Accuracy of soluble human leukocyte antigen-G for predicting pregnancy among women undergoing infertility treatment: meta-analysis

M.J. Vercammen1, A. Verloes1, H. Van de Velde2 and P. Haentjens3,4,5

1Laboratory of Haematology, Universitair Ziekenhuis Brussel (UZ Brussel), Laarbeeklaan 101, B-1090, Brussels, Belgium; 2Centre for Reproductive Medicine, UZ Brussel, Belgium; 3Centre for Outcomes Research, UZ Brussel, Laboratory for Experimental Surgery, Vrije Universiteit Brussel, Laarbeeklaan 101, B-1090 Brussels, Belgium; 4Centre for Evidence-Based Medicine, Belgian Branch of the Cochrane Collaboration (CEBAM), Belgium

5Correspondence address. Tel: +32-2-477-64-18; Fax: +32-2-477-64-28; E-mail: patrick.haentjens@uzbrussel.be

BACKGROUND: There have been concerns about validity and accuracy of the measurement of sHLA-G in embryo culture supernatants. In this systematic review, we quantified the diagnostic accuracy of sHLA-G for predicting the ability to achieve clinical pregnancy in women who are undergoing infertility treatment. METHODS: Medline and Embase were searched up to 7 September 2007, for full English and non-English articles concerning cohort studies evaluating sHLA-G in embryo culture for predicting clinical pregnancy in women undergoing IVF and ICSI. RESULTS: Eleven studies including 1813 patients met our inclusion criteria. In the individual studies, sensitivity ranged from 0.01 to 0.97, specificity from 0.18 to 0.98, the positive likelihood ratio from 0.34 to 3.21 and the negative likelihood ratio from 0.08 to 1.01. These values were highly heterogeneous with, in each case, I² values of >75%, and P-values for the Q statistic of <0.001, arguing against generating a pooled estimate for these diagnostic test properties. The diagnostic odds ratios (DORs) ranged from 0.92 to 24.82 in the individual studies with an I² value of 49% indicating moderate heterogeneity. Therefore, the meta-analysis combined the logs of the DORs, which are derived from sensitivity and specificity. A random-effects model yielded a summary DOR of 4.38 (95% CI, 2.93–6.55), consistent with modest diagnostic accuracy. Interestingly, an a priori defined subgroup analysis restricted to six studies with good quality embryos showed a better diagnostic performance with a DOR of 12.67 (95% CI, 3.66–43.80) to predict the ability to achieve clinical pregnancy in women undergoing infertility treatment. CONCLUSIONS: Further research is needed with single-embryo culture, single-embryo transfer and highly sensitive detection techniques to determine the potential application of measuring sHLA-G in culture supernatant.

Keywords: soluble human leukocyte antigen-G; clinical pregnancy; meta-analysis; diagnostic accuracy; infertility treatment

Introduction

Human leukocyte antigen (HLA)-G is thought to play a role in implantation. This non-classical HLA class Ib molecule has been demonstrated in human preimplantation embryos at the mRNA and protein level as well as in their culture supernatant (Jurisicova et al., 1996a,b; Yao et al., 2005), where the presence of soluble HLA-G (sHLA-G) has been reported a prerequisite for implantation (Fuzzi et al., 2002).

At the maternal–fetal interface, fetal-derived trophoblast cells interact with maternal uterine tissue. In humans, a particular type of uterine natural killer (uNK) cells seems to play a major role in vascular remodelling during decidualization, and in maternal immune tolerance towards the invading trophoblast (Hanna et al., 2006). However, uNK cells secrete growth factors, chemokines and cytokines in response to a balanced stimulation of their killer-cell immunoglobulin-like receptors, leukocyte immunoglobulin-like-receptors and natural cytotoxicity receptors. Ligands for these receptors are HLA antigens expressed on decidual stromal cells and trophoblast, one of these ligands being HLA-G (Hanna et al., 2006; Rajagopalan et al., 2006). HLA-G is also an immunomodulatory molecule with mainly suppressive effects on NK cells, CD4+ and CD8+ T cells, B lymphocytes and antigen presenting cells such as macrophages and dendritic cells (reviewed in Hunt et al., 2007).

HLA-G expression was initially thought to be restricted to the reproductive system, but was later shown to be inducible in other tissues in the presence of viral infections, chronic inflammation, tumours and allotransplantation (reviewed in LeMaoult et al., 2003). HLA-G single mRNA is alternatively spliced in seven distinct mRNA’s leading to soluble (HLA-G5 to HLA-G7) and membrane-bound (HLA-G1 to HLA-G4) isoforms (Ishitani and
Geraghty, 1992; Fuji et al., 1994; Kirszenbaum et al., 1994; Moreau et al., 1995; Paul et al., 2000). sHLA-G in culture supernatant of embryos originates from soluble and shed membrane-bound isoforms. Membrane-bound HLA-G has been shown to induce uNK cell proliferation and interferon-γ secretion (Van der Meer et al., 2004). Endocytosis of sHLA-G by the KIR2DL4 NK cell receptor has been shown to induce a unique profile of proinflammatory/proangiogenic mediators and cytokines favouring implantation and placentation (Rajagopalan et al., 2006).

sHLA-G has been proposed as a non-invasive marker for embryo selection that could be used, in addition to the morphological scoring system, to improve the implantation rate of assisted reproduction. The availability of such a marker is especially important considering the evolution towards single-embryo transfer in order to prevent complications of multiple births. Although a number of reports have demonstrated that sHLA-G can be detected in the culture medium of human IVF embryos and that levels correlate with the potential of an embryo to implant, there have been concerns about the validity and accuracy of the measurement of sHLA-G in embryo supernatants for predicting the ability to achieve clinical pregnancy in women who are undergoing infertility treatment (Sargent et al., 2007).

In the present systematic reviews and meta-analyses, we quantify the diagnostic accuracy of sHLA-G for predicting the ability to achieve clinical pregnancy in women who are undergoing infertility treatment.

Methods

Search strategy

We searched for English and non-English articles using Embase and Medline (OVID and PubMed), with the last computerized search undertaken on 7 September 2007. To avoid missing any relevant study, we used broadly defined medical subject heading terms and text words, including the following: ‘HLA-G’ or ‘soluble HLA-G’ and ‘implantation’ or ‘embryo’. The computerized search was supplemented by a manual search of the bibliographies of all retrieved articles. Potentially relevant articles were assessed for inclusion against pre-specified eligibility and exclusion criteria (Stroup et al., 2000).

Study eligibility

We included longitudinal (cohort) studies published in full that evaluated sHLA-G in embryo supernatants for determining the likelihood of clinical pregnancy in patients undergoing ICSI or IVF. Studies were required to evaluate sHLA-G for predicting treatment outcome in an infertility clinic population, and to report clinical pregnancy, which was defined as observation of at least a gestational sac on ultrasonography. If a particular patient population was reported in more than one publication, we selected the article that provided the most complete dataset. Reviews, letters to the editor, case–control studies, uncontrolled studies, studies not reporting clinical pregnancy as a separate outcome measure and studies providing insufficient data to permit completion of a 2 × 2 contingency table were excluded.

Data extraction

The following data were abstracted: the first author’s name; the publication year; the country of origin; the study design details (prospective versus retrospective); the number of women; women’s age; infertility treatment (ICSI versus IVF); clinical pregnancy rate; embryo transfer characteristics (days after ovum pick-up; single or multiple); culture conditions (single or group; volume; duration); variations in assay protocols (detection technique; supernatant volume; ratio of test/culture volume; antibodies capture detection; standard type); method used to dichotomize sHLA-G test results (detection limit versus optical density) and whether the authors provided separate data according to women’s age, infertility treatment or embryo quality. Data were independently extracted by two of us (P.H. and M.V.) and checked for accuracy in a second review. Consensus was achieved for all data.

Statistical analyses

Outline of principles

Our first aim was to provide, for each eligible study, a summary of the important methodological characteristics, and to show the key diagnostic test properties, i.e. sensitivity, specificity, positive likelihood ratio, negative likelihood ratio and diagnostic odds ratio (DOR). Next, we estimated heterogeneity, determined which diagnostic test properties could be combined into an average summary measure of accuracy, and examined whether diagnostic accuracy depended on methodological characteristics of the studies. Finally, we conducted separate meta-analyses within predetermined subgroups to assess whether diagnostic accuracy differed according to embryo quality or among specific patient groups.

Computational details

Data from all eligible studies were abstracted into 2 × 2 tables expressing the presence or absence of sHLA-G in at least one culture supernatant (single or group culture) of the transferred embryo (single or multiple) versus the presence or absence of clinical pregnancy. The threshold for sHLA-G expression was the one determined by the authors of the individual studies based on the technical performance of their sHLA-G assay. The following diagnostic test properties together with 95% confidence intervals (CIs) were calculated for each study using standard methods: sensitivity, specificity, positive likelihood ratio, negative likelihood ratio and DOR (Walter, 2002; Glas et al., 2003). If a 2 × 2 table contained a cell value of 0 (zero), that study was not omitted, but 0.5 was added to each cell value of the 2 × 2 table of that study (Sutton et al., 2000; Deeks et al., 2001). For each eligible study, sensitivity and specificity were plotted in a receiver operating characteristics (ROC) curve of true positive values on false-positive values (Sutton et al. 2000). Heterogeneity of reported sensitivity, specificity and other diagnostic test properties was assessed by visually examining the ROC curve and by the Cochran’s Q test for heterogeneity (Deeks et al., 2001). Between-study heterogeneity was judged to be excessive when the P value for the Q statistic was <0.10 (Deeks et al., 2001). Heterogeneity was quantified by the I² statistic, the proportion of variability across studies that is due to heterogeneity rather than chance, with I² values <25% indicating low, 25–50% indicating moderate and >50% indicating high heterogeneity. An I² value of 75% or more indicates very high heterogeneity, and suggests that the studies are too different to combine and generate a pooled estimate (Higgins and Thompson, 2002; Higgins et al., 2003).

With low or moderate heterogeneity (I² values of 50% or less), weighted average summary estimates of the diagnostic test properties and the corresponding 95% CI were calculated with weights equal to the inverse variance. Since some heterogeneity was anticipated, only random effects models are presented in the results (DerSimonian and Laird, 1986; Deeks et al., 2001). As a general rule, the
The DOR is the ratio of the odds of a positive test result in a pregnant woman compared with a non-pregnant woman, and it can be estimated from sensitivity and specificity. The DOR is a measure of overall accuracy and has the advantage of allowing assessment of the effects of study methodology characteristics (covariates) on diagnostic accuracy by means of stepwise multiple regression analysis, with the natural log of the DOR as the dependent variable (Moses et al., 1993; Glas et al., 2003). In this regard, we identified, a priori, the following continuous covariates for inclusion in the stepwise multiple regression analysis: the publication year; the number of women; clinical pregnancy rate; number of days after ovum pick-up; culture volume, culture duration, supernatant volume and ratio of test/culture volume. Likewise, the following categorical covariates were included in the regression analysis: study design details (prospective versus retrospective); embryo transfer (single versus multiple); culture conditions (single versus group) and method used to dichotomize sHLA-G test results (detection limit versus optical density).

Finally, we postulated, also a priori, that the findings would be affected by predetermined subgroup characteristics, and thus constructed separate 2 × 2 tables for studies providing separate data pertaining to, in turn, ICSI infertility treatment, IVF infertility treatment, women of <38 years old, women of 38 years old or more and embryos of good quality (defined according to the number of blastomeres and the fragmentation). Accordingly, we conducted five separate random-effects subgroup meta-analyses.

Since DORs have no direct clinical relevance, the positive predictive value (PPV) and negative predictive value (NPV) of sHLA-G in embryo supernatants were calculated and plotted as a function of varying prevalence (pre-test probability) of clinical pregnancy. The following formulae based on Bayes’ theorem were used to perform the calculations:

Positive predictive value = (prevvalence)(sensitivity) / (prevvalence)(sensitivity) + (1 − specificity)(1 − prevvalence).

Negative predictive value = (1 − prevvalence)(specificity) / (1 − prevvalence)(specificity) + (1 − sensitivity)(prevvalence).

Potential publication bias was explored by funnel plot (DerSimonian and Laird, 1986; Sutton et al., 2000), the Begg and Mazumdar test and the Egger test (Begg and Mazumdar, 2004; Egger et al., 1997).

Results

Study characteristics

Our initial search identified 316 unique publications, of which 11 studies from six research groups met all our inclusion criteria (Fig. 1) (Fuzzi et al., 2002; Sher et al., 2004; Criscuoli et al., 2005; Noci et al., 2005; Sher et al., 2005a,b; Yie et al., 2005; Desai et al.; 2006; Fisch et al., 2007; Rebmann et al., 2007; Sageshima et al., 2007). We excluded 42 abstracts; 13 editorial comments, opinion statements or newsletters; 72 narrative reviews; 3 workshop reports; 172 original studies, published in full, but not relevant to our research question and 3 studies published in full that did not consider pregnancy as an outcome measure. Thus, 11 full papers, all published in English, were included in the current review (Table I). Considered together, the cohorts included 1813 women undergoing ICSI or IVF. The overall clinical pregnancy rate ranged from 18% to 70% with a mean of 45%. All but one study used the ELISA sandwich technique. Test threshold was based on either a detection limit or optical density (Table I).

Quantitative data synthesis

According to the data in the individual studies, sensitivity ranged from 0.01 to 0.97, specificity from 0.18 to 0.98, the positive likelihood ratio from 0.34 to 3.21 and the negative likelihood ratio from 0.08 to 1.01 (Table II). Sensitivity, specificity, positive and negative likelihood ratio values were highly heterogeneous. I2 values in each case were >75%, and P values for the Q statistic in each case were <0.001, suggesting that the individual studies were too different to combine and generate a pooled estimate for these diagnostic test properties. The ROC plot also showed wide scatter between the studies, but some of the points approached the left upper corner where there would be good discrimination between the likelihood of success or failure with respect to clinical pregnancy (Fig. 2).
Table I. Characteristics of the cohort studies included in the meta-analysis.

<table>
<thead>
<tr>
<th>First author, year published</th>
<th>Country</th>
<th>Type of study*</th>
<th>Number of women</th>
<th>Infertility treatment</th>
<th>Clinical pregnancy rate (%)</th>
<th>Embryo transfer</th>
<th>Culture</th>
<th>Volume (µl)</th>
<th>Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fuzzi, 2002</td>
<td>Italy</td>
<td>Prospective</td>
<td>101</td>
<td>IVF or ICSI</td>
<td>18%</td>
<td>2 or 3</td>
<td>Multiple</td>
<td>Group</td>
<td>600</td>
</tr>
<tr>
<td>Sher, 2004</td>
<td>USA</td>
<td>Retrospective</td>
<td>201</td>
<td>ICSI</td>
<td>51%</td>
<td>3</td>
<td>Multiple</td>
<td>Single</td>
<td>50</td>
</tr>
<tr>
<td>Criscuoli, 2005</td>
<td>Italy</td>
<td>Prospective</td>
<td>31</td>
<td>IVF or ICSI</td>
<td>29%</td>
<td>3</td>
<td>. .</td>
<td>Single</td>
<td>600</td>
</tr>
<tr>
<td>Noci, 2005</td>
<td>Italy</td>
<td>Prospective</td>
<td>66</td>
<td>IVF or ICSI</td>
<td>14%</td>
<td>3</td>
<td>Multiple</td>
<td>Single</td>
<td>600</td>
</tr>
<tr>
<td>Sher, 2005a</td>
<td>USA</td>
<td>Prospective</td>
<td>107</td>
<td>ICSI</td>
<td>48%</td>
<td>3</td>
<td>Multiple</td>
<td>Single</td>
<td>50</td>
</tr>
<tr>
<td>Sher, 2005b</td>
<td>USA</td>
<td>Prospective</td>
<td>482</td>
<td>ICSI</td>
<td>49%</td>
<td>3</td>
<td>Multiple</td>
<td>Single</td>
<td>50</td>
</tr>
<tr>
<td>Yie, 2005</td>
<td>Canada</td>
<td>Prospective</td>
<td>137</td>
<td>ICSI</td>
<td>44%</td>
<td>3</td>
<td>Multiple</td>
<td>Group</td>
<td>400</td>
</tr>
<tr>
<td>Desai, 2006</td>
<td>USA</td>
<td>Prospective</td>
<td>83</td>
<td>ICSI</td>
<td>55%</td>
<td>3</td>
<td>Multiple</td>
<td>Single</td>
<td>20</td>
</tr>
<tr>
<td>Fisch, 2007</td>
<td>USA</td>
<td>Prospective</td>
<td>209</td>
<td>ICSI</td>
<td>64%</td>
<td>3</td>
<td>Multiple</td>
<td>Single</td>
<td>20</td>
</tr>
<tr>
<td>Rehmann, 2007</td>
<td>Germany</td>
<td>Prospective</td>
<td>313</td>
<td>IVF or ICSI</td>
<td>25%</td>
<td>2, 3, or 4</td>
<td>Multiple</td>
<td>Single or Group</td>
<td>50</td>
</tr>
<tr>
<td>Sageshima, 2007</td>
<td>Japan</td>
<td>Prospective</td>
<td>109</td>
<td>IVF or ICSI</td>
<td>75%</td>
<td>2, 3, 4 or 5</td>
<td>Both**</td>
<td>Single or Group</td>
<td>50 or 1000</td>
</tr>
<tr>
<td>First author, year published</td>
<td>Assay Characteristics</td>
<td>Method for dichotomizing sHLA-G results***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------------</td>
<td>------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Detection technique</td>
<td>Supernatant volume (µl)</td>
<td>Test volume / culture volume</td>
<td>Antibodies capture detection</td>
<td>Standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuzzi, 2002</td>
<td>ELISA sandwich</td>
<td>100</td>
<td>17</td>
<td>MEM-G9, w6/32</td>
<td>HLA-G transfected LCL 721.221</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sher, 2004</td>
<td>ELISA sandwich</td>
<td>50</td>
<td>100</td>
<td>MEM-G1, w6/32</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criscuoli, 2005</td>
<td>ELISA sandwich</td>
<td>100</td>
<td>17</td>
<td>MEM-G9, w6/32</td>
<td>HLA-G transfected LCL 721.221</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noci, 2005</td>
<td>ELISA sandwich</td>
<td>50</td>
<td>8</td>
<td>MEM-G9, w6/32</td>
<td>HLA-G transfected LCL 721.221</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sher, 2005a</td>
<td>ELISA sandwich</td>
<td>50</td>
<td>100</td>
<td>MEM-G9, w6/32</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sher, 2005b</td>
<td>ELISA sandwich</td>
<td>50</td>
<td>100</td>
<td>MEM-G1, w6/32</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yie, 2005</td>
<td>ELISA sandwich</td>
<td>100</td>
<td>25</td>
<td>4H84, 3C/G4</td>
<td>Placenta</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desai, 2006</td>
<td>ELISA sandwich</td>
<td>20</td>
<td>100</td>
<td>MEM-G9, anti-beta-2 m</td>
<td>. . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fisch, 2007</td>
<td>ELISA sandwich</td>
<td>50</td>
<td>250</td>
<td>MEM-G9, w6/32</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rebmann, 2007</td>
<td>Luminex</td>
<td>10</td>
<td>20</td>
<td>MEM-G9, anti-beta-2 m</td>
<td>HLA-G5 and beta2 m transfected SF9 cells</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sageshima, 2007</td>
<td>ELISA sandwich</td>
<td>50</td>
<td>5 or 100</td>
<td>MEM-G9, w6/32</td>
<td>HLA-G transfected LCL 721.221</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Data as provided by the authors in the original paper; . . . no data.
**Single or multiple.
***DL, detection limit; OD, optical density.
<table>
<thead>
<tr>
<th>First author, year published</th>
<th>Sensitivity (95% confidence interval)</th>
<th>Specificity (95% confidence interval)</th>
<th>Positive likelihood ratio (95% confidence interval)</th>
<th>Negative likelihood ratio (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fuzzi, 2002</td>
<td>0.97 (0.78 to 1.00)</td>
<td>0.32 (0.22 to 0.43)</td>
<td>1.42 (1.21 to 1.67)</td>
<td>0.08 (0.01 to 1.31)</td>
</tr>
<tr>
<td>Sher, 2004</td>
<td>0.85 (0.77 to 0.92)</td>
<td>0.57 (0.46 to 0.67)</td>
<td>1.96 (1.55 to 2.49)</td>
<td>0.26 (0.16 to 0.43)</td>
</tr>
<tr>
<td>Criscuoli, 2005</td>
<td>0.93 (0.50 to 1.00)</td>
<td>0.66 (0.38 to 0.87)</td>
<td>2.70 (1.33 to 5.48)</td>
<td>0.11 (0.01 to 1.61)</td>
</tr>
<tr>
<td>Noci, 2005</td>
<td>0.95 (0.62 to 1.00)</td>
<td>0.46 (0.33 to 0.59)</td>
<td>1.75 (1.33 to 2.30)</td>
<td>0.20 (0.01 to 1.65)</td>
</tr>
<tr>
<td>Sher, 2005a</td>
<td>0.75 (0.60 to 0.86)</td>
<td>0.77 (0.64 to 0.87)</td>
<td>3.21 (1.94 to 5.31)</td>
<td>0.33 (0.20 to 0.54)</td>
</tr>
<tr>
<td>Sher, 2005b</td>
<td>0.79 (0.74 to 0.84)</td>
<td>0.57 (0.50 to 0.63)</td>
<td>1.83 (1.57 to 2.15)</td>
<td>0.36 (0.28 to 0.48)</td>
</tr>
<tr>
<td>Yie, 2005</td>
<td>0.87 (0.75 to 0.94)</td>
<td>0.38 (0.27 to 0.49)</td>
<td>1.39 (1.14 to 1.70)</td>
<td>0.35 (0.17 to 0.72)</td>
</tr>
<tr>
<td>Desai, 2006</td>
<td>0.80 (0.66 to 0.91)</td>
<td>0.43 (0.27 to 0.61)</td>
<td>1.42 (1.03 to 1.94)</td>
<td>0.45 (0.23 to 0.90)</td>
</tr>
<tr>
<td>Fisch, 2007</td>
<td>0.88 (0.81 to 0.93)</td>
<td>0.18 (0.10 to 0.29)</td>
<td>1.08 (0.95 to 1.22)</td>
<td>0.65 (0.34 to 1.26)</td>
</tr>
<tr>
<td>Rebmann, 2007</td>
<td>0.47 (0.36 to 0.58)</td>
<td>0.76 (0.70 to 0.81)</td>
<td>1.96 (1.41 to 2.72)</td>
<td>0.70 (0.56 to 0.87)</td>
</tr>
<tr>
<td>Sageshima, 2007</td>
<td>0.01 (0.00 to 0.05)</td>
<td>0.98 (0.84 to 1.00)</td>
<td>0.34 (0.01 to 16.6)</td>
<td>1.01 (0.96 to 1.07)</td>
</tr>
<tr>
<td><strong>Heterogeneity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cochran’s Q (df = 10)</td>
<td>39.43</td>
<td>162.21</td>
<td>54.85</td>
<td>124.11</td>
</tr>
<tr>
<td><em>P</em> value for Cochran’s Q</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><em>I</em>² statistic</td>
<td>75%</td>
<td>94%</td>
<td>82%</td>
<td>92%</td>
</tr>
</tbody>
</table>
The DORs ranged from 0.92 to 24.82 in the individual studies with an $I^2$ value of 49% indicating moderate heterogeneity (Fig. 3). Therefore, the meta-analysis combined the logs of the DORs, which are derived from sensitivity and specificity. The DerSimonian and Laird random-effects summary DOR was 4.38 (95% CI, 2.93–6.55). With the numbers available, none of the pre-defined categorical and meta-regression analyses conducted to explore heterogeneity was statistically significant (all $P$ values >0.10). More specifically, none of the a priori variables, such as publication year of study, clinical pregnancy rate, women’s age, sample size or infertility treatment (ICSI or IVF), could explain the heterogeneity between studies.

Interestingly, the predefined analyses restricted to 2 × 2 tables including only good quality embryos (six studies) showed a better diagnostic performance (Fig. 4), with a summary DOR of 12.67 (95% CI, 3.66–43.80).

Since DORs are not directly relevant to clinical practice, PPV and NPV based on varying baseline pregnancy rates (pre-test probability) are plotted in Fig. 5.

We found no evidence of publication bias. The funnel plot of DOR versus precision was nearly symmetrical (Fig. 6). Formal statistical tests for publication bias, including the rank correlation test of Begg and Mazumdar (Begg and Mazumdar, 2004) ($z = 0.24; P = 0.41$) and the test of Egger et al. (1997) (intercept, 0.309; 90% CI, −1.195 to 1.814; $P = 0.36$), were not significant.

Discussion

The results of the current systematic review and diagnostic meta-analysis, to the best of our knowledge the first on this topic, indicate that sHLA-G in embryo culture supernatants is moderately helpful to predict the ability to achieve a pregnancy in women undergoing infertility treatment. If the embryos are of good quality, however, sHLA-G has a much better diagnostic performance. Our findings will hopefully stimulate further research with single-embryo culture, single-embryo transfer and a highly sensitive detection technique at the pg/ml level in order to
determine the potential of sHLA-G in culture supernatant for predicting implantation.

The likelihood ratios from the individual studies and the summary DOR (sDOR) suggest that the diagnostic performance of sHLA-G in embryo supernatants is modest. A guide on how to use likelihood ratios in clinical practice suggests that positive likelihood ratios of \( >10 \) or negative likelihood ratios of \( <0.1 \) can provide high diagnostic accuracy, whereas those above \( 5 \) and below \( 0.2 \) give strong diagnostic evidence (Jaeschke et al., 1994). Likewise, a DOR of \( >100 \) indicates high accuracy, \( 25-100 \) indicates moderate accuracy and \( <25 \) indicates an unhelpful test (Jaeschke et al., 1994; Deeks et al., 2001). Yet, a formal judgement upon the magnitude of summary statistics of a diagnostic test also depends on the context to which the test is applied, and on the pre-test probability (Broekmans et al., 2006; Collins et al., 2007).

In the context of assisted reproductive technology, the sDOR of 4.38 indicates that the sHLA-G test is moderately helpful in predicting the likelihood of clinical pregnancy in a group of women undergoing ICSI or IVF. This summary diagnostic statistics is superior to those of some other diagnostic tests that have become part of the routine diagnostic procedure for infertility patients who will undergo assisted reproductive technology, such as, e.g. sperm DNA integrity, early-follicular-phase blood hormones, antral follicle count, ovarian volume and -blood flow, the Clomiphene Citrate Challenge Test, the exogenous FSH ovarian reserve testing and the gonadotrophin agonist stimulation test (Broekmans et al., 2006; Collins et al., 2007).

With a mean baseline clinical pregnancy rate of 45% (pre-test probability) in the current infertility population, the use of the sHLA-G test would theoretically raise the post-test probability of pregnancy to 72% (PPV). Such an improvement in clinical pregnancy rate might be clinically relevant, especially if sHLA-G expression is viewed as a secondary test to distinguish among embryos with an equally good morphology. Other non-invasive tests have been described for this purpose including the metabolic profiling of the culture medium and the gene expression profile of granulosa cells (McKenzie et al., 2004; Seli et al., 2007).

At the time being, the presence of sHLA-G in embryo culture supernatant measured using ELISA systems with a detection limit of 1 ng/ml might not be useful to predict the ability to achieve clinical pregnancy. The amount of sHLA-G the embryo can produce is a matter of debate; 1 ng/ml sHLA-G in the embryo culture supernatant would represent an important proportion of its protein production (Menezo et al., 2006). sHLA-G production might be lower or fluctuating around the actual detection limit, making it difficult to discriminate between true levels and background signals and masking the eventual association between the presence of sHLA-G and clinical pregnancy. Signals above background but below the detection limit have
been found by Sageshima et al. (2007) and ourselves (unpublished data), and considered to be negative. The variation coefficient at the low level is generally not mentioned in the published studies. In order to reliably calculate the relationship between sHLA-G and clinical pregnancy, single-embryo culture and single-embryo transfer are a prerequisite and the ELISA should be made more sensitive by lowering the detection limit as has been done for the first time by Rebmann et al. (2007). Although limited by the small sample volume, the aim should be to titrate the unknown molecule by measuring at least two distinct points since calculating concentrations by using one single point on a slope is tricky.

The timing of measurement during preimplantation development could be critical. It is generally accepted that the maternal-embryonal genome switch in humans occurs at the 4- to 8-cell stage (Braude et al., 1988). Therefore, it is unlikely that significant amounts of any protein produced by the embryo could be detected at day 2 of culture. If any, it would be of maternal origin. Moreover, as Yao et al. (2005) reported that HLA-G mRNA and protein amounts increase during preimplantation development particularly from day 3 onwards, it would be more useful to measure at days 5–6, prior to implantation when the trophoderm, the source of sHLA-G, is cohesive and the blastocyst is fully expanded.

The overall findings presented in our analysis may have important implications for embryo transfer in standard day-to-day clinical practice. According to the data in the individual cohorts, the measures of diagnostic performance always fell below the current standards required to consider a diagnostic test to be effective, even with assay protocols considered to be of very good methodological quality. The published evidence available today indicates that clinical embryologists should continue to select embryos on the basis of currently widely accepted morphological criteria. The presence or absence of sHLA-G, using the currently available methods, is not a valid selection criterion.

Our subgroup meta-analysis suggests that the sHLA-G test might be much more informative in clinical practice for embryos of good quality. A likely explanation is that these embryos are able to produce higher quantities of sHLA-G, resulting in true positive measurements, confirming the above hypothesis, as shown by Rebmann et al. (2007). However, if, and to what extent, the combination of the two modalities (sHLA-G test results and morphologic criteria) might increase accuracy in the diagnostic process remains to be established.

A number of potential limitations should be taken into account. Overall, the validity of any meta-analysis is critically dependent on the internal validity of the studies involved, and we acknowledge that we could only use the aggregated information provided at the study level. In this regard, our analyses could not take into account unmeasured or unreported differences in patient or study characteristics, such as, e.g. underlying causes of infertility or varying infertility treatment modalities. Another limitation of our study is the generalizability (external validity) of these data, which were obtained mainly from individuals from North America and Europe. Given the small number of studies and the small sample sizes available, we were unable to detect statistically significant differences when we explored potential sources of heterogeneity, e.g. infertility treatment. More specifically, the potential impact of the male factor (ICSI) needs to be further investigated. As a general rule, subgroup and meta-regression analyses are always explanatory and hypothesis generating. All the above findings therefore need to be interpreted in this context.

The strengths of our study, on the other hand, include the use of standard meta-analytic procedures for retrieval, and statistical processing of the data collected among a case-mix of women undergoing infertility treatment. All estimates were based on papers published in full in peer-reviewed journals within the past decade, minimizing any effect of secular trends and changes in clinical practice. All studies used a cohort design, and the value of the sHLA-G test was determined using the same ‘gold’ or ‘reference’ standard (clinical pregnancy), limiting verification (work-up) bias. When we formally tested for publication bias, we found none, indicating that our exhaustive search strategy likely captured most relevant studies. Finally, our systematic review and meta-analysis also confirms that quantification and exploration of heterogeneity as a formal goal of meta-analysis presents opportunities to increase the relevance of the published results and enhances the scientific understanding of the studies reviewed.

In summary, at the time being, sHLA-G in embryo culture supernatants is moderately helpful to predict the ability to achieve a pregnancy in women undergoing infertility treatment. If the embryos are of good quality, however, sHLA-G has a much better diagnostic performance.

Our findings underscore the need to address the critically important issues related to single-embryo transfer, single culture condition, and sHLA-G detection threshold.

**Author’s Role**

M.J.V. had the idea for the study. P.H. and M.J.V. designed the study, collected the data, had full access to all of the data, take responsibility for the integrity of the data and the accuracy of the data analysis and are the guarantors for the paper. P.H. did the statistical analysis. All authors provided advice on interpretation of the results and drafted the paper. All authors revised the paper critically and approved the final manuscript.

**Funding**

This work was supported by grants from the Willy Gepts Foundation, UZ Brussel, to M.J.V. M.J.V. is Senior Clinical Investigator of the Fund for Scientific Research—Flanders, Belgium (F.W.O.—Vlaanderen).

**References**


Ishitani A, Geraghty DE. Alternative splicing of HLA-G transcripts yields proteins with primary structures resembling both class I and class II antigens. Proc Natl Acad Sci USA 1992;89:3947–3951.

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;271:703–707.


Submitted on October 31, 2007; resubmitted on January 9, 2008; accepted on February 6, 2008.