The diagnosis of rupture of fetal membranes (ROM): a meta-analysis

Abstract

Aim: The aim of this study was to compare the performance of tests based on the detection of insulin-like growth factor binding protein 1 (IGFBP-1) and placental α-microglobulin-1 (PAMG-1) in diagnosing rupture of fetal membranes (ROM) across different patient populations.

Methods: A meta-analysis was conducted on prospective observational or cohort studies investigating ROM tests based on the detection of IGFBP-1 and PAMG-1 meeting the following criteria: (1) performance metrics calculated by comparing results to an adequate reference method; (2) sensitivity thresholds of the investigated tests matching those of the currently available tests; (3) study population, as a minimum, included patients between 25 and 37 weeks of gestation. Sensitivities, specificities, and diagnostic odds ratios were calculated.

Results: Across all patient populations, the analyzed performance measures of the PAMG-1 test were significantly superior compared with those of the IGFBP-1 test. Of particular clinical relevance, PAMG-1 outperformed IGFBP-1 in the equivocal group, which comprised patients with uncertain rupture of membranes (sensitivity, 96.0% vs. 73.9%; specificity, 98.9% vs. 77.8%; PAMG-1 vs. IGFBP-1 tests, respectively).

Conclusions: Compared with its performance in women with known membrane status, the accuracy of the IGFBP-1 test decreases significantly when used on patients whose membrane status is unknown. In this latter clinically relevant population, the PAMG-1 test has higher accuracy than the IGFBP-1 test.

Keywords: Clinical relevance; insulin-like growth factor binding protein (IGFBP-1); placental α-microglobulin 1 (PAMG-1); premature rupture of fetal membranes (PROM); uncertain rupture.

Introduction

Over the years, more than 100 different approaches have been proposed in obstetrical practice for the diagnosis of premature or prelabor rupture of the fetal membranes (PROM) [13, 45]. The mere number of such attempts signals the importance of making an accurate diagnosis of PROM. PROM is encountered in 10% of all pregnancies, with up to 5% of those cases occurring preterm [referred to as preterm PROM (PPROM)]. The latter group accounts for up to 40% of all spontaneous preterm births [3, 20]. It follows that an accurate diagnosis of PROM is essential in guiding the clinical management and allowing for the early and timely administration of antibiotics, corticosteroids, and other interventions to help reduce the effects of prematurity [35].

Despite the high number of proposed methods for the diagnosis of PROM, most have not entered into or persisted in routine clinical practice. Some were just impractical, whereas others performed poorly in unknown cases (membrane status unknown at the time of presentation with suspicion of ROM), despite good performance previously demonstrated in known samples or unequivocal cases.

A good example of a test that is impractical for routine use is the intra-amniotic injection of indigo carmine dye;
consequently, its use remains very limited [4]. Well-known examples of methods with poor performance in unknown cases and whose role was limited to that of a supportive test rather than a confirmatory one are the fern test [10, 48] and the fetal fibronectin (IFN) test [15, 22].

The more recent literature has focused prevalently on two ROM biomarker tests: the AmniSure® ROM Test (AmniSure® International LLC, Boston, MA, USA), based on the detection of placental α-microglobulin 1 (PAMG-1), and the Actim® Prom Test (Oy Medix Biochemica Ab, Kauniainen, Finland), based on the detection of insulin-like growth factor binding protein (IGFBP-1) [12, 25, 27, 39]. The test based on PAMG-1 is the more recent of the two, with the first study on it being published in 2005, vs. 1996 for the test based on IGFBP-1 [8, 41]. The objective of this systematic review was to compare the performance of these two tests in relevant patient populations.

**Methods**

**Data source**

The literature published in any language between 1990 and 2011 was searched for papers on the diagnosis of premature or prelabor rupture of the fetal membranes. We searched the MEDLINE bibliographic database using a combination of keywords, including “rupture of membranes”, “insulin-like growth factor binding protein”, “IGFBP-1”, “placental α-microglobulin 1”, and “PAMG-1”. All references in the retrieved articles were screened for further papers. Editorials, proceedings of meetings, and reviews, although not included in the analysis, were scanned for relevant studies not quoted by the database.

**Study selection**

Only prospective observational or cohort studies that met the following criteria were included in the meta-analysis: (1) the performance metrics were calculated by comparing the results with an adequate reference method for the diagnosis of ROM as defined later; (2) the investigated test(s) had sensitivity thresholds matching those of the currently available tests for the respective antigens: 5 ng/mL in vivo for PAMG-1 [8] and 400 ng/mL in vivo for IGFBP-1 (i.e., 25 ng/mL in vitro) [41]; and (3) the study population included (but was not confined to) patients between 25 and 37 weeks of gestation.

An adequate reference method is one that is expected to be accurate, such as (a) visible leakage from the cervical os or intraamniotic injection of indigo carmine dye or (b) a chart review of the patient’s clinical course from initial diagnosis that includes outcome measures closely linked to the clinical pathology of ROM (e.g., duration of latency period, time to delivery, results of repeat examinations, signs of fetal distress, chorioamnionitis) [45]. On the contrary, an inadequate reference method is one with limited accuracy, such as the diamine oxidase (DAO) or IFN test.

To establish clinical homogeneity among patient populations from the included studies, two main groups were created to compare performance metrics:

1. **Known group**: cases with unequivocally ruptured membranes (e.g., artificially ruptured membranes, gross leakage of amniotic fluid, or known amniotic fluid samples used in the study) or unequivocally not ruptured membranes (e.g., asymptomatic women presenting for routine antenatal screening without complaints of leakage).

2. **Unknown group**: patients presenting with signs and symptoms of ROM with unknown membrane status at the time of study enrollment.

One study may appear in more than one patient population group when more than one set of performance metrics relevant to different patient populations were used in the study.

From each study, the following data were extracted: the total number of patients and the number of true-positive, true-negative, false-positive, and false-negative results for the diagnosis of ROM. The performance measures for the PAMG-1 and IGFBP-1 tests were sensitivity, specificity, and the diagnostic odds ratios. Sensitivity and specificity tests assessed diagnostic accuracy without being influenced by the different prevalence of ROM within the different patient population groups, and the diagnostic odds ratio is one of the better measures of overall accuracy, as it makes the most efficient use of all data points. The performance of the same test in different patient populations was compared as well as the performance of the two tests in the same patient population.

Weighted least squares regressions on the logits of each measure were performed, with the weights inversely proportional to the variance of the logits. Significance was determined at the 0.05 level through t-tests on the coefficients. In cases where a false-negative or false-positive value was 0, 0.5 was added to that value, whereas 0.5 was subtracted from the true-positive or negative-value, depending on the measure being calculated [17].

**Results**

**Study selection**

The search yielded 36 articles, eight of which related specifically to the PAMG-1 test [5, 8, 29–31, 36, 37, 43], 24 to the IGFBP-1 test [1, 6, 9, 11, 14, 16, 18, 19, 21, 23, 24, 26–28, 32–34, 38, 40, 42, 46, 47, 49, 50], and seven that were related to both [2, 7, 13, 25, 39, 44, 45]. The supplementary search of proceedings of perinatal meetings yielded three abstracts, all of which related specifically to the PAMG-1 test [31, 36, 43]. Together, 39 studies were identified and evaluated further for inclusion into the meta-analysis. Figure 1 illustrates the study selection algorithm.

The first filter isolated 21 prospective observational studies that investigated one or both of the biomarkers for their ability to diagnose ROM. The 18 studies that were excluded were review articles, investigated a property of the testing device itself (e.g., reproducibility of testing...
results), or investigated the ability of the test for an alternative indication (e.g., prediction of preterm delivery).

The second filter isolated 14 of the 21 remaining studies based on the inclusion criterion stipulating that the tests detection thresholds should match those of the commercially available tests for the respective antigens. The 7 studies that were eliminated were specifically related to the detection of IGFBP-1 and did not match the detection threshold of the commercially available kit (400 ng/mL in vivo; 25 ng/mL in vitro). All of the 14 selected studies included patients between 25 and 37 weeks of gestation, satisfying all the inclusion criteria.

Finally, 2 more studies were eliminated from further evaluation on account of an inadequate reference method to generate performance metrics. In one study [1], the IGFBP-1 test was compared with the results of DAO, a method that is not considered accurate in diagnosing ROM [45], and in the other [21], the performance metrics for
the IGFBP-1 test were based on a heterogeneous outcome measure (delivery within 2 weeks) that does not allow for a direct association with ROM. It is noted that this study could have been eliminated from further evaluation during the second filter instead of the third because although the IGFBP-1 test used had a detection threshold matching that of the commercially available kit for IGFBP-1, the test was performed by placing the testing strip directly into the cervical os and posterior fornix of the vagina without the use of the collection swab [1]. Given that Rutanen et al. [41] highlighted that the swab is responsible for a 1:16 dilution of the sample, the use of the test without the swab lowers the detection threshold of the test quite substantially.

Grouping performance metrics by patient population group

Table 1 outlines the grouping of the various performance metrics by the patient population group from which they were derived. From the included 12 studies, 16 sets of performance metrics were extracted. Six of these sets were for the PAMG-1 test (unknown group) and 10 were for the IGFBP-1 test (known and unknown group). No sets of performance metrics were identified for the PAMG-1 test in patients that had unequivocally ruptured membranes or unequivocally not ruptured membranes (i.e., the known group). For both the PAMG-1 and IGFBP-1 test, six sets of performance metrics were identified that were derived from patients presenting with suspected ROM but with unknown membrane status at the time of presentation (i.e., the unknown group).

Comparison of performance metrics between tests and within patient population groups

Across all patient population groups (known and unknown), the PAMG-1 test performed significantly better than the IGFBP-1 test with respect to all performance measures ($P<0.01$; Table 2).

For the unknown group specifically, the PAMG-1 test performed significantly better than the IGFBP-1 test, with respect to all performance measures ($P<0.012$; Table 2). Figure 2 illustrates how each test performed in the unknown group using the averages of the measures.

<table>
<thead>
<tr>
<th>GA range</th>
<th>Study</th>
<th>PPG</th>
<th>TP</th>
<th>FN</th>
<th>TN</th>
<th>FP</th>
<th>n</th>
<th>SN (%)</th>
<th>SP (%)</th>
<th>DOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>24–42</td>
<td>Silva et al. [43]$^b$</td>
<td>Unknown</td>
<td>21</td>
<td>0</td>
<td>42</td>
<td>0</td>
<td>63</td>
<td>100</td>
<td>100</td>
<td>3403</td>
</tr>
<tr>
<td>17–42</td>
<td>Birkenmaier et al. [5]$^b$</td>
<td>Unknown</td>
<td>51</td>
<td>3</td>
<td>143</td>
<td>2</td>
<td>199</td>
<td>94.4</td>
<td>98.6</td>
<td>1216</td>
</tr>
<tr>
<td>15–42</td>
<td>Cousins et al. [8]</td>
<td>Unknown</td>
<td>90</td>
<td>1</td>
<td>112</td>
<td>0</td>
<td>203</td>
<td>98.9</td>
<td>100</td>
<td>20,048</td>
</tr>
<tr>
<td>11–42</td>
<td>Lee et al. [30]</td>
<td>Unknown</td>
<td>157</td>
<td>2</td>
<td>21</td>
<td>3</td>
<td>183</td>
<td>98.7</td>
<td>87.5</td>
<td>550</td>
</tr>
<tr>
<td>17–37</td>
<td>Tagore and Kwek [44]</td>
<td>Unknown</td>
<td>38</td>
<td>3</td>
<td>59</td>
<td>0</td>
<td>100</td>
<td>92.7</td>
<td>100</td>
<td>1475</td>
</tr>
<tr>
<td>16–41</td>
<td>Albayrak et al. [2]</td>
<td>Unknown</td>
<td>83</td>
<td>7</td>
<td>77</td>
<td>2</td>
<td>167</td>
<td>94.3</td>
<td>97.5</td>
<td>639</td>
</tr>
<tr>
<td>Average$^c$</td>
<td>Unknown</td>
<td>440</td>
<td>14</td>
<td>454</td>
<td>7</td>
<td>915</td>
<td>96.9</td>
<td>98.5</td>
<td>2038</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GA range</th>
<th>Study</th>
<th>PPG</th>
<th>TP</th>
<th>FN</th>
<th>TN</th>
<th>FP</th>
<th>n</th>
<th>SN (%)</th>
<th>SP (%)</th>
<th>DOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>25–42</td>
<td>Darj and Lyrenäs [9]$^b$</td>
<td>Unknown</td>
<td>46</td>
<td>19</td>
<td>30</td>
<td>4</td>
<td>99</td>
<td>70.8</td>
<td>88.2</td>
<td>18</td>
</tr>
<tr>
<td>22–42</td>
<td>Jeurgens-Borst et al. [24]$^b$</td>
<td>Unknown</td>
<td>22</td>
<td>5</td>
<td>40</td>
<td>16</td>
<td>83</td>
<td>81.5</td>
<td>71.4</td>
<td>11</td>
</tr>
<tr>
<td>24–39</td>
<td>Martinez et al. [34]</td>
<td>Unknown</td>
<td>19</td>
<td>3</td>
<td>20</td>
<td>7</td>
<td>49</td>
<td>86.4</td>
<td>74.1</td>
<td>18</td>
</tr>
<tr>
<td>16–41</td>
<td>Albayrak et al. [2]</td>
<td>Unknown</td>
<td>79</td>
<td>9</td>
<td>77</td>
<td>2</td>
<td>167</td>
<td>89.8</td>
<td>97.5</td>
<td>338</td>
</tr>
<tr>
<td>15–41</td>
<td>Kubota and Takeuchi [26]</td>
<td>Unknown</td>
<td>18</td>
<td>1</td>
<td>27</td>
<td>2</td>
<td>48</td>
<td>94.7</td>
<td>93.1</td>
<td>243</td>
</tr>
<tr>
<td>17–37</td>
<td>Tagore and Kwek [44]</td>
<td>Unknown</td>
<td>35</td>
<td>5</td>
<td>51</td>
<td>3</td>
<td>94</td>
<td>87.5</td>
<td>94.4</td>
<td>119</td>
</tr>
<tr>
<td>Subaverage$^b$</td>
<td>Unknown</td>
<td>219</td>
<td>42</td>
<td>245</td>
<td>34</td>
<td>540</td>
<td>83.9</td>
<td>87.8</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>20–42</td>
<td>Erdemoglu and Mungan [14]</td>
<td>Known</td>
<td>35</td>
<td>1</td>
<td>34</td>
<td>1</td>
<td>71</td>
<td>97.2</td>
<td>97.1</td>
<td>1190</td>
</tr>
<tr>
<td>15–41</td>
<td>Kubota and Takeuchi [26]</td>
<td>Known</td>
<td>40</td>
<td>2</td>
<td>38</td>
<td>4</td>
<td>84</td>
<td>95.2</td>
<td>90.5</td>
<td>190</td>
</tr>
<tr>
<td>24–39</td>
<td>Martinez et al. [34]</td>
<td>Known</td>
<td>20</td>
<td>0</td>
<td>13</td>
<td>1</td>
<td>34</td>
<td>100</td>
<td>92.9</td>
<td>500</td>
</tr>
<tr>
<td>15–37</td>
<td>Rutanen et al. [41]</td>
<td>Known</td>
<td>55</td>
<td>0</td>
<td>71</td>
<td>4</td>
<td>130</td>
<td>100</td>
<td>94.7</td>
<td>1939</td>
</tr>
<tr>
<td>Subaverage$^c$</td>
<td>Known</td>
<td>150</td>
<td>3</td>
<td>156</td>
<td>10</td>
<td>319</td>
<td>98.0</td>
<td>94.0</td>
<td>780</td>
<td></td>
</tr>
<tr>
<td>Average$^c$</td>
<td>Unknown</td>
<td>369</td>
<td>45</td>
<td>401</td>
<td>44</td>
<td>859</td>
<td>89.1</td>
<td>90.1</td>
<td>75</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Performance measures by patient population group.
GA=gestational age, PPG=patient population group, TP=true-positive, FN=false-negative, TN=true-negative, FP=false-positive, n=total number, SN=sensitivity, SP=specificity, DOR=diagnostic odds ratio. *A value of 0.5 was added to an FN or FP of 0, and 0.5 was subtracted from the true-positive or true-negative value depending on the measure being calculated. **Patient population consisted of only those presenting with suspicion of ROM who did not have gross or obvious ruptures, i.e., the equivocal group. Averages were calculated for each diagnostic measure using the pooled TP, FN, TN, and FP numbers from the studies within the specified group.
Ramsauer et al., Diagnosis of ROM meta-analysis

(96.9% vs. 83.9%; specificity, 98.5% vs. 87.8%; PAMG-1 test and IGFBP-1 test, respectively).

The IGFBP-1 test performed significantly better in the known group than in the unknown group with respect to sensitivity ($P=0.008$; Table 3) and the diagnostic odds ratio ($P=0.017$; Table 3). Figure 3 illustrates how the IGFBP-1 test performed in both patient population groups using averages of the measures (sensitivity, 98.0% vs. 83.9%; specificity, 94.0% vs. 87.8%; known and unknown IGFBP-1 groups, respectively). Because no studies were found investigating the performance of the PAMG-1 test in known samples or unequivocal patient cases, it was not possible to compare the performance of this test between the groups.

A subgroup analysis was also performed on the performance of each test in patients presenting with suspected ROM but for whom leakage from the cervical os could not be visualized. We called this subgroup of the unknown group, the “equivocal group”. For the equivocal group, the PAMG-1 test performed significantly better than the IGFBP-1 test with respect to the diagnostic odds ratio ($P=0.019$; Table 4). Figure 4 illustrates how each test performed in the equivocal group using averages of the measures (sensitivity, 96.0% vs. 73.9%; specificity, 98.9% vs. 77.8%; PAMG-1 test and IGFBP-1 test, respectively).

Lastly, the IGFBP-1 test performed significantly better for the known group than it did for the equivocal group with respect to sensitivity ($P=0.042$; Table 4) and the diagnostic odds ratio ($P=0.018$; Table 4). Figure 5 illustrates how the IGFBP-1 test performed in the known and equivocal patient population groups using averages of the measures (sensitivity, 98.0% vs. 73.9%; specificity, 94.0% vs. 77.8%; known and equivocal IGFBP-1 groups, respectively). Because no studies were found investigating the performance of the PAMG-1 test in known samples or obvious patient cases, we were unable to compare the performance of this test for the equivocal group to that for the known group.

### Discussion

Although it was found that all studies investigating the PAMG-1 test were conducted solely on patients with unknown membrane status, many of studies specifically focusing on the IGFBP-1 test included a patient population for whom there existed no question about the status of their membranes (i.e., the known group) [18, 21, 32, 36]. When the known and unknown groups were compared, we found that the sensitivity and diagnostic odds ratio for the IGFBP-1 test were lower in patients with unknown membrane status compared with those whose membrane

<table>
<thead>
<tr>
<th>Description</th>
<th>Performance measure</th>
<th>Interpretation of statistically significant results$^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAMG-1 vs. IGFBP-1: unknown and known</td>
<td>SN 0.009$^<em>$ SP 0.005$^</em>$ DOR 0.001$^*$</td>
<td>The PAMG-1 test performed better than the IGFBP-1 test overall.</td>
</tr>
<tr>
<td>PAMG-1 vs. IGFBP-1: unknown</td>
<td>SN 0.005$^<em>$ SP 0.011$^</em>$ DOR 0.003$^*$</td>
<td>The PAMG-1 test performed better than the IGFBP-1 test in the unknown group.</td>
</tr>
</tbody>
</table>

Table 2 Test comparison within and between patient population groups statistical analysis.

<table>
<thead>
<tr>
<th>Description</th>
<th>Performance measure</th>
<th>Interpretation of statistically significant results$^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGFBP-1: unknown vs. known</td>
<td>SN 0.008$^<em>$ SP 0.504 DOR 0.017$^</em>$</td>
<td>The IGFBP-1 test performed better in the known group than in the unknown group.</td>
</tr>
<tr>
<td>PAMG-1: unknown vs. known</td>
<td>N/A SP N/A DOR N/A</td>
<td>Not applicable because the known group does not exist for the PAMG-1 test.</td>
</tr>
</tbody>
</table>

Table 3 Individual test performance between patient population group statistical analyses.

SN=sensitivity, SP=specificity, DOR=diagnostic odds ratio. $^*$Significance level, $P<0.05$. 

Unauthenticated Download Date | 7/3/17 11:47 PM
status was known. This finding has practical implications because in obstetrical care, the only clinically relevant population to test is that of women for whom the status of the membranes is not obvious at the time of presentation [9, 28].

Similarly, the classical fern test was found by de Haan et al. [10] to perform better in obvious or known cases than in non-laboring patients suspected of ROM but with unknown membrane status. Coupled with the practical difficulties of maintaining microscopes and preparing samples, the poorer performance of the fern test in clinically relevant patient populations led to its eventual disuse in most European countries [5].

For the group of patients that were suspected to have had ROM but whose membrane status was unknown at the time of inclusion into the study (i.e., the unknown group), the PAMG-1 test performed significantly better than the IGFBP-1 test with respect to sensitivity, specificity, and the diagnostics odds ratio (Table 2).

The PAMG-1 test was also compared with the IGFBP-1 test with respect to their performance in the equivocal group (i.e., patients presenting with suspected ROM but for whom leakage from the cervical os could not be visualized). As Figure 4 shows, the PAMG-1 test performed better than the IGFBP-1 test in the equivocal group across sensitivity and specificity measures (sensitivity, 96.0% vs. 73.9%; specificity, 98.9% vs. 77.8%; PAMG-1 and IGFBP-1 tests, respectively) and also in the diagnostic odds ratio (P=0.019). In vitro studies attempting to simulate the clinically relevant patient cases in which membrane rupture is not obvious have demonstrated that the PAMG-1 test will remain positive for several serial dilutions of amniotic fluid beyond the level at which the IGFBP-1 test first reads negative [7, 39]. The disparate in vivo sensitivities of the two tests found in the present study in patients for whom membrane rupture is suspected, but not obvious, agree with the findings of the in vitro simulations of this same patient group.

<table>
<thead>
<tr>
<th>Description</th>
<th>Performance measure</th>
<th>Interpretation of statistically significant results*</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAMG-1 vs. IGFBP-1: equivocal</td>
<td>0.073 0.071 0.019*</td>
<td>The PAMG-1 test performed better than the IGFBP-1 test in the equivocal group.</td>
</tr>
<tr>
<td>IGFBP-1: equivocal vs. known</td>
<td>0.042* 0.214 0.018*</td>
<td>The IGFBP-1 test performed better in the known group than in equivocal group.</td>
</tr>
</tbody>
</table>

Table 4 Equivocal subgroup statistical analysis.  
SN=sensitivity, SP=specificity, DOR=diagnostic odds ratio. *Significance level, P<0.05.
Conclusion

Compared with its performance in women for whom membrane status is known, the performance of the IGFBP-1 test decreases significantly when used on patients for whom membrane status is unknown. In this latter clinically relevant population, the PAMG-1 test has higher accuracy than the IGFBP-1 test.

Received October 19, 2012. Accepted November 16, 2012. Previously published online December 25, 2012.

References


The authors stated that there are no conflicts of interest regarding the publication of this article.