Author's response to reviews

Title: Proportional odds ratio model for comparison of diagnostic tests in meta-analysis

Authors:

Mir S Siadaty (mirSiadaty@virginia.edu)
Jianfen Shu (js8jv@virginia.edu)

Version: 2 Date: 17 November 2004

Author's response to reviews: see over
Responses to questions of referee 1 (Dirk Stengel):
Referee 1: http://www.biomedcentral.com/imedia/1078607947480157_comment.pdf
Thank you for your comments.

You asked about length of the paper, and its redundancy with our previous papers:
We have rearranged and rewritten sections of the paper, to reduce its size. We have omitted text, graphs, and tables, to reduce overlaps between our recent published papers.

You asked “I am confused about some discrepancies between the present and previous results. For example, referring to Table 4 in the original meta-analysis published in Clin Chem, the Auto Dimertest, the Dimertest Gold EIA, the Nephelotex, and others showed a relative DOR >1 compared to VIDAS, whereas in the current comparison, all of these tests fared significantly worse. Did I make a mistake?”

The discrepancy in results is mainly due to fitting slightly different models. In other words we do expect different results by fitting different models. In the Clin Chem paper, the Disease*Paper effect is dropped to allow estimation of the three covariate effects. In the current analyses we kept the Disease*Paper effect while dropping the covariates. The incomplete structure of data (with majority of cells missing) does not allow estimation of all the effects simultaneously. For the Clin Chem we have used the model (6) published in the JCE paper

$logit(\text{Result}_{pt}) = \beta_0 + \beta_1*\text{Disease}_{pt} + \beta_2*\text{Test}_{pt} + \beta_3*\text{Prvlnc}_{pt} + \beta_4*\text{Gold}_{pt} + \beta_5*\text{Setting}_{pt} + \beta_6*\text{Disease}_{pt}*\text{Test}_{pt} + \beta_7*\text{Disease}_{pt}+\text{Prvlnc}_{pt} + \beta_8*\text{Disease}_{pt}*\text{Gold}_{pt} + \beta_9*\text{Disease}_{pt}+\text{Setting}_{pt}$. For the current analyses we used

$logit(\text{Result}_{pt}) = \beta_0 + \beta_1*\text{Disease}_{pt} + \beta_2*\text{PaperID}_{pt} + \beta_3*\text{Disease}_{pt}+\text{PaperID}_{pt} + \beta_4*\text{TestID}_{pt} + \beta_5*\text{Disease}_{pt}+\text{TestID}_{pt}$.

Also, please note some of your readings are not consistent with the papers. In the 1st version of the current paper, under the new model, Nephelotex is NOT significantly different from VIDAS (look at Figure 4). Note Table 3 is comparing each test with the Overall LOR, NOT with VIDAS.

You asked “The numerous symbols contained in complex graphs (for example, the funnel, Galbraith and L’Abbe plot) are impossible to distinguish. Figures transporting virtually similar information should be deliberately deleted.”

When trying to shorten the paper, we have omitted the figures.

“The authors should add the graphical analysis of the POR assumption suggested on p12, which may help readers familiar with proportional hazards to keep on the track.”

It is Figure 5 in the revised (2nd) version of the paper we are submitting.
“The notation of the given equations is unconventional, but suits the former provided in the Journal of Clinical Epidemiology. However, if I remember my statistics correctly, the general mixed effects model formula is notated as \( \hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u_i + e_i \), with \( u_i \) and \( e_i \) being normally distributed.”
We have slightly modified the notations. Please note we are NOT proposing a general mixed effects model.

“Yet, the assumption of a “general applicability” of the proposed methods is not justified (see title and conclusions).”
We have rewritten a section “2.3. Expanding the Proportional Odds Ratio Model” where we explain how the domain of the POR model may be expanded. Also, note we have omitted the phrase “POR generalizes ...” from the title and conclusion.

“The authors stress the suitability of the POR for therapeutic meta-analyses, which has been demonstrated by Whitehead and co-workers (Whitehead A, Omar RZ, Higgins JP, Savaluny E, Turner RM, Thompson SG. Meta-analysis of ordinal outcomes using individual patient data. Stat Med 2001;20:2243-2260).”
The paper by Whitehead et al is using the ‘proportional odds’ model, explained in McCullagh 1980. Please note the difference between “proportional odds model” and “proportional odds RATIO model”. In the first, there are two “odds” and the model assumes they are proportional, in other words, that the ratio of the two odds is constant across levels of a variable. In the second, there are two “odds ratios” and the model assumes the “ratio of odds ratios” is constant. This has the advantage of eliminating assumption of “OR-homogeneity”.
While there is ample literature about proportional odds model, we found no literature for modeling ratio of odds ratios. Would you please inform us of your possible encounters with papers explaining Proportional Odds Ratio model?

“Conventional aggregation of a set of diagnostic studies by SROC fails in case of heterogeneity. Although I am not a promoter of the diagnostic odds ratio, I love the idea of handling the differences in LOR without making assumptions about OR0. There are valid points in this paper, that will become clearer if the authors get rid of the ballast.”
We thank the reviewer for his effective suggestion. We have rearranged, rewritten, or omitted several sections of the paper, both to shorten the paper and to make it clearer.
Responses to questions of referee 2 (Karel G Moons):
Referee 2: http://www.biomedcentral.com/imedia/5366568095020602_comment.pdf
We thank you for your interesting and constructive questions.

1. The authors focus very much on the use of the diagnostic odds ratio. There are some limitations on this method as e.g. described in the paper by Glas et al (J Clin Epidemiol 2003). Notably, that the summary estimates of the sensitivity and specificity are not directly available. What are the comments of the authors on this?

   Given an OR, there are a range of pairs of [sensitivity, specificity] that correspond to it. This means an OR does not characterize a specific pair (this is our rewording for your question). However, please note the way we implement our model is via cell counts of the 2-by-2 tables. One may choose to extract the OR for each test from each paper and use it in the meta-analysis (as in a random-effects model assuming LOR being normally distributed), but we are not doing that. The result is that our model contains sufficient information to estimate measures like sensitivity, specificity, etc separately. To demonstrate how our model can estimate such measures, consider the following logistic regression:

   \[
   \text{logit}(\text{Result}) = \beta_0 + \beta_1 \text{Disease} + \beta_2 \text{PaperID}
   \]

   Then we have \( \log(\text{true positive/false negative}) = \beta_0 + \beta_1 + \beta_2 \text{PaperID} \). Substituting a value for the covariate (here PaperID) such as a modal or average value, and using the model estimates for the betas, one gets the log-odds. Then one exponentiates it to get the TP/FN, call it \( Q \). Now it is easy to verify that sensitivity = \( Q/(1+Q) \).

   Likewise we have \( \log(\text{false positive/true negative}) = \beta_0 + \beta_2 \text{PaperID} \), that we call = \( \log(W) \). Then specificity = \( 1/(1+W) \).

   Although OR does not distinguish between the two types of diagnostic mistakes, our model does, as it enters them separately (using the indicator variable Disease). One can easily apply separate weights to the log(true positive/false negative) and log(false positive/true negative), to balance the true positive and false positive rates for decision making in clinical practice.

2. Given comment 3, I also wonder what the authors think of the method for direct pooling of sensitivity and specificity using bivariate models, as described for trials by van Houweling H in Stat Med 2002 and 1993. This bivariate method for diagnostic meta-analysis is also discussed at the Cochrane meetings by Reitsma JB et al. I wonder to what extent the authors’ method corresponds or differs to this bivariate approach, and what the pros and cons of both are?

   The pooling method by Mantel-Haenszel assumes OR-homogeneity. The random-effects model tries to relax this assumption. The way the random-effects model is usually implemented is by extracting OR from each paper, and assuming LOR being normally distributed. Then the distinction between the two types of mistakes (FNR and FPR, or equivalently TPR and FPR) is lost (as you asked in your first question), since one enters the LOR as datapoints into the model. The bivariate model by Houwelingen et al tries to
fix this, by entering two datapoints into the model for each test from each paper. The model we are proposing uses the cell counts of the 2-by-2 tables as datapoints. This means the two types of mistakes are included in the model (as we demonstrated responding to your first question). Moreover, the proposed model improves a few other flaws/disadvantages that the bivariate model couldn’t.

1. The method (of weighting) to incorporate study sample sizes into the model is not optimal, as shown by Mosteller & Chalmers [Statistical Science, 7, 227-236, 1992]. We use a grouped binary data structure, hence incorporating the study sample sizes into the model quite naturally.

2. The assumption of normality might not be realistic. We use a binomial distribution, which is a more realistic assumption. The way they try to relax the normality assumption is via mixture models (either non-parametric or a mixture of two normals). However they realize that it is an ill-posed problem and reliable estimates are hard to obtain. We have demonstrated how to use a marginal approach to estimate our proposed model, where the estimation is more stable and the software is easier to work with.

3. The effective sample size for a meta-analysis by the bivariate model is the number of papers included, which is usually quite small. There is a great danger for overfitting. And the number of explanatory variables one could include in the model is very restricted. Since we use the grouped binary data structure, the patients are the effective sample size, hence much bigger degrees of freedom. In the case where the meta-analysis is trying to compare several diagnostic tests, where each test is studied in one or more papers, where each paper has studied a few tests but not all, one has to deal with ‘very’ incomplete data structure with lots of missings. The value of the proposed model to expand each paper to its original sample size is more realized then.

Note our proposed model incorporates risk of events in the control group via a predictor, such as observed prevalence, hence a ‘control rate regression’.

Also note the marginal implementation of the proposed model handles measures that are not independent (via the GEE).

3. The paper includes a toolkit (including syntaxes) for performing a diagnostic meta-analysis using their method, plus additional explanation. However, the paper is very long and therefore hard to read and keep on track. It would be appealing if the authors could reduce the paper by 10 pages?

We have rearranged, rewritten, or omitted several sections of the paper, both to shorten the paper and to make it clearer.

4. Furthermore, I do not understand why the authors split up the two examples in section 2.4 and 3.2. It would be more clear to me if first the theory would be
explained (at once). And finally, the entire approach illustrated by the same clinical example study. So why not integrating 2.4 and 3.2.

We have integrated the subsections.