Should the post-coital test (PCT) be part of the routine fertility work-up?

Jan W.van der Steeg1,2,3, Pieternel Steures1,2,3, Marinus J.C.Eijkemans1, J. Dik Habbema1, Fulco van der Veen*, Patrick M.M.Bossuyt3, Peter G.A.Hompes4 and Ben W.J.Mol2,5

1Department of Public Health, Erasmus Medical Center, Rotterdam, 2Center for Reproductive Medicine and 3Department of Biostatistics and Epidemiology, Academic Medical Center, Amsterdam, 4Department of Obstetrics and Gynaecology, Vrije Universiteit Medical Center, Amsterdam and 5Department of Obstetrics and Gynaecology, Máxima Medical Center, Veldhoven, The Netherlands

BACKGROUND: This study aimed to determine whether medical history and semen analysis can predict the result of the post-coital test (PCT). METHODS: A previously reported data set of Dutch patients collected between 1985 and 1993 was used. Our study was limited to just patients with an ovulatory cycle. Data were complete for medical history, semen analysis and PCT. We performed logistic regression analysis to evaluate whether these factors could predict the result of the PCT (PCT model). Furthermore, we evaluated the additional contribution of the PCT in the prediction of treatment-independent pregnancy (pregnancy model). RESULTS: Thirty-four percent (179 out of 522) had an abnormal PCT. The PCT model contained previous pregnancy [odds ratio (OR) 2.1; 95% confidence interval (CI) 1.3–3.5], semen volume (OR 0.88; 95% CI 0.77–0.99), sperm concentration (OR 0.96; 95% CI 0.94–0.97), sperm motility (OR 0.97; 95% CI 0.96–0.98) and sperm morphology (OR 2.7; 95% CI 1.2–6.8). The area under the ROC curve of the model was 0.81. In the pregnancy model, the result of the actual PCT could be replaced by the predicted result of the PCT model in about half of the couples, without compromising its predictive capacity. CONCLUSION: The medical history and semen analysis can predict the result of the PCT in ~50% of the subfertile couples with a regular cycle, without compromising its potential to predict pregnancy.

Key words: model/PCT/pregnancy/prognosis/subfertility

Introduction

The use of the post-coital test (PCT) in the basic fertility work-up has been subject to debate over the last 10 years (Griffith and Grimes, 1990; Eimers et al., 1994; Oei, 1998; Cohlen et al., 1999). The PCT traditionally is used as a test to diagnose cervical factor subfertility. An abnormal PCT in the presence of normal semen is considered to reflect cervical hostility. In the absence of tubal occlusion, some argue that cervical factor subfertility can be treated with intrauterine insemination (IUI). Among the five randomized studies in which IUI is compared with no treatment, two clearly reported a beneficial result of IUI (te Velde et al., 1989; Check et al., 1995), whereas three other studies indicated the absence of such a beneficial result (Friedman et al., 1989; Chaffkin et al., 1991; Kirby et al., 1991). If IUI were to be effective, the PCT has potential value in the work-up for subfertility. However, the routine use of the PCT has been reported to lead to more treatment without an associated increase in pregnancy rates (Oei et al., 2001). If it were to be effective, the PCT has potential value in the work-up for subfertility. However, the routine use of the PCT has been reported to lead to more treatment without an associated increase in pregnancy rates (Oei et al., 2001). In this view, the routine use of the PCT in the basic fertility work-up would be unnecessary. Apart from the diagnostic aspect of the PCT, some authors have also emphasized its prognostic value in the prediction of treatment-independent pregnancy. An abnormal PCT would decrease the probability of treatment-independent pregnancy 2- to 3-fold (Eimers et al., 1994; Oei et al., 1995a; Snick et al., 1997). However, despite these findings, Eimers et al. (1994) advocated the use of the PCT in the routine fertility work-up as a prognostic test, whereas Oei et al. (1996) argued against the routine performance of the PCT. Only one study made a distinction between couples with a long duration of subfertility and couples with a shorter duration of subfertility (Glazener et al., 2000). It was concluded that the PCT has additional prognostic value, but only in couples with a duration of subfertility of <3 years.

Therefore, there is at present no clear-cut evidence that the PCT has diagnostic or prognostic value as a routine test for subfertile patients. However, the PCT might be of use for some couples, whereas it is not useful for others. Therefore, if one were able to select a group of subfertile couples, based on their history and semen analysis, in which the PCT has limited additional value, the routine performance of this test would not
be necessary. In the remaining couples, the routine performance of this test could be evaluated by means of a randomized controlled trial.

The present study aims to evaluate whether the results of the PCT can be predicted from the patients’ history and the semen analysis. Furthermore, we assessed whether omission of the PCT in couples in whom the PCT result can be predicted would compromise the capacity to predict treatment-independent pregnancy from data obtained at the basic fertility work-up.

Materials and methods
We used the data reported on previously by Snick et al. (1997). This data set contains information on consecutive couples referred by general practitioners, presenting with subfertility between January 1, 1985 and December 31, 1993 to the Walcheren hospital, Vlissingen, The Netherlands. Subfertility was defined as failed conception after 1 year of unprotected intercourse. Since the PCT cannot be performed in anovulatory women, we limited data to patients with an ovulatory cycle. The menstrual cycle was considered ovulatory if the duration of the menstrual cycle was <8 weeks and if serum progesterone was >18 nmol/l in the luteal phase. Data had to be complete for medical history and semen analysis. The medical history consisted of data on male and female age, subfertility being primary or secondary, and history and semen analysis. The semen analysis consisted of data on sperm concentration, volume, and motility, and on the sperm morphology.

| Table I. Baseline characteristics of the data of Snick et al. (1997); couples with an ovulatory cycle, and results of the univariable and multivariable regression analysis |
|---------------------------------|-----------------|-----------------|-----------------|
| **Result of the PCT**           | **Abnormal 179 (34%)** | **Normal 343 (66%)** | **OR 95% CI** |
| **Univariable analysis**        |                  |                  |                |
| Male age (mean, range)          | 28.8 (18–40)     | 29.9 (20–42)     | 0.95 (0.91–0.99) |
| Female age (mean, range)        | 30.7 (22–60)     | 31.7 (21–50)     | 1.0 (0.96–1.04)  |
| Duration of subfertility in years (median, range) | 1.5 (0.9–11.5) | 1.3 (0.8–11.5) | 1.0 (0.89–1.2) |
| Primary subfertility            | 149 (83%)        | 221 (64%)        | 2.7 (1.7–4.2)   |
| Semen characteristics            |                  |                  | 2.1 (1.3–3.5)  |
| Volume (ml) (median, range)     | 3 (0–8)          | 3 (0–12)         | 0.85 (0.76–0.95) |
| Concentration (10^6/ml) (median, range) | 12 (0–300) | 47 (0–400) | 0.94 (0.92–0.95) |
| Motility (%) (median, range)    | 15 (0–80)        | 40 (0–85)        | 0.96 (0.95–0.97) |
| Abnormal morphology (<20%)      | 25 (14%)         | 10 (3%)          | 4.9 (2.3–10.6)  |

OR for sperm concentrations from 0 to 40 × 10^6/ml. Higher concentrations have the same OR as concentrations of 40 × 10^6/ml.
The chance of an abnormal PCT can be calculated from the multivariable model with the formula: probability = 1/[1 + exp(±0.74 + semen volume × –0.13 + sperm concentration × –0.041 + motility × –0.03 + abnormal morphology × 0.99)].

To evaluate the performance of the logistic model, the area under the receiver operating characteristic (ROC) curve was calculated. Sensitivity was defined as the fraction of patients with an abnormal PCT that was predicted correctly, whereas specificity was defined as the fraction of patients with a normal PCT that was predicted correctly.

Internal validation was performed with bootstrapping. Bootstrapping is a technique to create new data sets by random drawing from the sample with replacement. In each of these new data sets (n = 1000), the same multivariable regression was assessed. By analysing the difference of the prognostic models, a shrinkage factor was calculated to reduce the overfit of the created model. Calibration was evaluated with the Hosmer and Lemeshow goodness-of-fit test statistic.
Prediction of pregnancy. To evaluate whether the PCT, performed in a selected group, can predict treatment-independent pregnancy, we used Cox proportional hazard analysis to develop three pregnancy prediction models. The initial model contained the variables duration of subfertility, age of the female partner, previous pregnancies for the couple, and the semen parameters volume, concentration, motility and morphology, but not the PCT (pregnancy model I). We then developed a second model, in which both semen analysis and PCT were always performed (pregnancy model II). Finally, we developed a third pregnancy prediction model (pregnancy model III), in which the PCT was only performed in cases where the probability of an abnormal PCT was in between the cut-off points 0.2 and 0.6. These points correspond to the chance of an abnormal PCT of 20 and 60%, calculated with the PCT model. In cases where the chance was <20% or >60%, pregnancy model III used the prediction for treatment-independent pregnancy made without the PCT.

The performances of the models I, II and III were assessed by comparing the area under the curve (AUC) of the three ROCs. The AUCs were calculated using the Kaplan–Meier method as proposed by Heagerty et al. (2000), and the 95% confidence intervals (CIs) were calculated by bootstrapping.

Results

The data set contained 726 patients of whom 587 (81%) couples had an ovulatory cycle. The PCT was available in 531 couples (90%). In nine couples, data were incomplete. The baseline characteristics of the 522 couples that were included in the study are shown in Table I. There were 179 (34%) couples with an abnormal PCT. Female age, male age as well as duration of subfertility were almost equal for the subgroup with a normal PCT and the subgroup with an abnormal PCT. After the PCT was performed, tubal pathology was diagnosed in 109 (22%) women, in whom 55 (11%) had reconstructive surgery of the Fallopian tubes. During the follow-up period, 228 spontaneous pregnancies were registered, comprising 31 spontaneous abortions, seven ectopic pregnancies and 190 ongoing pregnancies.

For female and male age, duration of subfertility, semen volume and sperm motility, we found a linear association with the result of the PCT on the logistic scale. In contrast, a non-linear association was observed for the sperm concentration and the result of the PCT on the logistic scale (Figure 1). We therefore decided not to distinguish between sperm concentrations >40 \times 10^6/ml, and to model lower values in a linear way.

Table I also shows the results of the univariable and multivariable analyses. Primary subfertility of the couple was associated with an increased probability of an abnormal PCT, as were low semen volume, lower sperm concentrations, low sperm motility and abnormal sperm morphology. These factors were therefore included in the PCT prediction model. Internal validation by bootstrapping showed a shrinkage factor of 0.91, which corresponds to 9% overfit of the initial model. All odds ratios (ORs) and the intercept of the multivariable model were adjusted for this 9% overfit.

The PCT prediction model had an area under the ROC curve of 0.81 (95% CI 0.77–0.85), indicating an excellent discriminative performance (Figure 2A). Figure 2B shows how the number of normal and abnormal PCTs depends on the probability of an abnormal PCT (divided into 10 groups of 10% prediction chance). This figure shows that 27 of the 209 patients (13%) in whom the calculated probability of an abnormal PCT was <20% indeed had an abnormal PCT. On the
other hand, the PCT result was abnormal in 83 of the 109 patients (76%) in whom the calculated probability of an abnormal PCT was 60% or higher. Tubal pathology was diagnosed in 59 couples that had a calculated probability of an abnormal PCT result of <20%, in 37 couples with a calculated probability between 20 and 60%, and in 13 couples with a calculated probability >60%.

Figure 3 shows the association between the probability of an abnormal PCT as predicted by the PCT model, and the mean observed test results.

Finally, we assessed the prognostic performances of the three pregnancy prediction models. Pregnancy model I, i.e. the model without the PCT, had an AUC of 0.61 (95% CI 0.55–0.67). Pregnancy model II, i.e. the model with the PCT, had an AUC of 0.65 (95% CI 0.59–0.71). This AUC was statistically significantly better (the difference in AUCs is 0.04; 95% CI 0.004–0.07). Pregnancy model III, including the PCT results only in those couples that had a chance of an abnormal PCT between 20 and 60%, had an AUC of 0.64 (95% CI 0.58–0.70). The improvement between pregnancy model I and pregnancy model III was borderline statistically significant (difference in AUCs is 0.03; 95% CI –0.006 to 0.06). The AUC of pregnancy model III was comparable with that of pregnancy model II, with a non-statistically significant difference (difference in AUCs is –0.01; 95% CI –0.03 to 0.01).

Figure 4A shows the comparison between the probability of treatment-independent pregnancy predicted from pregnancy model I and pregnancy model III. Couples with an abnormal PCT had a poorer prognosis calculated with pregnancy model III than with pregnancy model I. The false-positive predicted PCTs are PCTs that are thought to be abnormal from the PCT model, but turn out to be normal when actually performed. Similarly, false-negative predicted PCTs are PCTs thought to be normal from the PCT model, but appear to be abnormal when performed in reality. The false-positive and false-negative predicted PCTs are spread equally over the scatter plot. Figure 4B shows the comparison between the chances of treatment-independent pregnancy predicted from pregnancy model II and pregnancy model III. For almost all couples, the probabilities of a treatment-independent pregnancy at 6 months are equal in model II and model III, except for the few outliers. These outliers are due to the false-positive and false-negative predicted PCTs.

Figure 2. (A) Receiver operating curve (ROC) of the multivariable logistic regression model for the prediction of an abnormal PCT. Points located on the curve correspond to the cut-off values of the predicted PCT (e.g. a cut-off value of 0.1 considers patients with a probability below 0.1 of having a normal PCT, and above 0.1 of having an abnormal PCT). (B) Clustered bar chart. For each probability of an abnormal PCT, the distribution of total observed normal and abnormal PCTs is shown.
Discussion

This study shows that medical history and semen analysis can predict the result of the PCT in ~50% of subfertile couples with a regular cycle. For the other 50% of the couples, i.e. patients with a chance of an abnormal PCT between 20 and 60%, the result of the PCT cannot be predicted reliably. Performance of the PCT might therefore have clinical relevance in these patients. We showed that replacing the result of the actual PCT by the predicted result of the PCT, in patients with a high or a low chance of an abnormal PCT, does not compromise the capacity to predict pregnancy from the basic fertility work-up. On the other hand, complete omission of the PCT from the prediction model results in a decreased prognostic performance.

Our study has some limitations. First, the data set of Snick et al. (1997) was collected >10 years ago. Since then, treatment has improved and the characteristics of the population have changed, with women starting to conceive at older age, and couples contacting fertility clinics at an earlier stage of subfertility. A second limitation is the fact that only one gynaecologist performed all PCTs in a rather homogeneous population. Consequently, in a clinic with a more heterogeneous population or in a clinic in which the PCT is performed in a less standardized setting, the capacity of the PCT prediction model might decrease. The problem of standardization of the PCT has been described before (Oei et al., 1995b).

In view of these limitations, we recommend that our PCT model should be evaluated in other populations.

A strength of this study is that application of the PCT model will lead to a reduction in the number of PCTs performed in clinical practice. With the prediction of the PCT model, the clinician can decide whether or not to perform the PCT. For example, a couple with secondary subfertility and semen volume of 4 ml, and a sperm concentration of 85 × 10^6/ml, 10% of which are progressively motile and with a normal morphology, will have a chance of having an abnormal PCT of 19%. Therefore, the result of the PCT will most probably be normal and offering the PCT to such a couple is not likely to have additional value. Consequently, a primary subfertile couple with a semen volume of 4 ml, sperm concentration of 10 × 10^6/ml, of which 30% are progressively motile and 30% have normal morphology, will have a chance of an abnormal PCT of 51%. Therefore, the result of the PCT cannot be predicted reliably (probability between 0.2 and 0.6). The performance of the PCT is of additive value for this couple. Thus, in this way, the clinician can decide, based on the PCT prediction, for which couples the performance of the PCT is of clinical relevance and for which it is not.

In this study, we have chosen cut-off values of 20% as the upper level for normal PCT results and 60% as the lower border for abnormal PCT results. We have set the cut-off values at 20 and 60% supported by the thought that below a chance of 20% on an abnormal PCT, more than five PCTs should be performed to detect one abnormal PCT (in our data, 13% of the patients in this group indeed had an abnormal PCT). In the opposite way, with a chance of >60% on an abnormal PCT, three out of five PCTs will be abnormal (in our data, 76% of the patients in this

Figure 4. (A) Scatter plot in which the chance to conceive independently of treatment, in a period of 6 months, predicted with pregnancy model III is compared with pregnancy model I. (pred. = predicted). (B) Scatter plot in which the chance to conceive independently of treatment, in a period of 6 months, predicted with pregnancy model III is compared with pregnancy model II.
group indeed had an abnormal PCT). With these cut-off values, the false-negative fraction will be twice as small as the false-positive fraction. False-positive test results can lead to overtreatment, while false-negative test results can lead to undertreatment. Therefore, these cut-off values for the PCT model result in a lower number of overtreated patients in comparison with the number of undertreated patients. Only when the exact treatment for patients with an abnormal PCT is known, together with its side effects and costs, can a more considered choice for the cut-off values be made. These cut-off values might be helpful thresholds guiding future trial design.

Snick et al. (1997) reported a better performance of his prognostic models than we found in the present study; for the model with the PCT, we found an AUC of 0.65 (prediction of treatment-independent pregnancy at 12 months) whereas Snick et al. reported 0.79. For the model without the PCT, we found the AUC to be 0.61 versus 0.76 reported by Snick et al. (1997). This difference might be explained by the fact that Snick et al. used all 726 patients, whereas we excluded women with ovulation disorders. Another explanation might be that we considered the data as censored survival data, whereas Snick et al., using a more conservative approach, considered pregnancy as a dichotomous event. We used a method recently introduced by Heagerty et al. (2000). The comparison of the method of Heagerty with the more conservative approach showed a decrease of 10% in the AUC of the three pregnancy models.

The effectiveness of the PCT has been assessed in a randomized clinical trial, which compared a strategy in which all couples had a PCT with a strategy in which none of the couples had a PCT (Oei et al., 1998). At first, this type of study is considered to result in the standard of evidence. This trial attempted to replicate the imperfections of clinical practice in order to be relevant (Smith et al., 2003). However, methodologically, this clinical management trial has been criticized by several authors (Cohlen et al., 1999; Hendry, 1999; Hull and Evers, 1999). It has been stated that the included sample was not sufficiently exclusive (20% ovulatory disorder), that the intervention was used incompletely (36% missing PCTs in the intervention group) and that a specific response to the test results was lacking. These methodological problems might have influenced the results of the trial (Bossuyt et al., 2000). Furthermore, the occurrence of multiple pregnancies was not studied and treatment with IUI was always performed in a stimulated cycle. In contrast, our PCT model enables us to select a subgroup of patients in which the performance of the PCT might still be of use, thereby potentially identifying patients with a cervical factor in whom treatment with IUI without stimulation could prevent the occurrence of multiple pregnancies without compromising pregnancy rates.

Based on our findings, we suggest that a rearrangement of the basic fertility work-up is worthy of further evaluation. After taking patients’ history and performing semen analysis, a prediction of PCT outcome could be calculated. Depending on this prediction, the clinician could decide to plan the PCT afterwards or leave it out, instead of routinely performing it in all patients. Before such an approach can be implemented in clinical practice, two other conditions have to be fulfilled. First, the present model has to be validated in other populations. Secondly, it has to be established whether the PCT has value in the selection of patients that will benefit from IUI without ovarian hyperstimulation, and patients that benefit from IUI with hyperstimulation. In contrast to male factor subfertility, in which there is evidence that IUI without hyperstimulation is as effective as IUI with hyperstimulation, evidence for the use or non-use of ovarian hyperstimulation in IUI for cervical factor subfertility is lacking. If IUI without hyperstimulation were to give results comparable with IUI with hyperstimulation, the PCT can prevent couples from additional side effects, e.g. ovarian hyperstimulation syndrome and twin pregnancies, and reduce costs.

In conclusion, we found that the result of the PCT can be predicted with data from history and semen analysis in about half of the couples with a regular cycle. If this is confirmed at external validation, and if the PCT remains an important prognostic and diagnostic test, our findings suggest that a trial to evaluate the PCT in the basic fertility work-up would involve performance of a PCT planned from the history and semen analysis.

Acknowledgements
The authors thank Herman Snick, who kindly agreed to the use of his data for the present study. The current project has been funded by ZonMW, grant 945-12-002.

References

Submitted on November 25, 2003; accepted on February 13, 2004