

A DATA BASE FOR BIOMETRIC RESEARCH ON CHANGES IN BASAL BODY TEMPERATURE IN THE MENSTRUAL CYCLE

L. Miolo, B. Colombo, J. Marshall

1. INTRODUCTION

In the Department of Statistical Sciences of the University of Padua a data base has been created which has unique features for biometric research on the menstrual cycle. In the following are described: 1) the kind of data available, their origin and the collection procedure; 2) the choices made for the structure of the base and its contents; 3) the size of the sample and of some of its subsets, according to various characteristics.

The purpose of this note is to give a unique term of reference for several studies ongoing or planned which take advantage of this documentation. And it is felt that it could help in promoting interest and in offering possibilities for further hypothesis testing in several directions.

2. THE ORIGIN OF THE DATA AND THEIR COLLECTION

The Catholic Marriage Advisory Council of England and Wales provides a counselling and educational service, including instruction in natural family planning (NFP), free of charge to those who wish to make use of its service irrespective of race, nationality or religious affiliation. The NFP Service began in 1950 using initially the calendar method. This was soon superseded by the basal body temperature (BBT) method which used the temperature shift associated with ovulation to determine the fertile and infertile phases of the menstrual cycle. To this was later added observation of changes which occur in cervical mucus in relation to ovulation.

The NPF service was first offered in face to face meetings between client(s) and teacher. When it became apparent that many seeking the service were prevented by considerations of time or travel from reaching a NPF teacher, instruction by correspondence was instituted as a pilot scheme. Clients were

provided with a simple booklet describing the principles of NFP and how to apply it in practice, together with specially designed charts and an ovulation thermometer. When they had recorded the BBT through one menstrual cycle, they sent in the chart which was returned to them annotated to show the rise of temperature associated with ovulation and the fertile and infertile phases of the cycle. This was accompanied by a letter of explanation and instruction to annotate the next chart themselves before sending it in for comment. In this way the ability of clients to record and interpret charts correctly was developed and checked. The scheme proved very successful. In a later field trial of BBT as a method of avoiding conception in which clients taught by face to face interview and those instructed by correspondence took part, there was no significant difference between the results from the groups (1).

Photocopies of all charts submitted were kept in the NFP office in order to facilitate answering questions from clients about changes in their charts by reference to all their past charts. Although this practice began as a service need, it was soon realised that a data base which could be of great value for research was being built up. Some clients, once they were competent and confident about interpreting charts (which usually was achieved in six to twelve cycles), ceased to send in their charts, whilst others continued to send them every three cycles. It is this data base which provides the basis of this report.

The clients came largely from England and Wales with a significant minority from the Republic of Ireland and Scotland and a few from other countries. Most of them were already married but a few began to record and submit charts during the months prior to marriage. Most were known to be fertile having already given birth to one child. Some had not yet conceived but had no reason to doubt their fertility (more than 20%). Although BBT recording and observations of cervical mucus can be helpful in the management of infertility, no such records are included in this series. The series consists therefore of healthy, fertile women using NFP to regulate births.

No systematic enquiry was made as to why clients chose NFP in preference to other methods of avoiding conception but the correspondence indicates the majority chose NFP for ethical reasons and a significant minority because of preference for 'natural' things in various aspects of their lives.

The fact that the series was collected by correspondence indicates that the clients were all literate, though the standards of literacy varied enormously. Information was collected systematically about age and parity; as this was a service rather than a research project, other data, which would have been of interest, were not collected in a systematic manner. However, a number of correspondents took part in a number of research projects; this provided some information about the clients contributing to this data base. In a subset of 502 couples 21 per cent had been married less than five years and 45 per cent for 5 to 10 years. 46 per cent were in the Registrar-General of England and Wales socio-economic category III (the categories range from I to V).

Definition of a significant shift of temperature indicating ovulation was as follows: the first time in the cycle that three consecutive daily temperatures were

recorded all of which were above the level of the immediately preceding six daily temperature recordings. This was known as the 'three over six rule'. One exception to the rule was permitted namely the presence of one 'spike' temperature among the six at the lower level; a spike temperature was defined as a tempera-

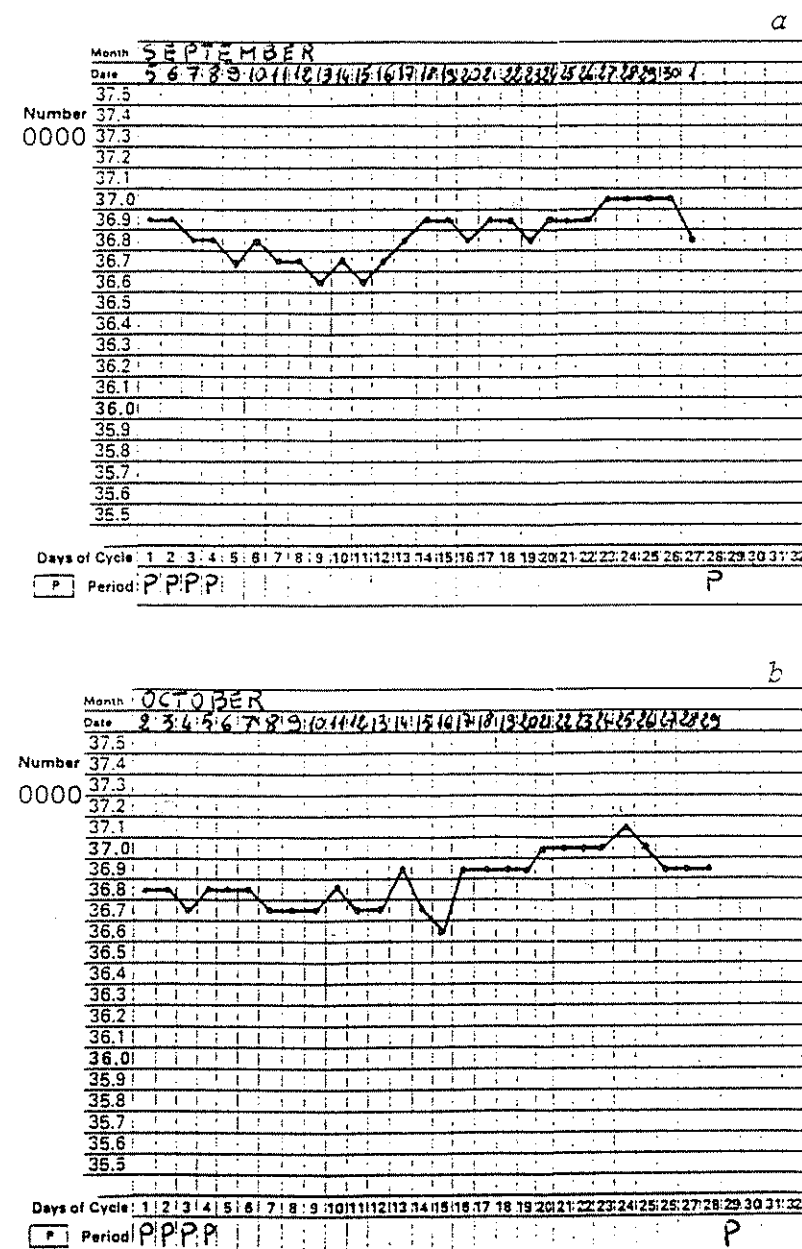


Figure 1 - BBT Charts a) for a normal case and b) for a case with a spike.

ture which was 0.2 degrees Centigrade or more above both its immediate neighbouring temperatures. The decision to permit this exception was made on empirical grounds. Charts had been observed showing a clear shift of temperature at the appropriate time and no evidence of disturbance apart from a 'spike' temperature. Couples ignoring this spike had not become pregnant hence this practice was confirmed as a permissible exception. The cause of the spike is likely to have been consumption of alcohol, a late night or other social event (2).

In figures 1a and 1b is presented a sketch of a chart for a normal case and of one for a case with a spike, respectively. From it, one can see that from each chart it was possible to know the dates of observation and the length in days of the menses and of the cycle. Further space allowed the client to give information on possible disturbances, but these were not recorded in the data base for electronic data processing (EDP). On each chart was also registered a four figures code number assigned to a woman. This code remained always the same for that client throughout all the period for which she remained in contact with the centre.

3. THE CREATION OF THE DATA BASE

3.1. Preliminary choices made

Photocopies of all the charts collected during many years were sent to the Department of Statistical Sciences of the University of Padua.

Here it was decided, for the purpose of building a base for EDP, to make a selection among the available paper documentation. The code number of the woman and the registered dates of the beginning and of the end of each cycle made it possible to order sequences of consecutive cycles for each woman. Among these, it was decided to consider only sequences of at least six cycles. This was done for two reasons:

- 1) to discard cases of occasional contacts and keep the records of clients who showed a proved interest in the use of the method;
- 2) to obtain always a minimum set of observations allowing longitudinal analyses.

Before the input, the sequentially ordered charts were examined one by one in order to determine on them the shift in the BBT. This was done visually following strictly the 'three over six' rule described above. In a small percentage of cases, when the recorded temperatures were not sufficient for that strict application, it was decided to accept the "clinical" evaluation of the beginning of the shift when this had been clearly indicated at the centre. That judgement could in fact be supported by a consideration of the whole available documentation on the preceding cycles for the same woman. This clinical determination was not accepted in cases in which the required information for the application of the 'three over six' rule was recorded and was in contradiction with the 'clinical' determination. This choice was made having in mind some specific

uses of the data base, while with the 'clinical' determination the purpose was to try to take the maximum advantage of the information in the light of circumstances of the individual couple. Obviously, it was not possible to determine the shift when the registration of the temperature was missing in days critical for a correct evaluation. In a certain number of cases also the cycle was clearly monophasic and sometimes it was explicitly defined on the chart as anovulatory. Instances of occasional illnesses disturbing the temperature pattern were observed and recorded as were illegible charts.

It was also possible in number of cases, when a single chart was missing, to determine the length of the corresponding cycle. This was possible from the information about the last date of the preceding cycle and the first of the following one. A special situation arose in a few cases with observed very short pre-shift durations. It was decided to accept as valid those with a pre-shift length of less than nine days only when the woman had usually short cycles (the two preceding had to be at most 27 days long).

At the headquarters of the Catholic Marriage Council a separate dossier was kept for each woman. Consulting this documentation made it possible to complement the information obtained from the charts. For more than half of the women it was possible to collect data on her reproductive history (in terms of pregnancies) and for 60% the age (at least approximately as the difference of calendar years of contact with the centre and of birth).

In table 1 of part three numerical documentation of what has been said is presented.

3.2. The structure of the base

Having in mind the utilization of the data with a variety of EDP instruments, it was decided to structure the data in a standard form, in fixed fields. The informations of interest, obtained from the paper base, and which were inputted, can be arranged in three sets. For the *woman*: code number, year of birth, date of marriage, number of pregnancies at first contact with centre, date of last previous reproductive event, date of all reproductive events while under observation, total number of recorded segments. For each *segment*: progressive number of recorded segments, total number of cycles. For the *cycle*: progressive number within a segment, total length (in days), length of the pre-shift phase (from the first day of the period through the last low temperature before the shift determined by application of the three over six rule), number of days of the period, type of measurement of the temperature (in Fahrenheit or Celsius), the recorded (or unrecorded) temperature for each day, a qualification of some special characteristics (with illness, monophasic, ...).

The basic unit - the record chosen - for the registration was the cycle.

The input through a data entry system allowed to insert a series of pre-checks for consistency in order to reduce the rate of errors in this phase. Also at the end several checks were made on the entire set of records. A few remaining mistakes were so eliminated.

4. THE SAMPLE, ITS SUBSET AND CHARACTERISTICS

The registered information comes from a sample of 1,798 women, each one of them – as said – represented by at least one sequence of six cycles. The total number of segments amounts to 2,397. For a woman, a maximum of eight distinct segments is recorded. The data base contains the lengths of 36,641 cycles, including 662 cases in which charts are missing.

Table 1 classifies the cycles according to some qualifications. In the 31,925 cases in which it proved applicable, the three over six rule did not allow in 6.6% of the total to determine the shift of the BBT.

TABLE 1
Frequency distribution of recorded cycles

| Cycles | Frequency | Percent |
|------------------------------------|-----------|---------|
| Charts missing | 662 | 1.8 |
| Charts illegible | 113 | 0.3 |
| With illness | 481 | 1.3 |
| Monophasic | 243 | 0.7 |
| With missing critical temperatures | 3,217 | 8.8 |
| With undetermined shift | 2,093 | 5.7 |
| With determined shift | 29,832 | 81.4 |
| Total | 36,641 | 100.0 |

In table 2 two frequency distributions are illustrated.

The first one (a) gives the number of cycles per woman. In almost 140 instances more than 50 cycles have been registered; in 19 more than one hundred, in one case 149.

The second (b), gives the number of cycles for each one of the quoted 2,397 distinct segments. The limits of the classes are here arranged so as to allow deriving approximately in years the length of the period during which a single case has been continuously under observation. One client registered regularly an uninterrupted sequence of 109 cycles. Section b) of the table shows the richness of the base for longitudinal analyses. Taking into account cases in which more than one segment has been collected, there is evidence of clients remaining in steady contact with the Council for more than ten years.

The wide range of the lengths of cycles compelled us, for better use of memory, to create distinct files for very long cycles. Table 3 gives a distribution constructed in a way to allow a detailed knowledge of lengths of maximum frequency and to offer the possibility of comparisons and evaluations with corresponding choices found in the relevant literature.

TABLE 2
Frequency distribution of the number of cycles a) per woman and b) per segment

| a) per woman | | | b) per segment | | | |
|---------------|-----------|------|----------------|-----------|------|----------------------|
| No. of cycles | Frequency | % | No. of cycles | Frequency | % | Retrocumulated freq. |
| 6 | 204 | 11.3 | 6 | 338 | 14.1 | 2397 |
| 7-10 | 540 | 30.0 | 7-13 | 1,189 | 49.6 | 2059 |
| 11-20 | 482 | 26.8 | 14-26 | 562 | 23.4 | 870 |
| 21-30 | 231 | 12.9 | 27-39 | 165 | 6.9 | 308 |
| 31-50 | 204 | 11.3 | 40-52 | 84 | 3.5 | 143 |
| 51-100 | 118 | 6.6 | 53-65 | 31 | 1.3 | 59 |
| ≥ 101 | 19 | 1.1 | ≥ 66 | 28 | 1.2 | 28 |
| Total | 1,798 | | Total | 2,397 | | |

TABLE 3
Frequency distribution of the number of cycles according to their length (in days)

| Length of cycle | Frequency | % | Cumulated frequency | % |
|-----------------|-----------|------|---------------------|-------|
| < 20 | 337 | 0.9 | 337 | 0.9 |
| 20-22 | 733 | 2.0 | 1,070 | 2.9 |
| 23-24 | 2,990 | 8.2 | 4,060 | 11.1 |
| 25 | 3,439 | 9.4 | 7,499 | 20.5 |
| 26 | 4,795 | 13.1 | 12,294 | 33.6 |
| 27 | 5,276 | 14.4 | 17,570 | 48.0 |
| 28 | 5,136 | 14.0 | 22,706 | 62.0 |
| 29 | 3,793 | 10.3 | 26,499 | 72.3 |
| 30 | 2,915 | 8.0 | 29,414 | 80.3 |
| 31 | 1,985 | 5.4 | 31,399 | 85.7 |
| 32-33 | 2,407 | 6.6 | 33,806 | 92.3 |
| 34-35 | 1,179 | 3.2 | 34,985 | 95.5 |
| 36-45 | 1,332 | 3.6 | 36,317 | 99.1 |
| ≥ 46 | 324 | 0.9 | 36,641 | 100.0 |
| Total | 36,641 | | | |

In 17 cases the cycle lasts for at least 100 days, up to a maximum of 286. These are exceptional cases about which similar evidence can be found elsewhere in published papers. According to what one considers, or defines, as "normal", from the table it is possible to find the derived fraction of anomalous instances.

The modal length of 27 days in table 3 underlines what can be defined as an observational bias in this kind of data collection. When women of different average cycle lengths are observed during by and large the same period, more cycles will come from women with shorter vs. women with longer cycles. This

fact has to be taken into account in some kinds of data analysis in order to avoid deriving biased estimates of parameters for the population as a whole.

As already stated, in a large number of cases – a total of 1074 women – it proved feasible to obtain the age of the woman. This was recorded in various ways: date of birth, age at first contact with the centre, year of birth. When there are for one woman other sequences of cycles after the first one, given the date of the new contact, there is a way to determine the age at the beginning of each segment. For purposes of standardization and ease of manipulation the base has been structured in a way that always permits us to derive the age at least approximately as a difference between calendar years.

The distribution of table 4 shows that, except for very young ages, there is ample documentation for any portion of the reproductive period.

TABLE 4

Frequency distribution of women by age at the first contact with the centre

| Age | Frequency | % |
|-------------|-----------|------|
| < 20 | 5 | 0.5 |
| 20-24 | 210 | 19.5 |
| 25-29 | 288 | 26.8 |
| 30-34 | 278 | 25.9 |
| 35-39 | 172 | 16.0 |
| 40-44 | 92 | 8.6 |
| 45 and more | 29 | 2.7 |

While in table 4 only the age at the first contact with the centre is given, obviously the data base allows to derive the same variable at the beginning – or at any moment – of each observed segment.

For a smaller number of clients, 977, the previous reproductive history, together with subsequent events while under observation, were also available. For almost all – in 898 instances – both variables have been recorded. Table 5 gives a cross-classification of the results.

TABLE 5

Cross-classification of women by age at first contact with the centre and number of pregnancies

| No. of pregnancies | Missing | < 20 | 20-24 | 25-29 | 30-34 | 35-39 | 40 and more | Total |
|--------------------|---------|------|-------|-------|-------|-------|-------------|-------|
| 0 | 14 | 4 | 102 | 67 | 22 | 8 | 8 | 225 |
| 1 | 24 | 1 | 34 | 59 | 35 | 23 | 5 | 181 |
| 2 | 19 | 0 | 30 | 58 | 53 | 32 | 15 | 207 |
| 3 | 11 | 0 | 7 | 31 | 55 | 34 | 30 | 168 |
| 4 | 7 | 0 | 2 | 19 | 39 | 25 | 15 | 107 |
| 5 and more | 4 | 0 | 0 | 8 | 30 | 27 | 20 | 89 |
| Total | 79 | 5 | 175 | 242 | 234 | 149 | 93 | 977 |

In evaluating the frequencies in the various cells of table 5, one has to remember that each case corresponds to a sequence of at least six cycles. The wealth of the data base for differential analyses according to the specific typologies of the subjects is evident.

Finally, another piece of information is worthwhile mentioning. In 14,231 cycles the temperature is registered every day, from the first to the last day of the cycle. In 14,520 only the first two temperatures are missing and in 14,564 only the last two. In 15,086 the temperature is regularly recorded except for the first and the last two days. In thousand of cases it is possible to have couples, and triplets, and more of cycles in which all the daily temperatures are present.

From this, and all the foregoing description of the data base, one can construct plans for a variety of research purposes.

*Dipartimento di Scienze statistiche
Università di Padova
Catholic Marriage Advisory Council,
London*

LAURA MIOLO
BERNARDO COLOMBO
JOHN MARSHALL

ACKNOWLEDGMENTS

The Authors are grateful to several undergraduate and graduate students who in various phases and ways gave their help in the construction and checking of the data base, and particularly to dr. Paola Melina for her generous, invaluable contribution.

REFERENCES

- J. MARSHALL (1968), *A field trial of the basal-body-temperature method of regulating births*, "Lancet", 2, pp. 8-10.
J. MARSHALL (1979), *Planning for a family. An atlas of mucothermic charts*, Faber and Faber, London, 2nd ed.

RIASSUNTO

Una base dati per l'analisi biometrica delle temperature basali nel ciclo mestruale

Nel Dipartimento di Scienze statistiche dell'Università di Padova, su documentazione messa a disposizione dal Catholic Marriage Advisory Council di Londra, è stata messa a punto una base di dati per elaborazione elettronica. L'articolo descrive innanzi tutto l'origine del materiale, raccolto in sede di consulenza sul ricorso a regolazione naturale della fertilità. Prosegue con chiarimenti sulle scelte fatte nella selezione dei dati da conservare, fra tutti quelli pervenuti, e nella strutturazione della base. Infine espone in breve sintesi dimensioni e caratteristiche del campione e di alcuni suoi sottoinsiemi. Lo scopo è di fornire una breve ma esauriente informazione su una base di dati che, per ampiezza associata ad alcune variabili, è unica al mondo. Ad essa potranno fare riferimento

pubblicazioni previste su ricerche già svolte, o in corso, o programmate per la verifica di una serie di ipotesi scientifiche.

SUMMARY

A data base for biometric research on changes in basal body temperature in the menstrual cycle

In the Department of Statistical Sciences of the University of Padua a data base for EDP has been created on data made available from the Catholic Marriage Advisory Council of London. This paper describes, first of all, the origin of the material, collected through consulting work on natural family planning. Clarifications are then provided about the selection of data to be preserved among those received and on the structure of the base. Finally, size and characteristics of the sample and of some of its subsets are presented. The purpose of the paper is to give brief but complete information on a data base which for size and some associated features is unique. To it reference will be made in a series of papers illustrating conclusions reached in various exercises of already performed, or ongoing, or planned hypotheses testing.