included in the review might be related to populations of node-negative patients, node-positive patients, metastatic patients. However, the impact of the meta-analysis results in terms of therapeutic choices might be different. If it is probably difficult to perform a subgroups analysis distinguishing node-negative and node-positive patients, it could be done with a distinction of studies including only metastatic patients. What is the meaning of the statement that "no obvious difference" was present for the effect of BCL2 in node negative and node positive patients?

Comments leading to minor revisions:

Why are false positive studies the result of a combination of low statistical power and publication bias?

Page 10, if the Q statistic for heterogeneity is 17.2 with 10 df, the p value is 0.07 (as reported in Table 3) and heterogeneity might be a concern

Fig 2 reports on the aggregation of 8 studies why it is 7 in the text. This should be clarified (although we know that the number of studies and the number of reports were not the same, this is a little bit confusing)

Table 1: median follow-up: unit of time should be reported, otherwise, very confusing, number of patients does not provide the power of the study, the number of events might be preferable

Table 2: I would adopt a policy of reporting 2 decimals for HR estimates and bounds of confidence intervals

Comments leading to discretionary revisions:

The wording "upper CI and lower CI" could be changed into lower bound of CI and upper bound of CI

Page 10: the meaning of the sentence "the adjusted HR estimates were generally close to unity and ranging from 1.10 to 3.26 and only five were statistically significant in 950 cases" is not clear to me

In figures 1 and 2, it might be specified that the combined HR reported are those from the random effects models
What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
I declare that I have no competing interests