## PEER REVIEW HISTORY



This paper was submitted to the BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | What is the best strategy for investigating abnormal liver function tests in primary care? Implications from a prospective study. |
| **AUTHORS** | Lilford, RJ; Bentham, Louise; Armstrong, Matthew; Neuberger, James; Girling, Alan |

## VERSION 1 - REVIEW

| | |
|---|---|
| **REVIEWER** | Ryder, Stephen<br>Nottingham |
| **REVIEW RETURNED** | 28-Jan-2013 |

| | |
|---|---|
| **GENERAL COMMENTS** | The changes to the manuscript have considerably improved it. The methodology section now is much more comprehensive and clear. The authors now have discussed other studies in the area in more detail in the discussion. The implications and outcomes are clear.<br><br>I still hate the title. Whats wrong with something anyone can understand such as "Whats the best strategy for investigating abnormal liver function tests in primary care?" Parsimonious strategies sound like something the Archbishop of Westminster would be better placed to think about. |

| | |
|---|---|
| **REVIEWER** | McLernon, David<br>University of Aberdeen, Division of Applied Health Sciences |
| **REVIEW RETURNED** | 05-Feb-2013 |

| | |
|---|---|
| **GENERAL COMMENTS** | 1. Originality - does the work add enough to what is already in the published literature? If so, what does it add? If not, please cite relevant<br>references.<br><br>As the study highlights in the introduction, there are no prospective primary care studies investigating abnormal LFTs from a full panel. Therefore the work is original.<br><br>2. Importance of work to general readers - does this work matter to clinicians, patients, teachers, or policymakers? Is a general Journal the right place for it?<br><br>I think this work will be of interest to primary care, health services researchers and policymakers interested in liver disease, |

| | |
|---|---|
| | hepatologists, and, since LFTs are measured for so many reasons, it should interest clinicians in general. The BMJ therefore seems appropriate. The article concludes that for the purposes of excluding liver disease, the LFT panel can be restricted to ALT and ALP. This will be much more cost effective compared with testing the whole panel of analytes given the frequency of testing. The article is therefore important with regards to saving money.<br><br>3. Research Question - clearly defined and appropriately answered?<br>The research question is clearly defined and appropriately answered.<br><br>4. Overall design of study - adequate ?<br>I see that this study has previously been reviewed and resubmitted following appeal. The changes made improve the paper. The methods used seem appropriate. However, I have some queries on the statistical methods mainly for clarification, so they should be relatively minor requests:<br><br>The analysis of analyte concentrations is a little unclear with regards to the 'hierarchical stepwise technique' that was performed. To me, this suggests that sets of independent variables were subjected to a stepwise technique separately before amalgamating the significant variables from each set into one model. What 'hierarchies' were used if this was the case? If it was not done like this, what do the authors mean by hierarchical stepwise? Was the stepwise regression done manually or automatically? The methods state that 1211 out of 1290 had a complete set of patient characteristics and were included in the analysis. However, I assume some of these 1211 patients had missing LFTs as well, meaning further exclusions from the analysis. Therefore, it might be useful to add a column to Table 4 with the total proportion and number of patients used in each of these models. Multiple imputation was performed only for the diagnostic potential part of the analysis. Why wasn't the multiple imputation ICE method applied to this analysis?<br><br>Regarding the diagnostic potential methodology, the authors state that they used 'stepwise logistic discrimination' to determine the best combinations of patient characteristics and analyte concentrations to distinguish between the non-specific diagnostic group and each of the three main liver disease groups. I have never heard of this method and a quick web search does not come up with anything either. Is this a term the authors have come up with themselves? It seems like stepwise logistic regression to me but I may be wrong. The article needs more information as to what this is.<br><br>5, Participants studied - adequately described and their conditions defined?<br>Yes, adequate and complete<br><br>Patients were recruited across eight primary care practices in Birmingham and three in Lambeth. Patients who had obvious or pre-existing liver disease were excluded. Those with one or more abnormal LFTs were included. Eight analytes were measured at the initial GP visit and then at the first follow-up session along with other appropriate data.<br><br>It is a shame however that patients with all normal analytes were not included as a control group and the authors do admit this in the discussion. Sensitivity and specificity of LFTs would have been a |

| | |
|---|---|
| | welcome addition.<br><br>Was there any difference in the way the analytes were measured in the laboratories i.e. were different lab systems used which could cause heterogeneity in analyte concentrations between practices?<br><br>6. Results - answer the research question? Credible? Well presented?<br><br>With the exception of the pending response to the above relatively minor methodological queries, the results appear to be credible. The authors highlight the weaknesses of the study in the discussion.<br><br>7. References – up to date and relevant? Any glaring omissions?<br><br>The authors appear to have cited relevant up to date references.<br><br>8. Abstract/summary/key messages/This week in BMJ - reflect accurately what the paper says?<br><br>The abstract and key messages reflect the content of the paper accurately. A summarising concluding paragraph at the end of the discussion would end the article off nicely.<br><br>Minor comments<br><br>On page 12, line 49, can the authors be more specific as to where '(see below)' refers to?<br><br>In the discussion, page 14 lines19-20, the recent record-linkage referenced looked at the PPV for mortality, not liver disease. Can the authors add this to the sentence?<br><br>In the discussion, page 14 lines 51-52, It states that 'in contrast to Donnan et al's record linkage study... GGT had a very high FP rate'. Donnan et al found a high FP rate for all the LFTs including GGT, although GGT had the best PPV. Perhaps the authors could revise this sentence as it reads as if Donnan et al found a high PPV for GGT which is not true. |

| | |
|---|---|
| **REVIEWER** | Rosenberg, William<br>University College London, Institute for Liver and Digestive Health |
| **REVIEW RETURNED** | 28-Feb-2013 |

| | |
|---|---|
| **GENERAL COMMENTS** | This is an interesting and well presented study that makes a significant contribution to the literature. The revised manuscript has greater clarity than the initial version. |

**VERSION 1 – AUTHOR RESPONSE**

Reviewer 1 – We thank the reviewer for the positive comments and have changed the title of the journal, as suggested, to make it more accessible.

Reviewer 2 – We thank the reviewer for their helpful comments. Regarding point 4, the reviewer queries the conduct of the analysis of analyte concentrations (p8), in particular the "hierarchical stepwise technique". However, the "technique" consists of the whole process explicitly described in the appropriate paragraph on p8. It was not a separate or additional step in the analysis. Our wording seems to have led to some confusion here, for which we apologise. We have made some changes to this paragraph to clear up this point. An extra column has been added to table 4 as requested, detailing the numbers of complete cases. Multiple imputation was not attempted here (a) because of the exploratory nature of the analyses; (b) because of the computational complexity involved in implementing the hierarchical procedure; and (c) given the high coverage achieved by complete cases (i.e. around 90% of the sample for each analyte). Conceptually, these analyses constitute a step on the path towards the discriminant analyses described under "Diagnostic Potential" and in which an imputation technique was applied. The reviewer is correct in surmising that "stepwise logistic discrimination" just means "stepwise logistic regression", and we have made this change.
Under point 5, there were three laboratories in all and the large majority of cases (82.4%) were analysed by a single laboratory. While there were no declared differences in lab procedures, we were aware of this possibility. In fact the analyses reported here have been adjusted, as a matter of course, for (multiplicative) laboratory effects. This is now reported.
We have also dealt with the minor comments the reviewer noted.

Reviewer 3 – We thank the reviewer for the positive comments.

In addition, we have included previous reviewer comments from the first submission (pages 5-11) and our reply (12-13) in the uploaded 'Review History' file; as well as the full NIHR Health Technology Assessment (HTA) report that has been accepted for publication (pending publication of this paper in a journal).

In conclusion, we very much hope you like this paper and look forward to hearing from you in due course.

Specific reviewer comments addressed above

Reviewer 2
…
4. Overall design of study - adequate ?
I see that this study has previously been reviewed and resubmitted following appeal. The changes made improve the paper. The methods used seem appropriate. However, I have some queries on the statistical methods mainly for clarification, so they should be relatively minor requests:

The analysis of analyte concentrations is a little unclear with regards to the 'hierarchical stepwise technique' that was performed. To me, this suggests that sets of independent variables were subjected to a stepwise technique separately before amalgamating the significant variables from each set into one model. What 'hierarchies' were used if this was the case? If it was not done like this, what do the authors mean by hierarchical stepwise? Was the stepwise regression done manually or automatically? The methods state that 1211 out of 1290 had a complete set of patient characteristics and were included in the analysis. However, I assume some of these 1211 patients had missing LFTs as well, meaning further exclusions from the analysis. Therefore, it might be useful to add a column to Table 4 with the total proportion and number of patients used in each of these models. Multiple

imputation was performed only for the diagnostic potential part of the analysis. Why wasn't the multiple imputation ICE method applied to this analysis?

Regarding the diagnostic potential methodology, the authors state that they used 'stepwise logistic discrimination' to determine the best combinations of patient characteristics and analyte concentrations to distinguish between the non-specific diagnostic group and each of the three main liver disease groups. I have never heard of this method and a quick web search does not come up with anything either. Is this a term the authors have come up with themselves? It seems like stepwise logistic regression to me but I may be wrong. The article needs more information as to what this is.

5, Participants studied - adequately described and their conditions defined?
Yes, adequate and complete

Patients were recruited across eight primary care practices in Birmingham and three in Lambeth. Patients who had obvious or pre-existing liver disease were excluded. Those with one or more abnormal LFTs were included. Eight analytes were measured at the initial GP visit and then at the first follow-up session along with other appropriate data.

It is a shame however that patients with all normal analytes were not included as a control group and the authors do admit this in the discussion. Sensitivity and specificity of LFTs would have been a welcome addition.

Was there any difference in the way the analytes were measured in the laboratories i.e. were different lab systems used which could cause heterogeneity in analyte concentrations between practices?