**Supplementary File 1**

**Model specification**

For a review of a volunteer’s scan the probability of left (*L*) and right ovary (*R*) being confirmed as ‘seen’ is modelled by a bivariate random effects probit model:

where denotes the cumulative distribution function of the bivariate standard normal distribution, is the linear portion of the model including random effectsand is the correlation between latent error terms. A probit model can be characterized by an underlying latent variable (here the propensity for an expert to confirm a sonographer’s positive scan) and so the observed outcome is defined as:

and the linear portion of the model is:

In more detail is the outcome of scan review for volunteer *i* (1, …, 354) viewed on occasion *j* (1, …, 3) looking at ovary *k* (*L*=left, *R*=right), where the outcome *y*=0 is not confirmed or *y*=1 confirmed by reviewer, is the value for the *a*th (1,…*z*) covariate (fixed effect) for volunteer *i* viewed on occasion *j* looking at ovary *k*. Categorical variables in the model are reviewer (7 parameters) and sonographer (6 parameters)

is the parameter value for the corresponding *a*th covariate for the *k*th equation

 are the correlated random intercepts for volunteer left ovary and volunteer right ovary respectively.

The vector of random intercepts is allowed to covary but is independent from the covariates and the error terms, and normally distributed:

The error terms for of each equation *k* (for the same volunteer *i* and same occasion *j*) are also correlated:

**Correlations**

The intraclass correlation coefficient (ICC) for ovary *k* was calculated using the probit model-based formula:

The correlation between the left and right ovary result for a given volunteer on a given review occasion is:

confidence intervals were derived using the Stata command nlcom which allows for non-linear combinations of estimated coefficients.

**Predictions**

Unfortunately, cmp does not currently offer population-averaged predictions whereby the prediction is based on integrating over the range of the random effect. Instead, it allows only fixed mean predictions, meaning just the fixed portion of the linear predictor is included and one must assume that the prediction is for a subject with a zero value for the random effect(s). In order to obtain these marginal predictions we were able to refit the model using Stata command gsem (for generalized structural equation modelling). However, a bivariate probit model fitted with this command results in rescaled estimates and standard errors.

To obtain the joint probability of a positive outcome for both left and right ovary we utilized the ability of the margins command to generalize the prediction to any function of the standard predictions allowed using the expression() option. Specifically, the expression included Stata’s own cumulative distribution function of the bivariate standard normal binorm, with the population-averaged linear prediction for left and right ovary and correlation (see above) plugged into the full expression. Note the term was represented as the function of the individual variance estimates using their official Stata legend names, rather than a simple scalar, in order to incorporate the appropriate level of uncertainty in the confidence intervals.