Three Highly Sensitive “Bedside” Serum and Urine Tests for Pregnancy Compared

Henrik Christensen,1 Hans H. Thysen, Ove Schebye, and Arne Berget

We examined three enzyme-linked immunosorbent assay (ELISA) kits for human chorionic gonadotropin (hCG) (pregnancy tests) for use with urine and serum samples: the Tandem Icon II hCG Urine and Tandem Icon II hCG Serum, the NovoClone Target hCG Test, and the Abbott TestPack hCG-urine and hCG-serum. Paired comparison of the results from each kit indicated that the NovoClone Target assay showed significantly lower diagnostic sensitivity (P <0.05) than did the Tandem Icon II or Abbott TestPack, both for urine and for serum samples. None of the products demonstrated any significant difference (P >0.05) in diagnostic specificity, but the NovoClone Target kit showed several serious false-negative results with both urine and serum. Paired testing of urine kits vs serum kits also showed no significant differences (P >0.05) in diagnostic sensitivity or specificity. We found the Abbott kits to be the most convenient to use and to read.

Reliable and fast detection of human chorionic gonadotropin (hCG) at low concentrations in urine or serum is important in diagnosing early pregnancy, ectopic pregnancy, and missed abortion. In recent years new sensitive "bedside" pregnancy tests, based on enzyme-linked immunosorbent assays, have been developed for qualitative detection of hCG in serum as well as in urine. We evaluated highly sensitive pregnancy tests, three for serum and three for urine, checking for the occurrence of false-positive and false-negative results, assessing their diagnostic sensitivity and specificity and convenience of use. Finally, we compared the diagnostic sensitivity and specificity of the urine kits with those of the serum kits by applying paired urine and serum samples from each patient, a study that, to our knowledge, has not been published previously.

Materials and Methods

We examined the following tests: Tandem Icon II hCG Serum and Tandem Icon II hCG Urine (Hybritech Inc., San Diego, CA 92121), NovoClone Target hCG Test (Novo Industri A/S, DK-2880 Bagsvaerd, Denmark), and Abbott TestPack hCG-serum and Abbott TestPack hCG-urine (Abbott Laboratories, Chicago, IL 60664).

In these assays, two different antibodies react with two different regions of the hCG molecule, so that the hCG molecules are fixed between the membrane-bound and the enzyme-bound antibodies. The enzyme produces a blue color, the intensity of which is proportional to the hCG concentration of the sample. These kits all contain a reference test zone, so that analysts can detect procedural errors or inactive reagents. The reference test zones of the Tandem Icon are calibrated to hCG concentrations of 25 int. units/L for serum and 50 int. units/L for urine. Thus positive results above and below 25 int. units/L for serum (50 int. units/L for urine) can be distinguished. Furthermore, The Tandem Icon II hCG serum kit contains a negative reference zone test, which in some cases may reveal the presence of immunologically interfering substances.

The NovoClone and Tandem Icon II kits may be used for serum samples as well as for urine samples, whereas the Abbott TestPack requires separate kits for analysis of serum and urine.

The manufacturers’ declared detection limits are shown for each kit examined (Table 1).

We tested 99 serum samples and 99 corresponding urine samples, collected from 27 women, who had had legally induced abortions during their first trimester. Samples of blood and urine were collected from one to three days before the induced abortion and once a week subsequently until all kits showed negative results for hCG or until the comparison method (see below) showed <10 int. units/L of serum. Twelve women failed to appear for some of the later checks; but owing to our study design, we still collected a large number of samples with low hCG concentrations. On all occasions the samples of blood and corresponding urine were collected on the same day with no special instructions for the urine collection. All samples were examined within a few hours after collection. Urine specimens were not centrifuged, but the assay samples were collected above sediments, if necessary. All kits were stored as prescribed by the manufacturer and allowed to reach room temperature before use.

A quantitative analysis for hCG in serum was performed by Statens Seruminstitut, Copenhagen, Denmark, as the comparison method. They measured hCG in serum, according to the manufacturer’s instructions, by time-resolved fluoroimmunoassay (Delfia; Pharmacia, Hillerod, Denmark). This two-site “sandwich” technique involves two monoclonal antibodies, one immobilized to microtiter wells, the other labeled with europium, the other labeled with europium. The sensitivity is 10 int. units/L (WHO 1st International Reference Preparation 1975 [IRP 75/537]) and the interassay precision (CV) is <10% within the range of the standard curve. The comparison method was applied to all serum samples, but not to the urine samples. Previous studies have shown a close statistical correlation between serum and urine in early pregnancy for intact hCG with a ratio of approximately 1:1 if corrected for the urinary creatinine concentration (J). We used the corresponding values of hCG in serum as controls for the urine samples. We calibrated the kits tested and the

<table>
<thead>
<tr>
<th>Table 1. Manufacturers’ Detection Limits for hCG Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kit</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Tandem Icon II</td>
</tr>
<tr>
<td>NovoClone Target</td>
</tr>
<tr>
<td>Abbott TestPack</td>
</tr>
</tbody>
</table>

* As measured vs WHO 1st IRP 75/537.
Results

In the 99 serum samples, the hCG concentration varied from 10 to 260 000 int. units/L; 65% of the samples had <1000 int. units/L. Table 2 lists the distribution of true and false readings, diagnostic sensitivity and specificity, and accuracy of the kits for all samples.

We performed a paired comparison of the individual kits to assess diagnostic sensitivity and specificity according to the McNemar's Test, making paired comparisons of the number of positive and negative readings, respectively, of the kits at different concentrations (Table 3).

The NovoClone kit showed significantly lower diagnostic sensitivity (P < 0.05) than did the Tandem Icon II and the Abbott TestPack for both urine and serum at hCG concentrations of 10–100 int. units/L and for serum at hCG concentrations of 26–100 int. units/L. There was no significant difference between the Tandem Icon II and the Abbott TestPack kits. The difference in the manufacturers' detection limits does not justify the lower diagnostic sensitivity of the NovoClone test with serum. Thus we conclude that the detection limit for the NovoClone test with serum is higher and for the Abbott TestPack hCG—serum is lower than declared by the manufacturers. Comparing the diagnostic specificity of the three products showed no significant difference (P > 0.05) for urine or for serum determinations. False results in assaying urine may be caused by small deviations from the detection limit (10 int. units/L), or by variations in urine dilutions because the controls for the urine results are actually hCG in patients' serum.

The NovoClone Target hCG Test with urine showed negative results for two samples for which the corresponding hCG concentrations in serum were 72 000 and 65 000 int. units/L. The other kits showed positive results for these samples with urine as well as with serum. With the control ring of the NovoClone test clearly visible, no procedural errors could be suspected. The two urine samples were further tested at Novo Biolabs (Novo Industri A/S, Bagsvaerd, Denmark): with one sample (serum hCG 72 000 int. units/L) the false-negative result could not be reproduced, and no sediment was found; sediment found in the other sample (serum hCG 65 000 int. units/L) caused a false-negative result when sucked into the test pipette.

We also found the most serious false-negative results for serum with the NovoClone Target hCG Test (hCG concentrations of 110, 150, and 710 int. units/L). Two serum samples also showed false-positive results with the Tandem Icon II and the Abbott TestPack, probably caused by detecting hCG at concentrations <10 int. units/L rather than by interference from some chemical substance.

Three of the 27 women were under medical treatment at the time of sample collection: one using terbutaline spray, another penicillin, and the third oxazepam. The number of false results was not increased by samples from these three women.

We made comparisons of urine kits vs serum kits for diagnostic sensitivity and specificity by McNemar's test, using the data in Table 3 as the starting points. No significant difference (P > 0.05) in diagnostic sensitivity or specificity was found between Tandem Icon II hCG urine vs Tandem Icon II hCG serum, between NovoClone Target hCG Test with urine vs NovoClone Target hCG Test with serum, and between Abbott TestPack hCG—urine vs Abbott TestPack hCG—serum.

With the Tandem Icon II, results for all 99 serum samples were compared with results for the positive reference zone, to differentiate between positive results that were greater than, equal to, or less than 25 int. units/L (Table 4). All false readings in Table 4 represent small deviations compared with the cutoff value of 25 int. units/L. For clinical use the reference zone will thus give a good estimate as to whether the hCG concentration of a sample is relatively high or low.

Finally, subjectively evaluating the general design of the kits, we found the Abbott kits to be the most convenient to use and to read.

Discussion

In studying urine pregnancy test kits, we found the most serious false-negative results with the NovoClone Target hCG Test (serum hCG concentrations of 72 000 and 65 000 int. units/L). Although precipitate in the urine might be the cause, similar false-negative results from this kit for urine samples have been reported previously (personal communication from Angelholm Hospital, Sweden, and Blaabjerg et al., Odense University Hospital, Denmark, 1987, unpublished). We found no previous studies of the Tandem Icon II kit or of the three serum kits. Most studies recommend an earlier and less sensitive version of the Tandem Icon hCG—urine kit as a reliable pregnancy test.

### Table 2. Clinical Performance of hCG Pregnancy Test Kits with 99 Urine and 99 Serum Samples

<table>
<thead>
<tr>
<th>Urine</th>
<th>Tandem Icon II</th>
<th>NovoClone Target</th>
<th>Abbott TestPack</th>
<th>Tandem Icon II</th>
<th>NovoClone Target</th>
<th>Abbott TestPack</th>
</tr>
</thead>
<tbody>
<tr>
<td>False-positive</td>
<td>0</td>
<td>1 *</td>
<td>0</td>
<td>2*</td>
<td>0</td>
<td>3*</td>
</tr>
<tr>
<td>False-negative</td>
<td>8*</td>
<td>25 *</td>
<td>13</td>
<td>19</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>True-positive</td>
<td>77</td>
<td>60</td>
<td>72</td>
<td>76</td>
<td>57</td>
<td>80</td>
</tr>
<tr>
<td>True-negative</td>
<td>14</td>
<td>13</td>
<td>14</td>
<td>12</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>Diagnostic sensitivity, %</td>
<td>90.6</td>
<td>70.6</td>
<td>84.7</td>
<td>89.4</td>
<td>67.1</td>
<td>94.1</td>
</tr>
<tr>
<td>Diagnostic specificity, %</td>
<td>100.0</td>
<td>92.9</td>
<td>100.0</td>
<td>85.7</td>
<td>100.0</td>
<td>78.6</td>
</tr>
<tr>
<td>Accuracy, %</td>
<td>91.9</td>
<td>73.7</td>
<td>86.9</td>
<td>88.9</td>
<td>71.7</td>
<td>91.9</td>
</tr>
</tbody>
</table>

* Results for control serum hCG, int. units/L, were as follows: * <10; * <10, <10; * <10, <10, <10; * 22, 17, 17, 13, 11, 12, 11; * 200, 64, 30, 15, 50, 22, 65, 22, 32, 10, 65, 17, 17, 13, 35, 16, 72 000, 11, 23, 11, 12, 15, 19, 50, 65 000; * 15, 22, 22, 10, 65, 17, 17, 13, 23, 11, 12, 15, 27; * 15, 17, 17, 13, 16, 11, 11, 12, 15; * 64, 30, 15, 50, 22, 65, 22, 32, 10, 110, 22, 65, 17, 710, 17, 150, 35, 13, 35, 16, 11, 23, 11, 12, 15, 19, 50, 27; * 17, 13, 11, 12, 15. 1 Diagnostic sensitivity = true positive × 100/(true positive + false negative). 2 Diagnostic specificity = true negative × 100/(true negative + false positive). 3 Accuracy = (true positive + true negative) × 100/total number of samples.
The Abbott TestPack hCG–urine and Tandem Icon kits are both recommended as reliable by Bandi et al. (3). Other works also describe an earlier and less-sensitive version of the Tandem Icon serum–hCG kits (7–9).

In conclusion, we recommend the Tandem Icon II and Abbott TestPack for pregnancy tests with serum as well as with urine. Although both showed false-negative results in our study, these represent small deviations from the declared detection limit. We cannot recommend the Novo-Clone Target hCG Test, with urine or with serum, because of its significantly lower diagnostic sensitivity than the Tandem Icon II and Abbott TestPack and because of serious false-negative results. The relatively high frequency of false-positive and false-negative results we found may be related to our testing of many samples with low hCG concentrations. The diagnostic sensitivity and specificity of the kits were not found to be significantly higher with serum specimens than with urine. This is interesting because as serum kits are often more expensive than urine kits, and serum samples are often more difficult to obtain and to handle than urine specimens.

Pregnancy test kits were generously donated by the manufacturers. We thank biostatistician S. Olesen Larsen, Statens Seruminstitut, Copenhagen, for his statistical assistance.

References