Impact of Contacting Study Authors on Systematic Review Conclusions: Diagnostic Tests for Hepatic Fibrosis
Research White Paper

Impact of Contacting Study Authors on Systematic Review Conclusions: Diagnostic Tests for Hepatic Fibrosis

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-2012-00014-I

Prepared by:
Pacific Northwest Evidence-based Practice Center
Oregon Health & Science University
Portland, OR

Investigators:
Shelley Selph, M.D.
Roger Chou, M.D.

AHRQ Publication No. 14-EHC004-EF
April 2014
This report is based on research conducted by the Pacific Northwest Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2012-00014-I). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials that are clearly noted in the document. Further reproduction of those copyrighted materials is prohibited without the specific permission of copyright holders.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

We welcome comments on this Methods Research Project. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov.

Richard G. Kronick, Ph.D.
Director
Agency for Healthcare Research and Quality

Yen-pin Chiang, Ph.D.
Acting Director
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang M.D., M.P.H.
Director, EPC Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Elisabeth Kato, M.D., M.R.P.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality
Acknowledgments

We would like to thank the following corresponding and other authors for responding to our request for additional study data:

Alfredo Alberti, M.D., Simona Bota, M.D., Marc Bourlière, M.D., Jérôme Boursier, M.D., Rafael Bruck, M.D., Mary J. Burton, M.D., Paul Calès, M.D., Laurent Castéra, M.D., Ramsey C. Cheung, M.D., Kin Jip Cheung, Ph.D., Dana Crisan, M.D., Ph.D., Mireen Friedrich-Rust, M.D., Naveen Gara, M.D., Edoardo G. Giannini, M.D., Ph.D., Mircea Grigorescu, M.D., Ph.D., Jérôme Guéchot, Pharm.D., Ph.D., Kwang-Hyub Han, M.D., Philippe Halfon, M.D. Ph.D., Pharm.D., Angelo Iacobellis, M.D., Dong Joon Kim, M.D., Sheng-Nan Lu, M.D., M.P.H., Ph.D., Gilles Morali, M.D., Joel Mossong, PhD., Rokaya Mohamed El-Sayed, M.D., Mohamed M. Omran, Ph.D., Jae Jun Park, M.D., Keyur Patel, M.D., Ph.D., Guillaume Pénaranda, M.Sc., William Rosenberg, Ph.D., Christoph Sarrazin, M.D., Leonardo L. Schiavon, M.D., Ph.D., Giada Sebastiani, M.D., Roxana Sirli, M.D., Ph.D., Roberto Testa, M.D., David L. Thomas, M.D., M.P.H., Hans Van Vlierberghe, M.D., Ph.D., Hanneke van Vuuren, Ph.D., Kentaro Yoshioka, M.D., Ph.D., Yusuf Yilmaz, M.D., and Jean-Pierre Zarski, M.D., Ph.D.
Impact of Contacting Study Authors on Systematic Review Conclusions: Diagnostic Tests for Hepatic Fibrosis

Structured Abstract

Background. In 2013, the Pacific Northwest Evidence-based Practice Center conducted a systematic review of screening and diagnostic tests for hepatic fibrosis or cirrhosis in patients with chronic hepatitis C viral infection. However, 17 of the 172 included studies reported diagnostic accuracy results that were discordant from 2x2 tables. In addition, 60 studies did not report adequate data to construct 2x2 tables or include in the analysis. This study explores the response rate and impact of contacting authors to provide data that were otherwise missing or discordant.

Objectives. To determine the efficacy and impact of contacting authors to clarify discordant data or obtain missing data for a systematic review on screening and diagnostic tests for hepatic fibrosis or cirrhosis in patients with chronic hepatitis C viral infection.

Methods. Sixty-six corresponding authors were sent letters requesting additional information or clarification of data from 77 studies that had discrepancies in the data reported or that provided insufficient data to construct 2x2 tables. Data received from authors were pooled with data included in the previous review and the diagnostic effect analyzed.

Results. Of the 66 authors, 45 (68%) were successfully contacted and 28 (42%) provided additional data for 29 studies. All authors who provided data did so by the third written request for information. Authors of more recent studies were significantly more likely to be located and provide data compared to authors of older studies. In general, inclusion of the additional 29 studies had only minor effects on the diagnostic accuracy meta-analyses. However, the additional data resulted in reclassification of the utility of three tests.

Conclusions. The results suggest that contacting authors to obtain additional data will likely be successful. However, there was no clear trend in the impact of new data on measures of diagnostic accuracy. As a result, it is unclear whether the time-intensive practice of contacting authors is worth the effort. It would be more effective to require authors of studies to provide 2x2 tables within the published manuscript for transparency and to facilitate additional analyses.
Contents

Background .................................................................................................................................................. 1
Methods ...................................................................................................................................................... 2
Results ......................................................................................................................................................... 3
   Effects on Diagnostic Accuracy .............................................................................................................. 3
Discussion .................................................................................................................................................. 10
References ................................................................................................................................................ 12
Abbreviations and Acronyms .................................................................................................................. 17

Tables
   Table 1. Diagnostic accuracy of tests for fibrosis ................................................................................. 5
   Table 2. Diagnostic accuracy of tests for cirrhosis ............................................................................ 8

Appendixes
   Appendix A. Studies Identified for Followup To Obtain Additional Data or Clarifications
   Appendix B. Sample of Letter Sent to Authors
Background

Liver fibrosis and cirrhosis refer to scarring of the liver, often due to viral hepatitis or chronic alcohol exposure. The gold standard for diagnosing liver fibrosis and cirrhosis is a liver biopsy. However, liver biopsies are not without risk, including pain, bleeding, infection, and accidental injury to a nearby organ. In addition, liver biopsies are subject to sampling errors and variability in interpretation. Blood tests that are accurate for evaluating the presence of fibrosis could spare patients the risks and discomfort involved with liver biopsy.

The Pacific Northwest Evidence-based Practice Center recently conducted systematic reviews of screening and diagnostic tests for hepatic fibrosis or cirrhosis in patients with chronic hepatitis C viral infection. We found evidence that a number of blood tests are moderately useful for identifying clinically significant fibrosis (platelet count, age-platelet index, aspartate aminotransferase-platelet ratio index [APRI], FibroIndex, FibroText, and Forns Index) or cirrhosis (platelet count, age-platelet index, APRI, and Hepascore), based on positive likelihood ratios of 5 to 10, suggesting a potential role as an alternative to liver biopsy.

However, our review of diagnostic blood tests had limitations. Out of the 172 studies included in our review, 17 provided results for measures of diagnostic accuracy that were discordant from 2 x 2 tables (i.e., number of true positives, false positives, true negatives, and false negatives) calculated based on the information provided in the studies. Although excluding such studies from the analyses had little impact on the overall conclusions, it is concerning that 10 percent of studies reported potentially incorrect data. Additionally, another 60 studies did not provide sufficient information to allow us to construct a 2 x 2 table at commonly reported cutoffs, or only provided area under the receiver operating characteristic (AUROC) analysis results, without reporting sensitivity or specificity at specific cutoffs. Results from these studies therefore could not be included in summary estimates for sensitivity and specificity, resulting in less robust and potentially biased estimates. While there is some support for contacting study authors to obtain unpublished data when conducting systematic reviews, evidence regarding the yield and impact of such efforts is sparse, particularly in the area of diagnostic tests. Research is needed to understand whether contacting study authors is worth the additional effort required, specifically with regard to how much additional information is obtained and how that information affects the conclusions of the systematic review, including estimates of diagnostic accuracy as well as the degree of confidence in the findings.

We sought to answer these questions by contacting the authors of 77 studies to request additional data and clarifications in the case of discrepancies.
Methods

We sent letters to 66 corresponding authors asking for additional information or clarification of data from 77 studies that had discrepancies in the data reported \( n=17 \) \(^{7-23} \) or that provided insufficient data to construct 2 x 2 tables at standard cutoffs \( n=60 \) \(^{9,24-43} \). See Appendix A for a list of studies included in this report and Appendix B for a sample of the letter sent to authors. We defined studies with discrepancies as those in which reported measures of diagnostic accuracy were inconsistent with measures of diagnostic accuracy calculated from 2 x 2 tables by values of \( >0.10 \) (e.g., reported positive predictive value of 0.85 vs. calculated positive predictive value of 0.70). (See “Table 3 of the Supplement. Discrepancies” published as an online supplement to the review.\(^5\)) For studies in which 2 x 2 table data were not provided, we calculated values for 2 x 2 tables for commonly reported cutoff values for a positive test, based on the reported sample size, prevalence of the condition of interest (fibrosis or cirrhosis), sensitivity, and specificity. Studies for which we could not construct 2 x 2 tables included those in which some measures of diagnostic accuracy were reported, but other necessary information was missing (e.g., sample size, prevalence of condition), studies in which sensitivity and specificity were reported at nonstandard cutoffs, and studies in which only an AUROC was reported, without sensitivity or specificity or other measures of diagnostic accuracy at specific cutoffs.

For studies with discrepancies and cases in which we could not construct a 2 x 2 table, we requested that authors provide the 2 x 2 data used to generate their estimates of diagnostic accuracy. For studies that provided only AUROC or did not report diagnostic accuracy at standard cutoffs, we asked that authors provide 2 x 2 data for diagnostic accuracy at standard cutoffs for the blood test or tests evaluated.

If there was no response to our initial email, after a minimum of 3 business days we sent a second reminder email to the corresponding author. If there was still no response after a minimum of 8 business days following the initial email, we sent a second reminder email. After a minimum of 10 business days with no response, we then attempted to contact authors by telephone. If still unable to reach corresponding authors, we attempted to contact last authors and statisticians, if identifiable. If corresponding authors forwarded our request to other authors, we sent reminders to these authors. After a minimum of 15 business days from our initial email, we sent a final email to authors. If we received an automated “out-of-office” response, we waited until the author had returned to send further reminders. In all cases, for the convenience of authors, we provided labeled 2x2 tables that they could fill in and send back to us. We tracked responses in a Microsoft Excel\(^\text{®} \) spreadsheet.

As in the original review, we considered clinically significantly fibrosis to be F2 to F4 on the Meta-analysis of Histological Data in Viral Hepatitis (METAVIR) scale, a score of 3 to 6 on the Ishak scale, or equivalent. We considered F4 on the METAVIR scale, a score of 5 to 6 on the Ishak scale, or equivalent to indicate liver cirrhosis. We recalculated median values and ranges for sensitivity and specificity at the cutoffs used in the original review using additional data obtained, and we compared differences between the updated and original findings. As in the original review, we categorized blood tests reporting a positive likelihood ratio of 5 to 10 or a negative likelihood ratio of 0.1 to 0.2 as moderately useful (no blood test was associated with a positive likelihood ratio of \( >10 \) or negative likelihood ratio \( <0.1 \)).\(^{44} \) We also re-rated the strength of evidence with the additional data, in accordance with the Agency for Healthcare Research and Quality “Methods Guide for Effectiveness and Comparative Effectiveness Reviews.”\(^{45} \)}
Results

Of the 66 authors, we were able to contact 45 (68 percent). Of those 45 authors, 28 (62 percent) provided additional data for 29 studies, including 4 who provided datasets. Among authors whom we were able to contact, reasons for not sending data included: no current access to the data and need for additional time to find and format the data (e.g., data stored on a floppy disk).

All authors who provided data did so by the third request for information (second reminder). We received information on 10 studies after only one request (34 percent). Two requests were required for 13 studies (45 percent), and three were required for 6 studies (21 percent). We received no additional information after three requests and received no additional data in response to telephone contact.

There was no difference in the likelihood of providing data between authors of studies in which 2 x 2 tables could not be calculated compared with authors of studies with discrepancies (40 percent vs. 54 percent, p=0.36). Of the 17 studies in which there was a discrepancy between reported results for diagnostic accuracy and constructed 2 x 2 tables, 7 of 13 authors (54 percent) provided data on 7 studies (41 percent). We were unable to contact four authors, and one declined telephone contact. Of the 60 studies missing information to generate 2 x 2 tables, 21 of 53 authors (40 percent) provided additional or confirmatory data on 22 studies (37 percent), including three datasets. Reasons for not providing data were similar to those for authors of studies with discrepancies. We were unable to locate 10 authors, and 2 did not have access to the data, 5 forwarded our request to a colleague who did not provide the data, or requested that we contact another author who did not provide the data, or was too busy to comply, 1 author had died, and data from 8 were still pending at the time this report was written. Authors of more recent studies were more likely to be located and provide data (p=0.02). The mean year of publication of studies for which we received additional data was 2010. The mean year of publication of studies by contacted authors who did not provide additional data was 2008, while the average publication year for authors of studies we could not locate was 2007. Country of publication did not appear to predict the likelihood of receiving data. We received data from authors based in Belgium (1 study), Brazil (1 study), Egypt (2 studies), France (6 studies), Germany (1 study), Israel (1 study), Italy (3 studies), Japan (1 study), Korea (2 studies), Luxembourg (1 study), the Netherlands (1 study), Romania (3 studies), Taiwan (1 study), Turkey (1 study) and the United States (4 studies).

Effects on Diagnostic Accuracy

For diagnosing fibrosis, additional data was provided for 12 blood tests, 5 of which were evaluated at two different cutoffs (Table 1). The number of additional studies for specific tests and cutoffs ranged from zero to nine (zero additional studies occurred when additional data were obtained, but only for studies with discrepancies, so that one set of data was replaced by another). For the Fibrotest at a cutoff of >0.70 or >0.80, the number of studies increased from 5 to 10, but there was little impact on the positive likelihood ratio (5.5 to 7.6) or the negative likelihood ratio (0.81 to 0.65). For the APRI at a cutoff of >0.5 to >0.55, the number of available
studies increased from 28 to 38 (40 samples), with little impact on median estimates. The number of additional studies ranged from zero to four for other blood tests. The largest impact on positive likelihood ratios was observed for the FIB-4 at a cutoff of >3.25, with the median increasing from 2.4 to 9.3, resulting in reclassification from less useful to moderately useful. For the negative likelihood ratio, the largest impact occurred with the FibroIndex at a cutoff of >1.25, with the median increasing from 0.15 to 0.62, resulting in reclassification from moderately useful to less useful. Although re-analysis with additional data also resulted in reclassification of the positive likelihood ratio for the age-platelet index at a cutoff of >6.0 and the negative likelihood ratio for the Fibrotest at a cutoff of >0.10 to >0.22, the actual change in the median estimates was small to minimal (5.1 to 4.5 and 0.21 to 0.17, respectively). For the Lok Index, no studies reported accuracy for diagnosis of fibrosis in the original report. Based on additional data from three studies, the median positive likelihood ratio was 2.9 (range 2.0 to 3.1) and the median negative likelihood ratio 0.53 (range 0.31 to 0.65).

For diagnosing cirrhosis, additional data was provided for eight tests, four of which were evaluated at two different cutoffs (Table 2). For the APRI at a cutoff of >1.0 or >=1.0 data was available for 10 additional studies, but the effect on median likelihood ratio estimates was minimal. The number of additional studies ranged from one to five for other blood tests and cutoffs. For the Lok Index, the negative likelihood ratio based on a cutoff of >=0.20 or >0.26 was re-classified from less useful to moderately useful, but the impact on the estimate was minimal (0.21 in the original analysis and 0.19 with additional data). Similarly, the positive likelihood ratios for the Lok Index based on a cutoff of >=0.5 or >0.6 and the AST/ALT ratio with a cutoff of >1.0 were reclassified from less useful to moderately useful, but the impact on the estimate was also relatively small (4.4 to 5.8; 4.5 to 5.6, respectively).

We compared the effects of additional data from studies with discrepancies with the effects of additional data from studies in which 2 x 2 tables could not be generated and found no clear pattern suggesting differential effects on median estimates. We also evaluated effects of additional data with respect to the original strength of evidence ratings. No test for which we obtained additional data was originally rated low strength of evidence. The two tests for which additional data resulted in the greatest changes, the FIB-4 and the FibroIndex, were both originally rated as moderate strength of evidence. For diagnosis of fibrosis, the number of studies increased from four to seven and from three to six, respectively. For tests originally rated as high strength of evidence (APRI, AST/ALT ratio, Fibrotest), new evidence had little impact on median estimates. Analyses on the effects of year of publication, study quality, country, or other factors on changes in estimates were too limited by the small number of studies available for most blood tests and cutoffs to draw reliable conclusions.

The overall strength of evidence rating did not change for any of the tests for which we obtained additional data, due to the relatively small number of studies providing additional data for most tests. All were rated as moderate or high strength of evidence in the original systematic review. We received the most additional data for the APRI, which was already rated high strength of evidence.
<table>
<thead>
<tr>
<th>Fibrosis Test (cutoff)</th>
<th>Number of Samples</th>
<th>Sensitivity (median, range)</th>
<th>Specificity (median, range)</th>
<th>Positive Likelihood Ratio (median, range)</th>
<th>Negative Likelihood Ratio (median, range)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Platelets</strong>&lt;140 to &lt;163</td>
<td>8</td>
<td>0.56 (0.28-0.89)</td>
<td>0.91 (0.69-1.0)</td>
<td>6.3 (2.3-14)</td>
<td>0.48 (0.16-0.78)</td>
</tr>
<tr>
<td><strong>With additional data</strong></td>
<td>10 b</td>
<td>0.57 (0.28-0.89)</td>
<td>0.91 (0.58-1.0)</td>
<td>6.3 (1.64-35)</td>
<td>0.48 (0.16-0.78)</td>
</tr>
<tr>
<td><strong>API</strong> &gt;3.5 ≥4.0</td>
<td>5</td>
<td>0.70 (0.52-0.82)</td>
<td>0.70 (0.51-0.77)</td>
<td>2.3 (1.7-2.7)</td>
<td>0.43 (0.34-0.67)</td>
</tr>
<tr>
<td><strong>With additional data</strong></td>
<td>6 b</td>
<td>0.64 (0.52-0.82)</td>
<td>0.69 (0.51-0.77)</td>
<td>2.0 (1.7-2.7)</td>
<td>0.53 (0.34-0.67)</td>
</tr>
<tr>
<td><strong>API</strong> ≥6.0</td>
<td>5/3 c</td>
<td>0.51 (0.19-0.75)</td>
<td>0.90 (0.58-0.96)</td>
<td>5.1 (1.8-7.3)</td>
<td>0.54 (0.43-0.94)</td>
</tr>
<tr>
<td><strong>With additional data</strong></td>
<td>6 b</td>
<td>0.54 (0.19-0.75)</td>
<td>0.88 (0.58-0.96)</td>
<td>4.5 (1.4-7.3)</td>
<td>0.52 (0.43-0.94)</td>
</tr>
<tr>
<td><strong>APRI</strong> ≥0.5 to &gt;0.55</td>
<td>28</td>
<td>0.81 (0.29-0.98)</td>
<td>0.55 (0.10-0.94)</td>
<td>1.8 (1.1-4.8)</td>
<td>0.35 (0.08-0.78)</td>
</tr>
<tr>
<td><strong>With additional data</strong></td>
<td>40 d</td>
<td>0.79 (0.29-0.98)</td>
<td>0.56 (0.10-1.0)</td>
<td>1.8 (1.0-7.5)</td>
<td>0.56 (0.07-0.93)</td>
</tr>
<tr>
<td><strong>AST:ALT Ratio &gt;1.0</strong></td>
<td>5</td>
<td>0.35 (0.08-0.45)</td>
<td>0.77 (0.62-1.0)</td>
<td>1.5 (1.1-15)</td>
<td>0.84 (0.84-0.98)</td>
</tr>
<tr>
<td><strong>With additional data</strong></td>
<td>8 b,d</td>
<td>0.36 (0.08-0.59)</td>
<td>0.80 (0.48-1.0)</td>
<td>1.7 (1.1-14)</td>
<td>0.81 (0.65-0.98)</td>
</tr>
<tr>
<td><strong>ELF &gt;8.75 &gt;9.0, or &gt;9.78</strong></td>
<td>3</td>
<td>0.85 (0.84-0.86)</td>
<td>0.70 (0.62-0.80)</td>
<td>2.8 (2.3-4.2)</td>
<td>0.21 (0.19-0.23)</td>
</tr>
<tr>
<td><strong>With additional data</strong></td>
<td>3 b</td>
<td>0.84 (0.62-0.85)</td>
<td>0.80 (0.70-0.86)</td>
<td>4.2 (2.8-4.4)</td>
<td>0.20 (0.19-0.45)</td>
</tr>
<tr>
<td><strong>FIB-4 &gt;1.26 or ≥1.45</strong></td>
<td>6</td>
<td>0.64 (0.62-0.86)</td>
<td>0.68 (0.54-0.75)</td>
<td>2.0 (0.88-2.6)</td>
<td>0.53 (0.21-1.3)</td>
</tr>
<tr>
<td><strong>With additional data</strong></td>
<td>9 d</td>
<td>0.64 (0.57-0.86)</td>
<td>0.68 (0.28-0.85)</td>
<td>2.0 (0.88-3.7)</td>
<td>0.53 (0.21-1.3)</td>
</tr>
</tbody>
</table>
Table 1. Diagnostic accuracy of tests for fibrosis\(^a\) (continued)

<table>
<thead>
<tr>
<th>Fibrosis Test (cutoff)</th>
<th>Number of Samples</th>
<th>Sensitivity (median, range)</th>
<th>Specificity (median, range)</th>
<th>Positive Likelihood Ratio (median, range)</th>
<th>Negative Likelihood Ratio (median, range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIB-4 &gt;3.25</td>
<td>4</td>
<td>0.50 (0.28-0.86)</td>
<td>0.79 (0.59-0.99)</td>
<td>2.4 (1.3-28)</td>
<td>0.63 (0.21-0.80)</td>
</tr>
<tr>
<td>With additional data</td>
<td>7(^d)</td>
<td>0.28 (0.11-0.86)</td>
<td>0.97 (0.59-1.0)</td>
<td>9.3 (1.3-28)</td>
<td>0.74 (0.21-0.89)</td>
</tr>
<tr>
<td>FibroIndex &gt;1.25</td>
<td>3</td>
<td>0.94 (0.62-0.97)</td>
<td>0.40 (0.40-0.48)</td>
<td>1.6 (1.2-1.6)</td>
<td>0.15 (0.08-0.79)</td>
</tr>
<tr>
<td>With additional data</td>
<td>6(^d)</td>
<td>0.64 (0.54-0.97)</td>
<td>0.57 (0.40-1.0)</td>
<td>1.5 (1.2-2.2)</td>
<td>0.62 (0.08-0.79)</td>
</tr>
<tr>
<td>FibroIndex &gt;2.25 or ≥2.25</td>
<td>3</td>
<td>0.30 (0.17-0.36)</td>
<td>0.97 (0.97-1.0)</td>
<td>10,12,∞</td>
<td>0.72 (0.66-0.83)</td>
</tr>
<tr>
<td>With additional data</td>
<td>4(^d)</td>
<td>0.24 (0.14-0.36)</td>
<td>0.99 (0.97-1.0)</td>
<td>10,12,∞</td>
<td>0.78 (0.66-0.87)</td>
</tr>
<tr>
<td>Fibrometer &gt;0.419 to &gt;0.59</td>
<td>3</td>
<td>0.69 (0.64-0.80)</td>
<td>0.81 (0.76-0.81)</td>
<td>3.6 (3.4-3.6)</td>
<td>0.38 (0.26-0.44)</td>
</tr>
<tr>
<td>With additional data</td>
<td>5(^d)</td>
<td>0.80 (0.64-0.87)</td>
<td>0.76 (0.64-0.81)</td>
<td>3.3 (2.4-3.6)</td>
<td>0.26 (0.21-0.44)</td>
</tr>
<tr>
<td>FibroTest &gt;0.10 to &gt;0.22</td>
<td>6</td>
<td>0.92 (0.88-0.97)</td>
<td>0.38 (0.27-0.56)</td>
<td>1.5 (1.3-1.9)</td>
<td>0.21 (0.11-0.28)</td>
</tr>
<tr>
<td>With additional data</td>
<td>9(^d)</td>
<td>0.92 (0.64-0.98)</td>
<td>0.46 (0.21-1.0)</td>
<td>1.7 (1.2-2.2)</td>
<td>0.17 (0.11-0.39)</td>
</tr>
<tr>
<td>FibroTest &gt;0.70 or &gt;0.80</td>
<td>5</td>
<td>0.22 (0.20-0.50)</td>
<td>0.96 (0.95-0.98)</td>
<td>5.5 (5.5-13)</td>
<td>0.81 (0.53-0.82)</td>
</tr>
<tr>
<td>With additional data</td>
<td>10(^b,d)</td>
<td>0.38 (0.20-0.94)</td>
<td>0.95 (0.36-0.98)</td>
<td>7.6 (1.4-13)</td>
<td>0.65 (0.12-0.82)</td>
</tr>
<tr>
<td>Forns Index &gt;4.2 to &gt;4.57</td>
<td>14</td>
<td>0.88 (0.57-0.94)</td>
<td>0.52 (0.20-0.77)</td>
<td>1.8 (1.2-2.2)</td>
<td>0.22 (0.12-0.64)</td>
</tr>
<tr>
<td>With additional data</td>
<td>16(^b,d)</td>
<td>0.89 (0.42-0.94)</td>
<td>0.51 (0.20-0.77)</td>
<td>1.8 (0.54-2.2)</td>
<td>0.23 (0.12-2.6)</td>
</tr>
<tr>
<td>Fibrosis Test (cutoff)</td>
<td>Number of Samples</td>
<td>Sensitivity (median, range)</td>
<td>Specificity (median, range)</td>
<td>Positive Likelihood Ratio (median, range)</td>
<td>Negative Likelihood Ratio (median, range)</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>------------------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>Forns Index &gt;6.9</td>
<td>10</td>
<td>0.36 (0.18-0.61)</td>
<td>0.94 (0.66-1.0)</td>
<td>6.5 (1.6-18)</td>
<td>0.68 (0.56-0.92)</td>
</tr>
<tr>
<td>With additional data</td>
<td>14(^d)</td>
<td>0.40 (0.18-0.81)</td>
<td>0.95 (0.33-1.0)</td>
<td>7.4 (1.2-18)</td>
<td>0.63 (0.22-0.92)</td>
</tr>
<tr>
<td>Hepascore &gt;0.46 to ≥0.55</td>
<td>5</td>
<td>0.66 (0.54-0.82)</td>
<td>0.79 (0.65-0.86)</td>
<td>3.1 (2.3-4.5)</td>
<td>0.43 (0.28-0.55)</td>
</tr>
<tr>
<td>With additional data</td>
<td>8(^d)</td>
<td>0.65 (0.54-0.82)</td>
<td>0.80 (0.65-0.86)</td>
<td>3.2 (2.3-4.5)</td>
<td>0.44 (0.28-0.55)</td>
</tr>
<tr>
<td>Lok Index &gt;0.17 or &gt;0.20</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>With additional data</td>
<td>3(^d)</td>
<td>0.58 (0.48-0.82)</td>
<td>0.80 (0.58-0.81)</td>
<td>2.9 (2.0-3.1)</td>
<td>0.53 (0.31-0.65)</td>
</tr>
</tbody>
</table>

ALT = serum alanine aminotransferase; API = age platelet index; APRI = aspartate aminotransferase to platelet ratio index; AST = aspartate aminotransferase; ELF = enhanced liver fibrosis.

\(^a\)Values in bold indicate a change to above or below a cutoff of 5.0 for positive likelihood ratio or 0.20 for negative likelihood ratio

\(^b\)Additional data for study(s) with discrepancy in reported data

\(^c\)5 samples for sensitivity, 3 for specificity

\(^d\)Additional data for study(s) without 2 x 2 tables
Table 2. Diagnostic accuracy of tests for cirrhosis

<table>
<thead>
<tr>
<th>Fibrosis Test (cutoff)</th>
<th>Number of Samples</th>
<th>Sensitivity (median, range)</th>
<th>Specificity (median, range)</th>
<th>Positive Likelihood Ratio (median, range)</th>
<th>Negative Likelihood Ratio (median, range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platelets &lt;140 to &lt;155</td>
<td>9</td>
<td>0.78 (0.41-0.93)</td>
<td>0.87 (0.84-0.94)</td>
<td>6.0 (2.8-93)</td>
<td>0.25 (0.07-0.63)</td>
</tr>
<tr>
<td>With additional data</td>
<td>10(^b)</td>
<td>0.77 (0.41-0.93)</td>
<td>0.86 (0.57-0.99)</td>
<td>5.5 (1.6-93)</td>
<td>0.27 (0.07-0.63)</td>
</tr>
<tr>
<td>API ≥6.0</td>
<td>5/3(^c)</td>
<td>0.67 (0.43-0.80)</td>
<td>0.87 (0.81-0.93)</td>
<td>5.2 (2.7-10)</td>
<td>0.38 (0.22-0.68)</td>
</tr>
<tr>
<td>With additional data</td>
<td>6(^b)</td>
<td>0.64 (0.12-0.80)</td>
<td>0.88 (0.81-0.99)</td>
<td>5.3 (2.7-17)</td>
<td>0.41 (0.22-0.88)</td>
</tr>
<tr>
<td>APRI &gt;1.0 or ≥1.0</td>
<td>19</td>
<td>0.77 (0.33-1.0)</td>
<td>0.75 (0.30-0.87)</td>
<td>3.1 (1.4-4.9)</td>
<td>0.31 (0-0.77)</td>
</tr>
<tr>
<td>With additional data</td>
<td>30/29(^c,d)</td>
<td>0.75 (0.13-1.0)</td>
<td>0.77 (0.30-1.0)</td>
<td>3.2 (1.4-10.6)</td>
<td>0.33 (0-0.89)</td>
</tr>
<tr>
<td>AST:ALT Ratio &gt;1.0</td>
<td>17</td>
<td>0.36 (0.12-0.78)</td>
<td>0.92 (0.59-1.0)</td>
<td>4.5 (1.0-31)</td>
<td>0.70 (0.47-1.0)</td>
</tr>
<tr>
<td>With additional data</td>
<td>19(^b)</td>
<td>0.39 (0.10-0.78)</td>
<td>0.93 (0.59-1.0)</td>
<td>5.6 (1.0-31)</td>
<td>0.66 (0.23-1.0)</td>
</tr>
<tr>
<td>FIB-4 &gt;1.45</td>
<td>1</td>
<td>0.90</td>
<td>0.58</td>
<td>2.1</td>
<td>0.17</td>
</tr>
<tr>
<td>With additional data</td>
<td>4(^d)</td>
<td>0.89 (0.87-1.0)</td>
<td>0.58 (0.40-0.70)</td>
<td>2.1 (1.7-2.9)</td>
<td>0.19 (0.0-0.23)</td>
</tr>
<tr>
<td>FIB-4 &gt;3.25</td>
<td>1</td>
<td>0.55</td>
<td>0.92</td>
<td>6.9</td>
<td>0.49</td>
</tr>
<tr>
<td>With additional data</td>
<td>5(^d)</td>
<td>0.49 (0.40-0.55)</td>
<td>0.93 (0.91-0.95)</td>
<td>6.4 (5.7-8.9)</td>
<td>0.60 (0.49-0.63)</td>
</tr>
<tr>
<td>FibroTest &gt;0.56 or &gt;0.66</td>
<td>2</td>
<td>0.85 &amp; 0.82</td>
<td>0.74 &amp; 0.77</td>
<td>3.3 &amp; 36</td>
<td>0.20 &amp; 0.23</td>
</tr>
<tr>
<td>With additional data</td>
<td>7(^d)</td>
<td>0.83 (0.27-0.91)</td>
<td>0.74 (0.65-1.0)</td>
<td>3.6 (2.6-3.6)</td>
<td>0.23 (0.11-0.73)</td>
</tr>
<tr>
<td>Fibrosis Test (cutoff)</td>
<td>Number of Samples</td>
<td>Sensitivity (median, range)</td>
<td>Specificity (median, range)</td>
<td>Positive Likelihood Ratio (median, range)</td>
<td>Negative Likelihood Ratio (median, range)</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>------------------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>FibroTest &gt;0.73, &gt;0.75, &gt;0.862</td>
<td>7</td>
<td>0.56 (0.30-1.0)</td>
<td>0.81 (0.24-0.96)</td>
<td>2.9 (1.2-10)</td>
<td>0.54 (0.0-0.79)</td>
</tr>
<tr>
<td>With additional data</td>
<td>10&lt;sup&gt;b,d&lt;/sup&gt;</td>
<td>0.49 (0.11-0.86)</td>
<td>0.89 (0.55-1.0)</td>
<td>4.3 (1.2-11)</td>
<td>0.57 (0.20-0.89)</td>
</tr>
<tr>
<td>Forns Index &gt;4.2</td>
<td></td>
<td>0.98</td>
<td>0.27</td>
<td>1.3</td>
<td>0.07</td>
</tr>
<tr>
<td>With additional data</td>
<td></td>
<td>0.66 (0.27-1.0)</td>
<td>0.31 (0-1.0)</td>
<td>1.4 (0.27-1.5)</td>
<td>0.07 (0-0.66)</td>
</tr>
<tr>
<td>Forns Index &gt;6.9</td>
<td></td>
<td>0.67</td>
<td>0.91</td>
<td>7.4</td>
<td>0.36</td>
</tr>
<tr>
<td>With additional data</td>
<td></td>
<td>0.66 (0.53-0.67)</td>
<td>0.87 (0.86-0.91)</td>
<td>5.2 (4.2-7.4)</td>
<td>0.39 (0.36-0.53)</td>
</tr>
<tr>
<td>Lok Index ≥0.20 or &gt;0.26</td>
<td></td>
<td>0.90 (0.67-1.0)</td>
<td>0.50 (0.30-0.82)</td>
<td>1.8 (1.0-4.8)</td>
<td>0.21 (0-0.94)</td>
</tr>
<tr>
<td>With additional data</td>
<td></td>
<td>0.90 (0.67-1.0)</td>
<td>0.53 (0.30-0.82)</td>
<td>1.9 (1.0-4.8)</td>
<td>0.19 (0-0.94)</td>
</tr>
<tr>
<td>Lok Index ≥0.5 or &gt;0.6</td>
<td></td>
<td>0.53 (0.40-0.79)</td>
<td>0.88 (0.60-0.95)</td>
<td>4.4 (1.3-11)</td>
<td>0.53 (0.24-0.80)</td>
</tr>
<tr>
<td>With additional data</td>
<td></td>
<td>0.53 (0.23-0.79)</td>
<td>0.91 (0.60-0.97)</td>
<td>5.8 (1.3-11)</td>
<td>0.52 (0.24-0.80)</td>
</tr>
</tbody>
</table>

ALT = serum alanine aminotransferase; API = age platelet index; APRI = aspartate aminotransferase to platelet ratio index; AST = aspartate aminotransferase; ELF = enhanced liver fibrosis.

<sup>a</sup>Values in bold indicate a change to above or below a cutoff of 5.0 for positive likelihood ratio or 0.20 for negative likelihood ratio

<sup>b</sup>Additional data for study(s) with discrepancy in reported data

<sup>c</sup>First number is number of samples for sensitivity/second number is number of samples for specificity

<sup>d</sup>Additional data for study(s) without 2 x 2 tables
Discussion

Our experience in attempting to contact authors and acquire additional information in a systematic review of diagnostic tests is encouraging. We were able to contact the majority of authors, particularly for papers published within the last few years. Most contacted authors provided data, and several more would have likely complied with our request had the data been more readily accessible to them. However, this effort was time consuming, not only for us, but also for study authors, who often had to first locate the data before being able to complete the 2 x 2 tables. In addition, despite our efforts, data to resolve discrepancies or calculate 2 x 2 tables at commonly used cutoffs for sensitivity and specificity could not be obtained for 48 of 77 (62 percent) studies, most frequently because authors could not be contacted or they did not have access to the data; this experience indicates that despite relatively extensive efforts to obtain additional data, unresolved discrepancies and missing data remain likely. All data were obtained with the first three out of five attempted contacts, suggesting that more extensive efforts may be of low yield. In particular, telephone contact did not produce any additional information.

The effects of requests for additional data on the results and conclusions regarding the utility of blood tests to identify patients with clinically significant fibrosis or cirrhosis were generally small. An exception was the FIB-4 at a threshold of >3.25 for fibrosis, for which the median positive likelihood ratio increased from 2.4 to 9.3 (moving into the moderately useful range), and the FibroIndex >1.25, for which the median negative likelihood ratio increased from 0.15 to 0.62 (dropping out of the moderately useful range). Although the additional information also affected the categorization for the Fibrostest, the Lok Index, the AST/ALT ratio, and the age-platelet index, the changes in the actual median likelihood ratio estimates were small, resulting in less certainty regarding the significance of the reclassifications. For the Lok Index, additional data also enabled estimation of accuracy for fibrosis. Consistent with its development as a tool to identify patients with cirrhosis, the Lok Index was only weakly predictive for fibrosis.

We did not identify a clear pattern for the directional impact of additional data on estimates of diagnostic accuracy. There were also no clear differences in the effects of additional data from studies with discrepancies versus those in which 2 x 2 tables could not be calculated. However, data to evaluate for such effects were too limited to draw strong conclusions.

Our results support the strength of evidence ratings as assigned in the original report, as additional data had little impact on tests rated as high strength of evidence. For tests originally rated moderate strength of evidence, effects of new data were more variable, with small changes except for the FIB-4 and FibroIndex. No test for which we obtained additional data was originally rated low strength of evidence, where one would expect the impact of additional data to be the greatest, so these results may not be applicable to systematic reviews where the majority of outcomes have low strength of evidence. The impact in such situations could be examined through simulation or modeling analyses in which random, smaller samples of the original data are used as the “base” case. However, such an analysis was outside the scope of this report. Additional data did not change strength of evidence ratings, which is unsurprising given the relatively small numbers of studies for most tests. The one test for which a large number of studies provided new data (APRI) was already rated high strength of evidence and therefore additional data could not increase the rating.

In summary, requests for additional information from authors of primary studies resulted in additional data for a systematic review of diagnostic accuracy of blood tests to identify clinically significant fibrosis or cirrhosis in patients with hepatitis C virus infection. The additional data enabled somewhat more robust estimates for diagnostic accuracy at commonly used cutoffs for a
number of blood tests. Although the effects on summary estimates were relatively small in most cases, the changes had important implications in assessing the clinical utility of two tests (the FIB-4 and the FibroIndex), in one case moving a blood test into the moderately useful range and in the other case moving it out of the moderately useful range. This suggests that while including previously unpublished data can result in clinically important changes in estimates, the magnitude and direction of impact may not be readily predictable. Despite relatively extensive efforts, we were unable to obtain data to resolve discrepancies or complete 2 x 2 tables for 48 of 77 studies. Given that three attempts were needed to obtain even that level of information, more efficient mechanisms of achieving better access to information are needed. Requiring authors of studies on diagnostic accuracy to provide the 2 x 2 tables at commonly used cutoffs in the original study publication (or in the results of publically available trial registries such as ClinicalTrials.gov) would save time, enable systematic reviewers to synthesize data more readily and completely, and enable more transparent verification of authors’ estimates of diagnostic accuracy.
References


44. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA. 1994 Mar 2;271(9):703-7. PMID: 8309035.


Abbreviations and Acronyms

API: age platelet index
APRI: aspartate aminotransferase/platelet ratio
AST: aspartate aminotransferase
AUROC: Area Under the Receiver Operating Characteristic
ELF: Enhanced Liver Fibrosis
METAVIR: Meta-analysis of Histological Data in Viral Hepatitis
Appendix A. Studies Identified for Followup To Obtain Additional Data or Clarifications


73. Viana MSVB. Use of AST platelet ratio index (APRI Score) as an alternative to liver biopsy for treatment indication in chronic hepatitis C. Ann Hepatol. 2009;8(1)PMID: 19221530.


Appendix B. Sample of Letter Sent to Authors

Dear Dr. ____________,

We are researchers at the Pacific Northwest Evidence-based Practice Center at Oregon Health & Science University and are conducting a pooled analysis of diagnostic tests for liver fibrosis/cirrhosis in patients with HCV infection. We included your research in our study (Blood Tests to Diagnose Fibrosis or Cirrhosis in Patients with Chronic Hepatitis C Virus Infection, Annals of Internal Medicine, 2013; 158: 807-820) but would like to perform additional analyses based on a more complete set of data, and need more information about your study, “(insert study title)”. 
We need data in order to complete the 2 X 2 table for APRI, FibroTest, FibroIndex, Forns Index and FIB-4 in patients with hepatitis C virus infection:

<table>
<thead>
<tr>
<th>Fibrosis/Cirrhosis Test</th>
<th>Patient HAS fibrosis/cirrhosis</th>
<th>Patient does NOT have fibrosis/cirrhosis</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>APRI &gt; 0.5 for F2-F4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>APRI &gt; 1.0 for F4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fibrotest cut-off 0.73 for F2-F4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fibrotest cut-off 0.73 for F4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Forns &gt; 4.2 for F2-F4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Forns &gt; 4.2 for F4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Forns &gt; 6.9 for F2-F4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Forns &gt; 6.9 for F4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From your publication, we were able to abstract AUROCs but need your help with the raw numbers for the 2 X 2 tables above. We understand that we are requesting a lot of information and that you may not have time to complete the entire table. Even completing the 2 X 2 table for a few of the tests would be most helpful to us and we would be extremely grateful.

We very much appreciate any assistance you could give us with this matter. We would be happy to answer any questions you may have regarding this research.
Thank you.

Sincerely,

Roger Chou, MD  
Director  
Pacific Northwest Evidence-based Practice Center  
Oregon Health & Science University  
Portland, Oregon  
USA

Shelley Selph, MD  
Research Fellow  
Pacific Northwest Evidence-based Practice Center  
Oregon Health & Science University  
Portland, Oregon  
USA