Human Reproduction, Vol.27, No.2 pp. 550–557, 2012
Advanced Access publication on November 29, 2011 doi:10.1093/humrep/der409

human reproduction

ORIGINAL ARTICLE *Reproductive endocrinology*

Monitoring of ovarian activity by measurement of urinary excretion rates of estrone glucuronide and pregnanediol glucuronide using the Ovarian Monitor, Part II: reliability of home testing

Leonard F. Blackwell[1,*], Pilar Vigil[2], Barbara Gross[†], Catherine d'Arcangues[3], Delwyn G. Cooke[4], and James B. Brown[†]

[1]Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand [2]Pontificia Universidad Católica de Chile, Santiago, Chile [3]World Health Organisation, Geneva, Switzerland [4]Manawatu Diagnostics, Palmerston North, New Zealand

*Correspondence address. Tel: +64-6-3569099; Fax: +64-6-3505682; E-mail: l.f.blackwell@massey.ac.nz

Submitted on June 22, 2011; resubmitted on October 20, 2011; accepted on November 1, 2011

BACKGROUND: The UNDP/WHO/World Bank/Special Programme of Research, Development and Research Training in Human Reproduction (Geneva) set up a study to determine whether it is feasible for women to monitor their ovarian activity reliably by home testing. Daily self-monitoring of urinary hormone metabolites for menstrual cycle assessment was evaluated by comparison of results obtained with the Home Ovarian Monitor by untrained users both at home and in study centres.

METHODS: Women collected daily data for urinary estrone glucuronide (E1G) and pregnanediol glucuronide (PdG) for two cycles, then the procedure was repeated in the women's local centre (in Chile, Australia or New Zealand) giving a total of 113 duplicate cycles. The tests were performed without the benefit of replicates or quality controls. The home and centre cycles were normalized and compared to identify assay errors, and the resulting home and centre menstrual cycle profiles were averaged.

RESULTS: Reliable mean cycle profiles were obtained with the home and centre excretion rates agreeing to within 36 $\pm$ 21 nmol/24 h for E1G and 0.77 $\pm$ 0.28 μmol/24 h for baseline PdG values (1–5 μmol/24 h). The cycles had a mean length of 28.1 $\pm$ 3.1 days ($n = 112$; 5th and 95th percentiles: 24 and 35 days, respectively), a mean follicular phase of 14.8 $\pm$ 3.1 days ($n = 107$; 5th and 95th percentiles: 11 and 21 days) and a mean luteal phase length of 13.3 $\pm$ 1.5 days ($n = 106$; 5th and 95th percentiles: 11 and 17 days), calculated from the day of the LH peak.

CONCLUSIONS: The study confirmed that the Ovarian Monitor pre-coated assay tubes worked well even in the hands of lay users, without standard curves, quality controls or replicates. Point-of-care monitoring to give reliable fertility data is feasible.

Key words: estrone glucuronide / menstrual cycle profiles / pre-coated tubes / pregnanediol glucuronide / urine home test

Introduction

The Ovarian Monitor is a device developed to measure estrone glucuronide (E1G) and pregnanediol glucuronide (PdG) in urine samples, using pre-coated assay tubes in a home or in a point-of-care setting. The assays utilize the principle of a homogeneous enzyme immunoassay (Brown *et al.*, 1988, 1989, 1991; Blackwell *et al.*, 2003) in which an antibody to a ligand specifically inhibits the enzyme activity of an enzyme–ligand conjugate. Knowledge of E1G and PdG excretion rates allows identification of the fertile window (Blackwell and Brown, 1992; Blackwell *et al.*, 1998) and also differentiates between ovulatory fertile cycles and other cycle variants (Brown,

[†] Deceased.

2011). Ovarian activity can change abruptly within a 24 h period, with profound consequences; thus a device that allows daily monitoring of hormone levels non-invasively is needed to monitor the ovarian cycle.

In its search for reliable systems that women can operate themselves to monitor their ovarian activity, the UNDP/WHO/World Bank/Special Programme of Research, Development and Research Training in Human Reproduction (Geneva) set up a study, at three locations (Santiago, Chile; Sydney, Australia and Palmerston North, New Zealand), to determine whether the home Ovarian Monitor was suitable for this purpose. The first part of this study involved a comparison of Monitor results with corresponding laboratory radioimmunoassay results, performed with quality controls, standard curves and replicates. The data for both assays were determined by trained personnel in a London laboratory (Blackwell *et al.*, 2003) for the same urine samples from 20 cycles provided by women at the study locations. The conclusion was that the Monitor pre-coated tube assays produced quantitative results that were as accurate as the laboratory assays at the levels encountered during the ovulatory cycle. The timing of the pre-ovulatory E1G peak and the post-ovulatory PdG rise (Blackwell *et al.*, 1998), agreed within experimental error between the Monitor and RIA systems.

The objective of the present study is to test, under controlled conditions, the hypothesis that the Ovarian Monitor pre-coated assay tubes also give reliable results for menstrual cycle assessment when carried out by lay users in the home or in a study centre, as single measurement assays, without the benefit of assay replicates, standard curves or quality controls. To test this hypothesis, we compared the shapes of 113 menstrual cycle profiles obtained by the participants using the Monitor at home, with those duplicated by lay users for the same urine samples in the study centre, at each of the three locations. If the Monitor results obtained by the home users are reliable, the individual home and centre profiles for both the E1G and PdG excretion rates, after normalization, should be identical within the experimental error. Agreement of the profiles would therefore demonstrate the reliability of the data obtained as single measurement assays at home, and disagreements would identify any assay errors.

Materials and Methods

Study protocols

Women using a recognized method of natural family planning were recruited at each of three locations (Palmerston North, New Zealand; Santiago, Chile; Sydney, Australia). A description of these women has been given previously (Blackwell *et al.*, 2003); the main recruitment criterion was that they were ovulating as judged by a history of regular (20–40 days) menstrual cycles and ovulatory patterns of cervical mucus production and/or basal body temperature, for the previous three cycles. All participants signed an informed consent form approved by WHO before entering this study (#90905). Protocols were conducted in accordance with the WHO ethical guidelines for research involving human subjects. After initial instructions on how to use the Ovarian Monitor pre-coated assay tubes that were produced by St Michael's Research Foundation (Melbourne, Australia), the women collected daily urine specimens for periods of at least 3 h and at any convenient time of the day as described in detail previously (Cooke *et al.*, 2007). For the 10 days around the expected time of ovulation, ~3 ml of urine was taken for laboratory analysis of LH before the bulk sample was diluted to 150 ml/h. The variations in hormone concentration due to changes in urine volume production are compensated for by time-dilution of all samples to a constant volume of 150 ml/h (Brown *et al.*, 1988, 1989, 1991; Blackwell *et al.*, 2003; Cooke *et al.*, 2007), and expressing the results as excretion rates (amount excreted per unit time). The women ($n = 62$) measured their E1G and PdG levels each day at home on their freshly collected and time-diluted samples using the pre-coated assay tubes for two complete menstrual cycles to give a total of 113 cycles. Their Monitor data were recorded as transmission change ($\Delta T = T_t - T_o$), where T_o is the transmission of the substrate solution at time = 0 and T_t is the transmission after the clearing reaction has run for t minutes (Blackwell *et al.*, 2003), and are referred to simply as Monitor readings, $\Delta T/20$ min for the E1G assay and $\Delta T/5$ min for the PdG assay. A longer time period was necessary for the E1G assay since the urinary E1G excretion rates are much lower than the PdG excretion rates. The women froze aliquots of these urines immediately in plastic tubes containing boric acid as a preservative (Blackwell *et al.*, 2003).

Aliquots of the daily samples from each woman were sent to the appropriate centres and re-analysed also using pre-coated assay tubes. The study centre repeat analyses were carried out by lay operators who had no previous experience with the use of the Monitor and who were given the same training in its use as the study participants received. These tubes were usually from a different batch to those used by the women at home, and since the urine samples were presented to each centre only at the end of each cycle, a version of the Monitor test was used, which included a heating block for simultaneous analysis of multiple samples, allowing the cycle to be analysed in a few runs (Blackwell *et al.*, 2003). The study coordinators at the Palmerston North and Sydney centres had some previous experience with the Ovarian Monitor and taught the lay users to operate it and to use the pre-coated assay tubes; however, the centre at the Santiago location successfully established the Monitor system and taught the lay users based solely on written instructions.

Assay methods

The details of the pre-coated tube assays are given in full in part I of this series (Blackwell *et al.*, 2003). The sensitivities of the assays are set to include the most important ranges of values required to identify the cyclical periods of fertility and infertility during the human ovulatory cycle (Blackwell *et al.*, 2003). Urine samples giving a Monitor reading (transmission change) >300 for either the E1G or the PdG pre-coated assay should have been diluted and re-analysed according to the study protocols, but this was rarely carried out.

Data analysis

Identification of one day spikes and troughs in the Monitor readings

A computer database program was constructed to read the data file provided by the UNDP/WHO/World Bank/Special Programme of Research, Development and Research Training in Human Reproduction (Geneva). The program was used to display the cycles individually and process the data, including identifying the cycles with one day spikes or troughs in metabolite concentrations. To identify the one day spikes and troughs, a standard normal variate transformation (Krieg *et al.*, 1999; Cooke *et al.*, 2007) was carried out on the Monitor readings first. This was necessary because the women at home and the operators in the local centre were not using standard curves to transform their Monitor readings into excretion rates. Since they were using different batches of pre-coated assay tubes, the ranges of the Monitor readings could differ considerably. The menstrual cycle profiles were normalized individually by subtracting the mean value for the complete cycle profile from each daily Monitor reading for the cycle and dividing the result by the standard deviation for the cycle (Cooke *et al.*, 2007). The normalized profiles for each

cycle should then overlap closely day to day, since the same urine samples were being analysed by both operators. Thus, one day spikes and troughs in either the home or centre values for a day were easily recognized. An individual day was flagged as a potential spike day if the normalized daily difference (value on day_n−value on day_{n-1}) in either the home or centre data were one or more standard deviations above the normalized values for the two flanking days. If the daily difference within either a home or centre profile was one or more standard deviations less than the values for both of the two flanking days, a potential trough day was flagged. To be accepted as a true one day spike or trough day, the absolute value for the normalized differences on a flagged day between the normalized home and centre results also had to be equal to or greater than 1SD. The corresponding non-flagged home or centre value was accepted as the valid raw result. A measure of the closeness of the fit between the normalized home and centre menstrual cycle profiles after removal of the one day spike or trough values for each cycle at each location was calculated as $\sqrt{(\sum d^2/n)}$, the root mean square (RMS) of the normalized home-centre daily differences (d) where n is the number of days.

Calculation of average daily hormone excretion rates

The computer program counted and removed the values that were one day spikes and troughs from further calculation and then converted the remaining valid Monitor readings into the corresponding E1G and PdG hormone excretion rates. This was done independently for both the home and centre menstrual cycle profiles for each cycle day, by interpolation from a Boltzman sigmoidal fit (Blackwell *et al.*, 2003) to the appropriate standard curve provided with each batch of pre-coated assay tubes from Melbourne. The E1G excretion rates were expressed as nmol/24 h and for the PdG excretion rates the units were μmol/24 h. These derived excretion rates can be converted into more familiar nmol/l and μmol/l by division by 3.6, the number of litres of urine that would be collected in 24 h at a rate of 150 ml/h. It should be noted however that the actual urine collections were not 24 h collections. The minimum collection time was 3 h and most of the urine collections were between 3 and 10 h. Finally, the valid home and centre E1G and PdG pairs of excretion rate data for each cycle day were compared, and the mean daily differences were calculated for the complete cycle. A measure of the typical difference (S) for a cycle between the home and centre operators was calculated as $S = \sqrt{(\sum d^2/2n)}$, where d is the difference between the home and centre excretion rate for each cycle day and n is the number of pairs of home-centre data in the cycle. Any pair of values in a cycle for which $d \geq 2.8S$ (or $\sim 2\sqrt{2}S$) implies, with a probability of ~95% (Snedecor and Cochran, 1989), that at least one of the values is an outlier. Thus, that pair of home and centre values could not be used to give an average for that cycle day. The remaining valid pairs of values were averaged to give the mean E1G and PdG excretion rates for each cycle day, and finally the average daily results for each cycle were displayed graphically. The range of valid excretion rates for E1G was 20–750 nmol/24 h and for PdG it was 1–36 μmol/24 h.

Definition of cycle parameters

The first/pre-ovulatory E1G rise day following menses was determined by eye as the first day that the E1G excretion rate increased significantly above preceding values by >50 nmol/24 h. The rise was also part of at least three consecutive rising values. The ovulatory E1G peak day was the last day of elevated E1G values before a sharp fall that preceded the post-ovulatory rise in the PdG excretion rate. Note that the first and pre-ovulatory rise days are the same if there are no early E1G peaks. The early E1G peaks are defined by Alliende (2002). The day of the PdG threshold (Blackwell *et al.*, 1998, 2003) was the day that the PdG value was ≥7.0 μmol/24 h. Because we observed multiple LH peaks in the data, sometimes it was necessary to choose the correct LH peak day by its proximity to the E1G peak day and its occurrence prior to the post-ovulatory rise in PdG values. Therefore, the LH peak day was defined as the day with the highest LH value close to the E1G peak day and on, or before, the first significant increase in PdG excretion rate. The follicular phase is the number of days from the first day of bleeding to the day of the LH peak inclusive. The luteal phase is the number of days following the LH peak to the day before the commencement of menses for the next cycle.

Statistical methods

The arithmetic means and standard deviations for n values of parameters such as the follicular phase, luteal phase and cycle lengths, the RMS values, ΔT values and excretion rates were all calculated by the column statistics package of Graphpad Prism, version 5. Statistical significance was calculated by comparison of pairs of means and standard deviations using the unpaired t-test package of Prism, which yielded a two-tailed P-value in each case. The significance of the percentage of errors encountered by the Monitor users, between home and centre in each location, and between locations, was assessed by a χ^2 calculation using the CHITEST function of Excel. The means and standard deviations used to construct the composite E1G, PdG and LH profiles were calculated by the computer database program for each cycle day when all of the cycles were aligned in columns according to the E1G peak day.

Results

Initial data screening

An example of the normalization procedure to identify the valid pairs of raw home and centre data produced by the lay users is shown for the E1G values for the first cycle from a participant from Santiago (Fig. 1A; 500 4S). The mean Monitor reading for the pre-coated assay tubes used at home was 245 ± 46 ΔT/20 min for the complete cycle and, for the pre-coated assay tubes used in the Santiago centre, the corresponding cycle mean was significantly less (168 ± 52 ΔT/20 min). Using these statistical values, the daily data for the two menstrual cycle profiles (home and centre) were normalized (Krieg *et al.*, 1999; Cooke *et al.*, 2007) and plotted (Fig. 1A). Three 1 day troughs in the normalized E1G values were seen clearly by visual inspection and flagged by the computer program, one occurring in the home data (Day 13) and two in the centre data (Days 9 and 23). It should be noted that for each flagged day, the corresponding home or centre partner value was not a spike or trough (Fig. 1A) but fitted in with the trend of the neighbouring values. The absolute difference between the flagged normalized values in one profile and their non-flagged partners in the other for each of these 3 days was greater than 1SD. Thus, they were all confirmed as measurement errors—one day trough days in this case. The three tagged normalized one day trough values were counted and removed to give the overlapping normalized menstrual cycle profiles shown in Fig. 1B. The RMS value of the differences for these two profiles was 0.63.

A similar analysis was applied to the PdG data, and examples are given for the second cycle from participant 22G from Sydney (Fig. 2A) and the first cycle from participant 23B from Palmerston North (Fig. 2B). In each case, the two normalized menstrual cycle profiles were identical within experimental error as reflected in their low RMS values (0.19 and 0.22, respectively), and no one day spike or trough days were identified in either cycle. The excellent agreement between the home and centre normalized profiles for all of the PdG

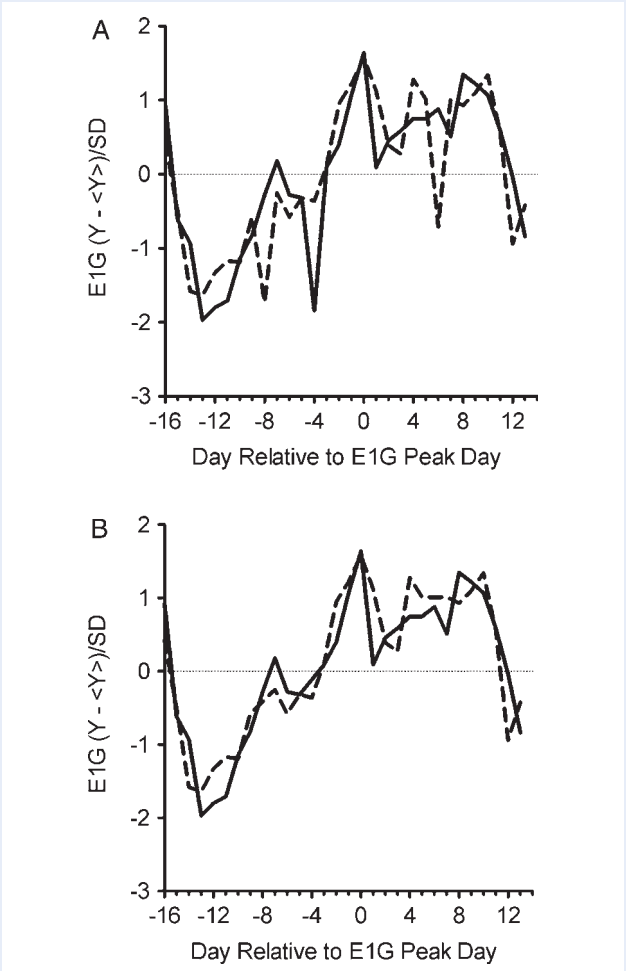

Figure 1 (**A**) Normalized menstrual cycle E1G profiles using subject's raw data for the first cycle from participant 500 4S from Santiago (ΔT/20 min, solid line) and the corresponding centre data (Δ/20 min, broken line). Normalization was carried out as described in the text (Y = value, $<Y>$ = mean value and SD = standard deviation). (**B**) The same data after removal of three troughs, one in the home data and two in the centre data.

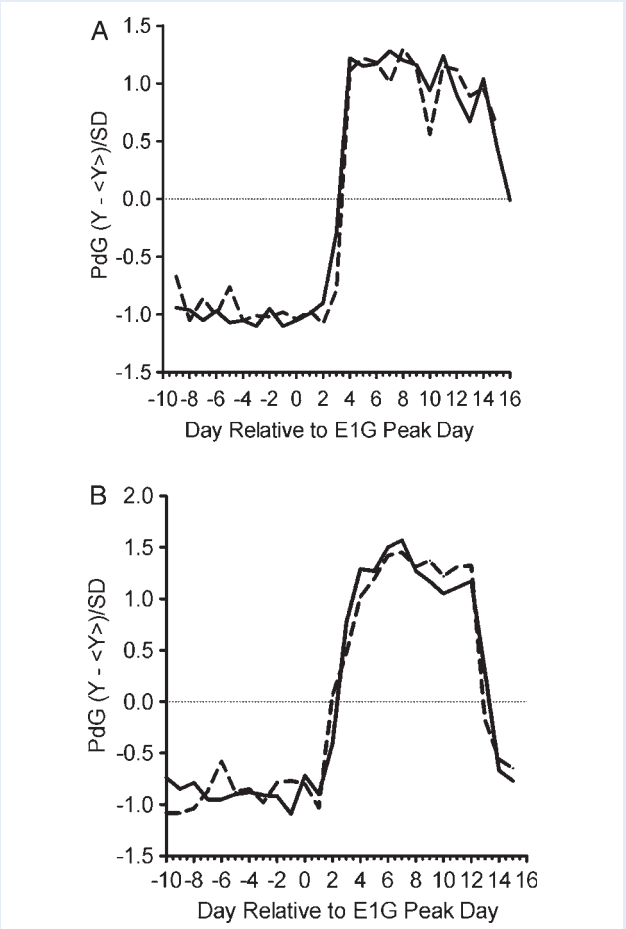

Figure 2 (**A**) Comparison of normalized PdG menstrual cycle profiles for the second cycle of participant 1495 22G from Sydney constructed from the subject's raw data (ΔT/5 min, solid lines) and the centre repeats (ΔT/5 min, broken line). (**B**) Comparison of normalized PdG menstrual cycle profiles for the first cycle of participant 1830 23B from Palmerston North constructed from the woman's raw data (ΔT/5 min, solid lines) and the centre repeats (ΔT/5 min, broken line).

data was similar to that shown in Fig. 2A and B irrespective of the study location. Similar analyses of the total E1G and PdG data for one day spikes and troughs was carried out for all cycles over the three study locations.

Comparison of home and centre Monitor readings

The total number of individual Monitor readings removed from each cycle for E1G and PdG as tube or operator faults for both the home and centre data is given in Table I for each location and overall. Overall, 4.9% of the E1G values measured by the women at home were classified as either one day spikes or troughs and hence removed from the calculations. For the data repeated in the local centres with different batches of pre-coated assay tubes, the comparable overall figure was 4.0% of the E1G values, which was not significantly different ($P = 0.091$). There was no significant difference between the percentages of one day spikes and troughs in the data collected at home for the three study locations (all $P > 0.05$). However, for the centre comparisons, although there was no difference ($P = 0.328$) between Santiago (4.8%) and Sydney (5.8%) for the E1G values, the Palmerston North centre E1G data had a much lower percentage (1.7%) of one day spikes or troughs than either (Table I; versus Santiago, $P \leq 0.0001$; versus Sydney, $P \leq 0.0001$). This was also true when the Palmerston North home data were compared with its corresponding centre data (Table I; $P \leq 0.0001$). A similar analysis was carried out for the PdG results, and overall only 1.5% of the pre-coated assay tubes used at home in all locations and 1.1% of the tubes used in the study centres at all locations produced either one day spikes or troughs. This difference was not statistically significant ($P = 0.107$). There were no significant differences between the home and centre data for Santiago ($P = 0.760$) and

Table I Summary of the number of one day spikes and troughs in the home and centre data removed for the 113 cycles.

Study location home or centre	E1G		PdG	
	No. of assays	No. of spike or trough days removed (%)	No. of assays	No. of spike or trough days removed (%)
Santiago				
Home	1018	53 (5.21)	1018	24 (2.36)
Centre	1020	49 (4.80)	1020	22 (2.16)
Sydney				
Home	929	45 (4.84)	922	9 (0.98)
Centre	932	54 (5.79)	922	8 (0.87)
Palmerston North				
Home	1086	51 (4.70)	1086	13 (1.20)
Centre	1085	18 (1.66)	1085	2 (0.18)
Total				
Home	3033	149 (4.91)	3026	46 (1.52)
Centre	3037	121 (3.98)	3036	32 (1.06)

Table II Summary of RMS differences between home and study centre menstrual cycle profiles.

Mean RMS difference between home and centre ΔT data (n)		
Location	**E1G**	**PdG**
Santiago	0.89 ± 0.21 (39)	0.58 ± 0.20 (39)
Sydney	0.96 ± 0.22 (35)	0.50 ± 0.17 (35)
Palmerston North	0.77 ± 0.22 (39)	0.40 ± 0.13 (39)
All locations	0.89 ± 0.23 (113)	0.51 ± 0.19 (113)

Sydney ($P = 0.807$), but the Palmerston North centre data were significantly better than the corresponding home data ($P = 0.004$). It should be noted that the spike and trough days identified in the home data were necessarily different from the days identified by the centre operators for the same samples, even though the percentages were similar.

The mean RMS values for overlap of the E1G and PdG menstrual cycle profiles resulting from the home-centre comparisons for the three locations individually and combined are given in Table II. The agreement between the home and centre profiles after removal of the one day spikes and troughs was similar for Sydney and Santiago for both E1G ($P = 0.16$) and PdG ($P = 0.07$). However, the Palmerston North home and centre profiles agreed slightly better as shown by significantly lower RMS values for both E1G (versus Santiago, $P = 0.016$; versus Sydney, $P = 0.0003$) and PdG (versus Santiago, $P = 0.0001$; versus Sydney, $P = 0.0056$).

Average E1G and PdG hormone excretion rates for the menstrual cycle profiles

An analysis of the differences between the duplicate pairs of calculated excretion rates for E1G and PdG (S) identified an extra 35 pairs of data points out of a total of 3037 (1.2%) where the difference was greater than 2.8S. Visual inspection of the home and centre excretion rate menstrual cycle profiles showed that the excretion rate of the pair of values that caused the high difference occurred 21 times in the home results and 14 times in the local centre duplicates. The 21 aberrant home values consisted of 15 E1G excretion rates and 6 PdG excretion rates. For the centre data, nine were E1G excretion rates and five were PdG excretion rates. Of the 24 total aberrant E1G excretion rates, 17 were due to high values that were calculated for the E1G peak day, 3 were due to one value being lower than the assay limit in the peri-ovulatory period and 4 were due to a high value occurring in the luteal phase. For the 11 aberrant PdG pairs, 3 had one of the duplicates with a low excretion rate, 7 had one of the pair with a high excretion rate in the mid-luteal phase and 1 was part of a cycle with poor PdG data. The high E1G excretion rates on the estrogen peak day and the high PdG excretion rates in the mid-luteal phase were accepted as valid duplicates. An example average cycle is given in Fig. 3A for the E1G excretion rate data calculated from the Monitor readings for the first cycle from participant 500 4S from the Chilean centre (Fig. 1B), together with the corresponding mean PdG excretion rates calculated for this cycle.

Composite average cycle profiles

Figure 3B shows the composite E1G, LH and PdG profiles drawn from the daily average home and centre excretion rates for each hormone calculated from the 112 cycles averaged and aligned relative to the LH peak day as day zero. (The cycle from a woman who became pregnant during the study was not included in this calculation.) The first E1G rise day, which marks the beginning of the fertile period (Blackwell and Brown, 1992), was Day −5 when calculated from the increase in E1G excretion rate 2 SDs above the mean baseline period (165 ± 8.5 nmol/24 h for Days −12 to −6). The mean E1G peak day (as defined in methods and shown in Fig. 3A) coincided with the mean LH peak day, and the mean peak values were 327 ± 107 nmol/24 h and 39.5 ± 39.5 mIU/ml, respectively. For the PdG data, following an initial period of declining PdG values (Days −16 to −10), the mean baseline excretion rate (Days −9 to −1) was 3.33 ± 0.05 μmol/24 h. The mean first rise in PdG levels above this follicular phase baseline was clearly discernible by eye on Day 0. The PdG threshold value of 7.0 μmol/24 h (Blackwell *et al.*, 1998) was exceeded on Day +3 when the mean value was 10.1 ± 6.0 μmol/24 h, and the mid-luteal phase maximum PdG value was on Day +7 (17.8 ± 7.2 μmol/24 h). The mean fertile window, defined as the number of days from the first rise in E1G excretion rate to the day before the PdG threshold day inclusive, therefore lasted for 8 days beginning on Day −5 and finishing on Day +2.

Mean lengths of cycle phases

Excluding the one pregnancy cycle ($n = 112$), the average length of the cycles for the population of averaged E1G and PdG menstrual cycle profiles was 28.1 ± 3.1 days and the 5th and 95th percentiles for the menstrual cycle length were 24 and 35 days, respectively. The

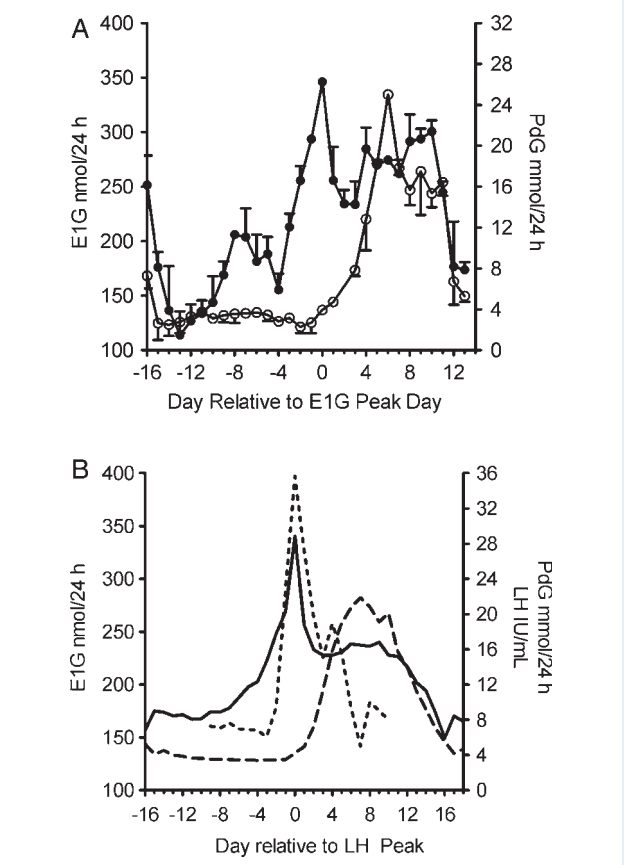

Figure 3 (**A**) Mean of the home and study centre E1G (nmol/24 h) and PdG (μmol/24 h) excretion rates for the first cycle of participant 500 4S from Santiago after conversion with the appropriate standard curves and removal of one day spikes and troughs in the data. Closed circles = E1G data; open circles = PdG data. (**B**) Composite cycle for the 112 mean cycle profiles. The E1G (nmole/24 h; solid line) and PdG (μmol/24 h; broken line) data are the means of the home and centre hormone excretion rates. The dotted line is the composite profile for the LH peak for all of the cycles.

mean day of the LH peak and the length of the follicular phase was 14.8 ± 3.1 days (n = 107; 5th and 95th percentiles: 11 and 21 days, respectively). The mean length of the luteal phase was 13.3 ± 1.5 days (n = 106; 5th and 95th percentiles: 11 and 17 days, respectively), and the mean day of the E1G peak was 14.5 ± 3.2 days (n = 113; 5th and 95th percentiles: 10 and 20, days respectively).

Agreement of key parameters between the home and centre duplicates

The difference between the home and centre results (home value minus centre value) for the pre-ovulatory rise day agreed to within −0.12 ± 0.90 days (n = 113), the E1G peak days agreed to within −0.03 ± 0.85 days (n = 113) and the PdG threshold days agreed to within 0.15 ± 0.99 days (n = 111). The negative sign indicates that the home results were ahead of the centre results. For the pre-ovulatory E1G rise day, 102/113 cycles (90.3%) agreed to within one day and the remaining 11 (9.7%) agreed to within 2 days. For the E1G peak day, 104/113 cycles (92%) agreed within one day while the remaining nine (8%) agreed to within 2 days. The PdG threshold day agreed to within one day for 98/111 cycles (88.3%). Of the remaining 11.7%, 9% (10/111) agreed to within 2 days and the last 2.7% agreed to within 3 days (3/111).

Discussion

The previous study (Blackwell *et al.*, 2003) showed that trained technicians using the pre-coated assay tubes could get clinically useful information on the women's menstrual cycles that was as accurate as that obtained by laboratory procedures. The present study confirmed that the Ovarian Monitor assay with pre-coated tubes worked well in the hands of lay users even without standard curves, quality controls or replicates. These were demanding conditions, but the assay was nevertheless suitable for home use, and valuable clinical information was obtained without significant degradation of the data compared with the trained technicians. Better performance of the tests was routinely obtained in the Palmerston North centre, where there was previous experience with the pre-coated assay tubes, and where repeat assays were carried out according to the study protocol. This suggested that an experienced supervisor is desirable at all stages of the application.

No reference assay was available to confirm the accuracy of the cycle profiles in this study. In a standard laboratory immunoassay, errors (or outliers) always occur but these are recognized by using replicates (usually three or four) for each sample. This strategy is not possible with a home test using singletons. Thus, to assess the performance of the pre-coated assay tubes in the hands of both sets of lay users, the following assumptions were made:

(i) the ovary is not an erratic organ and therefore, one day spikes or troughs in the data are not physiological;
(ii) an aberrant value (a one day spike or a trough) in a home assay result not duplicated in the corresponding centre data for the same urine sample means the home datum is a measurement error, and *vice versa*;
(iii) the relatively smooth menstrual cycle profiles are unique for each cycle and must be the same within experimental error irrespective of the operator or the batch of pre-coated assay tubes and
(iv) the home and centre excretion rates for each location and each analyte should be the same after conversion of the data, thus they should be duplicates and may be averaged.

Application of these principles to the normalized daily data as described here resulted in easy identification of the one day spikes and troughs by overlap of the home and centre profiles. After removal of the aberrant values, a good agreement between the home and centre profiles was obtained, within experimental error, for both E1G and PdG (as shown in Fig. 1B and Fig. 2A and B). The greater RMS values for the E1G profiles reflect the facts that the excretion rates for E1G are approximately 20–100 times smaller than for PdG, and also that the differences between E1G excretion rates in the follicular phase are only ∼40% per day (Brown, 2011), which is close to the stated coefficient of variation of 15% (Blackwell *et al.*, 2003) for the E1G assay.

However, it was noteworthy that the normalized PdG profiles showed better agreement than the E1G profiles for the whole study irrespective of the operator (see Fig. 2 and Table II). This was also reflected in the lower percentage of one day spikes and troughs for

the PdG assays with home use and in all centres (Table I) highlighting the robustness and value of the PdG assay. The greater variability in the shapes of the E1G profiles compared with the PdG profiles is consistent with data on variability of follicular growth patterns and the associated hormone production (Gore et al., 1995; Baerwald et al., 2003a,b; Santoro et al., 2003). As the low percentage of aberrant values overall was similar to that encountered previously (1–5%) with the laboratory-trained users (Blackwell et al., 2003), this implied that the home and centre lay users were successfully performing the assays according to the protocols of the study. It also implies that the one day spikes and troughs originated from faulty tube production and are not due to the operators.

Removal of the one day spikes and troughs from the Monitor readings before conversion to excretion rates did not prevent a few pairs of the calculated home and centre values differing by more than expected on the basis of the experimental variation. Most of these extra discrepancies were the result of one of the Monitor readings being outside the working range of the assays (<50 or >300). The high E1G values mostly occurred during the rising values that made up the ovulatory peak and in fact defined the ovulatory E1G peak day. These values were accepted because even if the actual values for the average excretion rate were uncertain, they were peak values and were definitely high. The same was true for the few high mid-luteal phase PdG values. Even though repeat assays were not done routinely in this study, the one day spikes and troughs in the data could be recognized easily by the users, since consecutive daily measurements were being taken. Thus, in home use any spike or trough can be identified and repeated at the time. Although undesirable, this level of repeats has not proved to be an obstacle to the use of the system (Brown et al., 1991; Cavero, 1995).

The agreement found in the timings of the key cycle parameters (the first E1G rise day, the ovulatory E1G peak day and the PdG threshold day) for the valid home and centre data when analysed separately was remarkable, generally being the same to within one day. This agreement is an important criterion of the accuracy of the results being obtained. Clearly, if the day a given event occurred in both the home and the centre data was within a day, then confidence could be placed on the average results (e.g. the day of the early E1G peak and the E1G peak day in Fig. 3A). For each parameter, this was true for nearly all of the averaged cycles. For the whole study, the average RMS of the differences between the duplicate home and centre E1G excretion rates was 36 $\pm$ 21 nmol/24 h, and for the baseline PdG values (1–5 μmol/24 h) the average difference was 0.77 $\pm$ 0.28 μmol/24 h. These differences correspond approximately to least significant differences of ~42 nmol/24 h and 0.55 μmol/24 h, respectively. Differences in E1G and PdG excretion rates less than these values, although not significant from zero statistically, may nevertheless lead to a different day being chosen for a given cycle event in a few cases as encountered in this study. Most of the timings of key cycle parameters between the home and centre data that differed by greater than $\pm$1 day occurred either where the hormone excretion rates were very low or where the E1G peak day excretion rate differed between the home and centre results by <42 nmol/24 h. In all but one case of the few discrepancies in the three key cycle parameters, no difficulties would have been experienced in the interpretations for either avoidance or achievement of pregnancy. However, averaging of the home and centre cycles as done in this work removed this problem and gave an accurate profile for both the E1G and PdG excretion rates.

The composite profiles derived from the averages of the lay user results were similar to published classical composite menstrual cycle profiles (Moghissi et al., 1972; Brown et al., 1988; Collins, 1989; Munro et al., 1991; Ecochard et al., 2001; Alliende, 2002; Baerwald et al., 2003a,b; Cole et al., 2009). The fact that the LH peak and E1G peak days were the same was unexpected since usually the LH peak lags after the E1G peak by about one day (Ecochard et al., 2001). However, they occurred on the same day in the data published by Alliende (2002). In both the present results and those published by Alliende (2002), multiple LH peaks were observed, which can bias the composite profile in the direction of an earlier mean LH peak day since the extra LH peaks tended to occur before the ovulatory LH peak. The reason for the multiple LH peaks is not yet clear but in another study on regularly cycling women, 47% had early urinary LH surges and 37% of these had multiple early LH peaks (Krotz et al., 2005).

It should be noted that real features of individual profiles, such as the decline in E1G and PdG excretion rates from the previous cycle and the early E1G peak shown in Fig. 3A, are not apparent in the composite profiles (Fig. 3B). This highlights the loss of information that can occur when populations of data are averaged (Alliende, 2002). The mean length of the fertile window of 8 days agreed well with the value of 7.6 days reported previously for a different population of women using the first rise in total estrogen excretion and a threshold of pregnanediol excretion (Blackwell et al., 1998). There was also excellent agreement between the mean follicular and luteal phase lengths calculated from the data using the day of the LH peak as the mid-cycle marker and with the corresponding values reported by Cole et al. (2009). The follicular phase concluded within 11–21 days of the first day of bleeding (Cole: 10–20 days) and the luteal phase lasted 11–17 days (Cole: 9–17 days). By the criteria given by Cole, all of these individual mean cycles would be judged as normal but it should be noted that they are not necessarily fertile.

The results clearly demonstrate that accurate point-of-care testing is possible with a device such as the Ovarian Monitor in the hands of lay users if a small level of assay repeats is acceptable. Thus, the Ovarian Monitor offers another possibility for reliable family planning to achieve or prevent a pregnancy (Freundl et al., 2010). However, if a point-of-care test is to be carried out without the need to run replicates, irrespective of whether they are done by lay or technically trained personnel, a method with a coefficient of variation of better than 15% is needed. A precision similar to the best laboratory assays (8–10%) is the objective, but thus far no point-of-care assay for E1G or PdG with this precision is available.

Further research should focus on evaluating the variability of the menstrual cycle profiles produced in this study, validating the early E1G peaks by daily ultrasound measurements and on comparing the hormonal points of change with changes in self-observed mucus and basal body temperature symptoms.

Acknowledgements

The authors thank the UNDP/WHO/World Bank/Special Programme of Research, Development and Research Training in Human Reproduction (Geneva) for the financial and administrative

support to carry out this study. The authors thank Drs Delwyn Cooke, Simon Brown, Guenter Freundl, Manuel Cortes, Maria Alliende and Paul van Look for reading the manuscript and making helpful suggestions after Prof. Brown's passing. The authors also thank Dr Robert Lambourne who wrote the computer programs. Special thanks are due to Prof. James Brown whose encyclopaedic understanding of the menstrual cycle contributed greatly to this study and its analysis. He was unflagging even as his health deteriorated.

Authors' roles

L.F.B. was a principle investigator (Palmerston North) and drafted the original proposal to WHO. P.V. (Santiago) and B.G. (Sydney) were principal investigators and executed the study protocols in their respective centres. C.D. was the coordinator of the study for WHO. D.G.C. participated in the execution, manuscript drafting and critical discussion. J.B.B. contributed to the conception, design and analysis and interpretation of data, manuscript drafting and critical discussion and provided the assay tubes.

Funding

This project was funded entirely by the UNDP/WHO/World Bank/Special Programme of Research, Development and Research Training in Human Reproduction, Project #90905.

Conflict of interest

There were no conflicts of interest when this study was being carried out. One of the author's (D.G.C.) now works for a small company trying to build alternatives to the Monitor.

References

Alliende ME. Mean versus individual hormonal profiles in the menstrual cycle. *Fertil Steril* 2002;**78**:90–95.

Baerwald AR, Adams GP, Pierson RA. A new model for ovarian follicular development during the human menstrual cycle. *Fertil Steril* 2003a; **80**:116–122.

Baerwald AR, Adams GP, Pierson RA, Baerwald AR, Adams GP, Pierson RA. Characterization of ovarian follicular wave dynamics in women. *Biol Reprod* 2003b;**69**:1023–1031.

Blackwell LF, Brown JB. Application of time-series analysis for the recognition of increases in urinary estrogens as markers for the beginning of the potentially fertile period. *Steroids* 1992;**57**:554–562.

Blackwell LF, Brown JB, Cooke D. Definition of the potentially fertile period from urinary steroid excretion rates. Part II. A threshold value for pregnanediol glucuronide as a marker for the end of the potentially fertile period in the human menstrual cycle. *Steroids* 1998;**63**:5–13.

Blackwell LF, Brown JB, Vigil P, Gross B, Sufi S, d'Arcangues C. Hormonal monitoring of ovarian activity using the Ovarian Monitor, part I. Validation of home and laboratory results obtained during ovulatory cycles by comparison with radioimmunoassay. *Steroids* 2003; **68**:465–476.

Brown JB. Types of ovarian activity in women and their significance: the Continuum (a reinterpretation of early findings). *Hum Reprod Update* 2011;**17**:141–158.

Brown JB, Blackwell LF, Cox RI, Holmes JM, Smith MA. Chemical and homogeneous enzyme immunoassay methods for the measurement of estrogens and pregnanediol and their glucuronides in urine. *Prog Clin Biol Res* 1988;**285**:119–138.

Brown JB, Blackwell LF, Holmes J, Smyth K. New assays for identifying the fertile period. *Int J Gynecol Obstet—Suppl* 1989;**1**:111–122.

Brown JB, Holmes J, Barker G. Use of the Home Ovarian Monitor in pregnancy avoidance. *Am J Obstet Gynecol* 1991;**165**: 2008–2011.

Cavero C. Using an ovarian monitor as an adjunct to natural family planning. *J Nurse Midwifery* 1995;**40**:269–276.

Cole LA, Ladner DG, Byrn FW. The normal variabilities of the menstrual cycle. *Fertil Steril* 2009;**91**:522–527.

Collins WP. Biochemical indices of potential fertility. *Suppl Int J Gynecol Obstet* 1989;**1**:35–43.

Cooke DG, Binnie JE, Blackwell LF. Validation of a reference ELISA for estrone glucuronide using urine samples normalized by dilution to a constant rate of urine production. *Steroids* 2007;**72**: 580–591.

Ecochard R, Boehringer H, Rabilloud M, Marret H. Chronological aspects of ultrasonic, hormonal, and other indirect indices of ovulation. *Int J Obstet Gynaecol* 2001;**108**:822–829.

Freundl G, Sivin I, Batar I. State-of-the-art of non-hormonal methods of contraception: IV. Natural family planning. *Eur J Contracept Reprod Health Care* 2010;**15**:113–123.

Gore MA, Nayudu PL, Vlaisavljevic V, Thomas N. Prediction of ovarian cycle outcome by follicular characteristics, stage 1. *Hum Reprod* 1995; **10**:2313–2319.

Krieg EF Jr, Kesner JS, Knecht EA. An algorithm for detecting features of the hormone profiles of the human menstrual cycle. *Comput Biol Med* 1999;**29**:229–242.

Krotz S, McKenzie LJ, Cisneros P, Buster J, Amato P, Carson S, Krotz S, McKenzie LJ, Cisneros P, Buster J *et al.* Prevalence of premature urinary luteinizing hormone surges in women with regular menstrual cycles and its effect on implantation of frozen-thawed embryos. *Fertil Steril* 2005;**83**:1742–1744.

Moghissi KS, Syner FN, Evans TN. A composite picture of the menstrual cycle. *Am J Obstet Gynecol* 1972;**114**:405–418.

Munro CJ, Stabenfeldt GH, Cragun JR, Addiego LA, Overstreet JW, Lasley BL. Relationship of serum estradiol and progesterone concentrations to the excretion profiles of their major urinary metabolites as measured by enzyme immunoassay and radioimmunoassay. *Clin Chem* 1991;**37**:838–844.

Santoro N, Isaac B, Neal-Perry G, Adel T, Weingart L, Nussbaum A, Thakur S, Jinnai H, Khosla N, Barad D *et al.* Impaired folliculogenesis and ovulation in older reproductive aged women. *J Clin Endocrinol Metab* 2003;**88**:5502–5509.

Snedecor GW, Cochran WG. *Statistical Methods*. Iowa, USA: Wiley-Blackwell, 1989.