The effectiveness of a fertility awareness based method to avoid pregnancy in relation to a couple’s sexual behaviour during the fertile time: a prospective longitudinal study

P.Frank-Herrmann1,8†, J.Heil2,†, C.Gnoth3, E.Toledo4, S.Baur5, C.Pyper6, E.Jenetzky7, T.Strowitzki1 and G.Freundl5

1Department of Gynaecological Endocrinology and Fertility Disorders, University of Heidelberg, Voßstrasse 9, 69115 Heidelberg, Germany, 2Department of Obstetrics and Gynaecology, University of Heidelberg, Voßstrasse 9, 69115 Heidelberg, Germany, 3Department of Obstetrics and Gynaecology, University of Cologne and Centre for Family Planning, Endocrinology and Reproductive Medicine, Rheyder Straße 143, 41515 Grevenbroich, Germany, 4Department of Preventive Medicine and Quality Management, Hospital Virgen del Camino, C/Irunlarrea 4, E-31008 Pamplona, Spain, 5Section Natural Fertility, German Society for Gynaecological Endocrinology, Voßstrasse 9, 69115 Heidelberg, Germany, 6National Perinatal Epidemiology Unit, Department of Public Health, University of Oxford, Old Road Campus, Oxford, OX3 7LF, Great Britain and 7Institute of Medical Biometry and Informatics, University of Heidelberg, Im Neuenheimer Feld 305, 69120 Heidelberg, Germany

8To whom correspondence should be addressed at: Department of Gynaecological Endocrinology and Fertility disorders, University of Heidelberg, Voßstrasse 9, 69115 Heidelberg, Germany. E-mail: petra.frank-herrmann@med.uni-heidelberg.de

BACKGROUND: The efficacy of fertility awareness based (FAB) methods of family planning is critically reviewed. The objective was to investigate the efficacy and the acceptability of the symptothermal method (STM), an FAB method that uses two indicators of fertility, temperature and cervical secretions observation. This paper will recommend a more suitable approach to measure the efficacy. METHODS: Since 1985, an ongoing prospective observational longitudinal cohort study has been conducted in Germany. Women are asked to submit their menstrual cycle charts that record daily basal body temperature, cervical secretion observations and sexual behaviour. A cohort of 900 women contributed 17 638 cycles that met the inclusion criteria for the effectiveness study. The overall rates of unintended pregnancies and dropout rates have been estimated with survival curves according to the Kaplan–Meier method. In order to estimate the true method effectiveness, the pregnancy rates have been calculated in relation to sexual behaviour using the ‘perfect/imperfect-use’ model of Trussell and Grummer-Strawn. RESULTS: After 13 cycles, 1.8 per 100 women of the cohort experienced an unintended pregnancy; 9.2 per 100 women dropped out because of dissatisfaction with the method; the pregnancy rate was 0.6 per 100 women and per 13 cycles when there was no unprotected intercourse in the fertile time. CONCLUSIONS: The STM is a highly effective family planning method, provided the appropriate guidelines are consistently adhered to.

Key words: contraception/efficacy/fertility awareness based methods/natural family planning/symptothermal method

Introduction

Background

Fertility awareness based (FAB) methods is a term that includes all family planning methods that are based on the identification of the fertile time. They are based on the woman’s observation of physiological signs of the fertile and infertile phases of the menstrual cycle. This knowledge can be used to plan or avoid pregnancy. FAB methods depend on two key variables: first the accurate identification of the fertile days of a woman’s menstrual cycle (the fertile time) and second the modification of sexual behaviour either to plan a pregnancy or to use this knowledge to avoid pregnancy. When couples use FAB methods of family planning to avoid pregnancy, they practise different sexual behaviour during the fertile time. When FAB methods involve sexual abstinence during the fertile time, this method is called natural family planning (NFP). When FAB methods involve naturally using a barrier method during the fertile time, the method is called FAB method with barriers. It must also be recognized that although many couples state they are practising a FAB method, sometimes they do not adhere to the guidelines and unprotected intercourse or other kinds of genital contact

†The first two authors contributed equally to this work.
occur during the fertile time. The efficacy of FAB methods to avoid pregnancy has been critically reviewed by several authors (Fehring et al., 1994; Frank-Herrmann et al., 1991; De Leizaola-Cordonnier, 1995; Barbato and Bertolotti, 1988; Hilgers and Stanford, 1998; Howard and Stanford, 1999; Kambic, 1999; The European Natural Family Planning Study Groups, 1999; Grimes et al., 2004). Several issues have been identified when attempting to compare the different FAB methods.

The first is that most FAB methods have evolved concurrently over the last 40 years in different countries; each has been lead by pioneers who have developed guidelines for their respective groups. This has resulted in many cases in a lack of evidence-based guidelines being developed and subsequently modified to conform to best scientific evidence.

The second issue is that efficacy rates may vary because they are derived from studies done with volunteers and researchers from different cultural backgrounds where motivation to avoid pregnancy and rigour of research methods may vary (World Health Organization, 1981a,b; Gomes and Congdon, 1988; Xu et al., 1994; Indian Council of Medical Research Task Force on Natural Family Planning 1996).

Third, the methods currently used to calculate the method effectiveness are questionable. Many investigators have recognized the importance of distinguishing between pregnancies attributable to user failure and method failure (method effectiveness). The standard procedure up to now was to compute separate method and use-effectiveness rates (pregnancies divided by exposure). In this procedure, all exposure from perfect and imperfect use is included in the denominator of both method and user failure rates. The common misinterpretation is that the resulting method effectiveness rate yields information about the inherent efficacy of the method. Inherent method efficacy can be measured only when the numerator (method failures) is assessed in relation to the proper risk set, i.e. the exposure only when the method is used perfectly. For this reason, method effectiveness rates computed by the standard procedure are biased downwards to an unknown extent (Trussell and Grummer-Strawn, 1990).

This problem is further confounded by the different ways an unintended pregnancy is classified. Some prospective studies ensure the couple’s intention to avoid a pregnancy is recorded at the beginning of each menstrual cycle. Other studies are retrospective and only question the couple’s intention after sexual intercourse has been recorded during the fertile time.

The fourth issue is that some new FAB methods are simplified methods that are often used in developing countries and very relevant for settings where cost of teaching is an issue and where continuation has a higher priority than efficacy (Thapa et al., 1990; Jennings and Sinai, 2001; Arevalo et al., 2004).

To be able to make an informed choice when selecting a family planning method, couples need to know the efficacy of a method when used consistently and imperfectly. Trussell and Grummer-Strawn are critical about how efficacy has been calculated in previous studies. They argue that previously published rates of method and user failure for all contraceptive methods suffer from methodological errors and are biased downwards. Trussell and Grummer-Strawn (1990, 1991) recommend a new model of calculating perfect and imperfect use pregnancy rates which up to now has rarely been applied. This is likely to be due to the fact that this approach requires documentation of all sexual behaviour during each cycle.

**Description of the method**

This paper describes a cohort of couples who used a method that consisted of recording the cervical secretion pattern, changes of basal body temperature and the application of a calculation rule. It is called the symptothermal method (STM) of NFP. The beginning and the end of the fertile time are identified by two parameters in order to have a double-check system.

The following two guidelines are given to each couple to identify the first fertile day—both guidelines are applied and the first fertile day is the earliest day identified (Figure 1):

(i) Change of cervical secretion: first appearance of cervical secretion.

(ii) Calculation guideline: the first fertile day is the sixth day of the cycle (In NFP methodology, this is called the ‘five days rule’ which states that the first five...
days of the cycle are infertile days.) or after a woman has completed 12 cycles of use this guideline is replaced by a calculation that takes the earliest temperature rise in the last 12 cycles and subtracts seven days to identify the first fertile day (In NFP methodology, this is called the ‘minus eight rule’: earliest temperature rise in the last 12 cycles minus eight days to identify the last infertile day.)

The following two guidelines are given to each couple to identify the last fertile day—both guidelines are applied and the last fertile day is the latest day identified (Figure 1):

(i) The evening of the third day after the cervical secretion peak day (The cervical secretion peak day is only recognized on the day following peak, when the secretions have become sticky again.)

(ii) The evening of third higher temperature reading, all three higher than the previous six readings, the last one 0.2°C higher than the previous six.

These evidence-based guidelines have been developed following extensive research that has been carried out over the last 20 years. They have been adopted widely by FAB groups who teach the STM method (Raith et al., 1999). Detailed guidelines of the STM methodology are described elsewhere (Arbeitsgruppe NFP, 2006).

**Objectives**

Our first objective was to analyse the overall use effectiveness of the STM method and to determine whether the effectiveness was different for different types of sexual behaviour during the fertile time according to the ‘perfect/imperfect-use model’ (Trussell and Grummer-Strawn, 1990). In particular, we specifically wanted to know if those women who only used the STM without combining with a barrier method during the fertile time had fewer unintended pregnancies, regarding the use-effectiveness rates, than those women who occasionally used barrier methods during the fertile time.

Our second objective was to investigate the acceptability of the STM method. We therefore reviewed those couples whose reasons for discontinuing the STM were due to dissatisfaction with the method or due to difficulties with using the STM.

**Materials and methods**

The study was conducted by the German NFP study centre in accordance with the ethical principles of the Declaration of Helsinki. The study centre is an integral part of the German Society of Gynaecological Endocrinology and Reproductive Medicine. Its main aim is to undertake research studies in the field of NFP and to evaluate NFP services. For the last 20 years, the study centre has coordinated this prospective, observational longitudinal cohort study. Between 1985 and 2005, the study enrolled 1599 women using the STM in different situations and collected data from 35 996 menstrual cycles. This data set has already addressed several questions (Gnoth et al., 1999, 2002, 2003). From this data set, a cohort of 900 women with 17 638 cycles met the effectiveness study selection criteria.

**Exclusion criteria**

All couples with a reason of potential sub- and infertility, or not being exposed to risk of conception or trying for pregnancy with the help of NFP were excluded. The reasons for excluding the women and the number of women excluded are given in Table I.

**Inclusion criteria**

The inclusion criteria are given in Table II in accordance with the recommendations from Tietze and Lewit (1974), Trussell and Kost (1987) and Potter (1996). A subcohort of 900 women with 17 638 cycles was selected out of the whole database.

The women had to state that they intended to avoid pregnancy and if they changed their intention they had to be willing to let the investigator know at once in order to have a clear definition of unintended pregnancy (discussed later). The women had to be willing to record all sexual behaviour, especially the occasional use of barrier methods to ensure clarity about the definitions for perfect/imperfect-use (discussed later). The couples must not use any contraceptive methods other than barrier methods. The study specifically only included those couples who were starting to use the STM method and commencing the first STM cycle. To ensure the participants were likely to have normal fertility, the women had to be between 19 years at entry into the study and less than 46 years at the end of the study. To ensure the women could become pregnant (i.e. were likely to have fertile ovulations), the average cycle length of the women had to be between 22 and 35 days (20% of the cycles of each study participant was allowed to deviate outside this range). Only those women with no previous history of infertility were included. There was no requirement for proven fertility in terms of the women already having a history of being pregnant, in order not to exclude the younger, potentially more fertile and sexually active women. In a previous data analysis, we found no significant difference in unintended pregnancy rate between those with and without proven

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants in the whole database</td>
<td>1599</td>
<td></td>
</tr>
<tr>
<td>Participants in the effectiveness study</td>
<td>900</td>
<td></td>
</tr>
<tr>
<td>Excluded</td>
<td>699</td>
<td>100</td>
</tr>
</tbody>
</table>

Table II. Inclusion criteria for the effectiveness study

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 19–45 years</td>
</tr>
<tr>
<td>Normal cycle lengths between 22 and 35 days (20% of cycle lengths could be outside this range)</td>
</tr>
<tr>
<td>Willing to record family planning intention at the start of each cycle</td>
</tr>
<tr>
<td>Willing to record sexual behaviour, including sexual intercourse, genital contact, withdrawal, occasional barrier use</td>
</tr>
<tr>
<td>Agreement not to use any other forms of contraception</td>
</tr>
<tr>
<td>No known history of subfertility or infertility</td>
</tr>
<tr>
<td>An established luteal phase of at least 10 days hyperthermic phase and at least 3 months following breastfeeding, oral contraception, post-partum, post-abortum</td>
</tr>
<tr>
<td>Willing to participate in the study for 12 months</td>
</tr>
</tbody>
</table>
fertility (Frank-Herrmann et al., 1991). Women who had either
delivered a child or breastfed or used oral contraceptives were only
included after 3 months of an established luteal phase, diagnosed by
an elevated temperature phase for at least 10 days. All the women
were asked to agree to participate for at least 12 cycles.

Teaching the STM
All the women who participated in the study were taught the STM by
accredited teachers from the ‘Arbeitsgruppe NFP’ which was founded
in 1981 with the aim of promoting NFP in Germany. In collaboration
with the German NFP study centre, the training and the teaching meth-
ology was standardized and adhered to strict guidelines. There were
comprehensive teaching materials that accompanied the personal
small group teaching sessions (Arbeitsgruppe NFP, 2006).

Recruitment of study participants
The participants were volunteers who had self-selected to join the
study following given standardized information about the study by
their STM teachers; all women gave their informed consent. Standard-
dized admission questionnaires were used to collect relevant data con-
cerning age, parity, family planning history and socio-demographic
background. All women were asked to send their cycle charts, after
each cycle was completed, that recorded basal body temperature,
quality of cervical secretions, cycle length, family planning intention
and sexual behaviour directly to the study centre. Those women,
whose cycle charts did not reach the study centre, were contacted on
three separate occasions by the study centre via the woman’s personal
teacher. If a woman did not respond to any of the requests for infor-
mation she was then classified as lost to follow-up.

Discontinuation
We were specifically investigating the acceptability of the STM. Therefore,
the most important reasons given for discontinuation were those due to dissatisfaction with the STM, and/or change to
other family planning method.

Table III lists all the reasons for discontinuation.

Data collection
The software used to handle the data was a Microsoft Access® rela-
tional database system called NFP DAT 1.0; it is described in a pre-
vious paper (Gnoth et al., 1999). It has an automated analysis system that follows-up the participant every 3 months.

Definition of the pregnancy
The definition of pregnancy was an elevated temperature of longer
than 18 days and clinical pregnancy test confirmed by the researcher.

All the pregnancy charts were reviewed and confirmed by the scienti-
fic committee of the NFP study centre.

Definition of the unintended pregnancy
Pregnancies were classified as intended or unintended on the basis of
the statements made by the women before conception. At the end of
each menstrual cycle, the woman was asked to state if she was plan-
ning to become pregnant the following cycle. This was documented
in the completed cycle chart. If she forgot to answer this question,
and if a pregnancy occurred in the next cycle, it was always classified
as an unintended pregnancy. If charts did not reach the study centre in
time, the last indicated family planning intention held at the study
centre was used to classify a pregnancy as intended or unintended.

Data analyses
Statistical analyses were carried out using the SAS® package, version
8. We used the non-parametric model of Kaplan–Meier, the survival
curve or actuarial curve, to estimate the rates of unintended pregnancy,
the drop out due to dissatisfaction and the women lost to follow-up
(Kaplan and Meier, 1958; Matthews and Farewell, 1996). We defined the ‘survival’ of a woman as the duration in the study until
she dropped out for the target event (= unintended pregnancy). Other dropouts are censored. The time unit was the menstrual cycle,
therefore, the estimated rates correspond to the life table approach of Tietze and Potter, often used in earlier family planning studies
(Potter, 1966; Tietze and Lewit, 1974). In contrast to the Pearl-index,
the actuarial curves according to Kaplan–Meier represent a time-
related hazard estimation. The results at observation cycle 13 can
roughly be compared to the Pearl-index (= number of unintended
pregnancies per 100 women years, defining 13 cycles to be one
woman year). To compare the actuarial curves of different parameters,
a logrank test was performed. Chi-squared-test was used for categori-
cal data.

The Kaplan–Meier approach was used to calculate the overall
effectiveness rates. Pregnancies due to both method and user failure
were included.

In order to calculate the method effectiveness, we used a modified
model of the ‘perfect/imperfect-use’ approach (discussed earlier,
Trussell and Grummer-Strawn, 1990): pregnancy rates were calcu-
lated according to sexual behaviour: all unintended pregnancies that
occurred during a defined mode of sexual behaviour were related to
all cycles of the corresponding type of sexual behaviour, i. e. unintended pregnancies that occurred in cycles with protected
intercourse during the fertile time were related to only those cycles
with protected intercourse in the fertile time. According to this
approach, we defined the following categories of sexual behaviour
in the fertile time: only abstinence, only protected intercourse,
protected and unprotected intercourse, only unprotected intercourse,
coitus interruptus or genital contact.

Results
To evaluate the overall use-effectiveness as well as the
method-related discontinuation and lost to follow-up, we studied the following groups:

(i) Out of the whole cohort of 900 women and 17 638
cycles, we calculated the overall use-effectiveness as
well as the method-related discontinuation, the lost to
follow-up and the overall duration of study
participation.

(ii) This whole cohort was divided into two subgroups: 322
women used only the STM (‘STM only’-group) and
509 women used the STM with occasional use of barriers in the fertile time (‘STM mix’-group). The ‘STM mix’-group used a barrier method in 53% of their cycles. Life table pregnancy rates have been analysed separately for these two groups. Sixty-nine women did not document their sexual behaviour and were therefore excluded from that analysis.

To analyse the pregnancies according to the modified ‘perfect/imperfect-use’ model, we formed different categories as described earlier.

Client profile

The socio-demographic characteristics of the study population are shown in Table IV. More than 60% of the women were between 19 and 29 years old. Nearly two-thirds of the women had a medium educational level (German baccalaureate or equivalent without a university degree), 52% were nulligravidas; about 20% had reached their desired family size and nearly 60% of the women wanted a further child in the future but not during the year of the study.

Overall unintended pregnancy rates (use-effectiveness)

For the whole cohort, we calculated an unintended pregnancy rate of 1.79 (+/− 0.52 standard error) per 100 women after 13 months of use (Table V); all unintended pregnancies due to method and user failure were included. There was no difference between the learning phase (first 3 months of use) and the subsequent months of use.

We compared the rates of unintended pregnancies between the two groups ‘STM only’ and ‘STM mix’ (Table VI). From a univariate point of view, we calculated slightly different rates of 1.62 (+/− 0.89) for the STM only-group versus 2.02 (+/− 0.72) for the ‘STM mix’ group at 13 cycles—which corresponds approximately to 1 year. At 24 cycles the differences were inversed. These differences were not found to be statistically significant at any time (Logrank test: \( \chi^2 < 0.31 \), hence \( P > 0.60 \)).

Figure 2 illustrates the overlapping standard errors.

**Pregnancy rates in relation to sexual behaviour in the fertile time**

In order to accurately estimate the true method effectiveness according to the ‘perfect/imperfect-use’ approach, every type of sexual behaviour had to be documented. Charting of sexual behaviour occurred in 85% of the cycles; analyses of these cycles showed that in more than a third the STM was used with abstinence during the fertile time, which reflects the ‘perfect-use’ scenario and true method effectiveness. For perfect use, the unintended pregnancy rate was 0.43 per 100 women and 13 cycles (Table VII).

In contrast, the rate of unintended pregnancies in cycles with unprotected intercourse during the fertile time the unintended pregnancy rate was significantly higher with 7.47 per 13 cycles (\( P < 0.00001 \)) and 100 women (Table VII and Figure 3). In 16 of the 22 pregnancy cycles, there was unprotected intercourse in the fertile time.

**Discontinuation**

Discontinuation for dissatisfaction or difficulties with the method, including change to another family planning method was an important parameter of acceptability. The overall discontinuation rate for this category was 9.2 per 100 women at 13 cycles of method-use (Table VIII).

The overall rate of lost to follow up was 6.7% after 13 cycles (Table VIII). Other reasons for discontinuation before cycle 13 given by 34% of the couples included: desire to achieve a pregnancy (8%); separation from partner (2%); medical reasons (4%) and most frequently (22%) because they wished to discontinue participating in the study, although they wished to continue to use the STM.

**Study population duration of participation**

Figure 3 shows the study population over time. The study started with 900 women, 322 of them using ‘STM only’ and 509 of them using ‘STM mix’. 69 women did not document their sexual behaviour. Figure 4 shows that almost 70% of the couples participated in the study for at least 12 cycles. Less than 25% of the couples remained in the study for longer than 24 cycles. Therefore, the time bias influencing the pregnancy rates according to the Trussell approach was not too serious, especially taking into account that all the participants who joined the study were new STM users.

**Discussion**

We believe that this is a significant prospective cohort study of a clearly defined STM method that has several distinctive...
features that has ensured its quality, these include: a large
database; relatively low lost to follow-up rate; inclusion of
the teaching phase; documentation of all sexual behaviour
and classification of the pregnancies as intended or unintended
according to the intention before conception. To ensure the
quality of this study, we have described in depth the recruit-
ment, teaching methods and follow-up of the study partici-
pants, including the inclusion and exclusion criteria. In
addition, we have clearly defined both ‘pregnancy’ and ‘unin-
tended pregnancy’ as well as ‘perfect use’ and ‘imperfect use’
behaviour. Finally, we have described how we analysed the
data from this large prospective cohort study.

While we recognize that critics may argue that this study was
not a randomized controlled trial, it should be recognized that very
few FAB method studies are randomized controlled trials. The
majority are observational studies. This study was a
prospective cohort study which could be seen to be placed
from an evidence-based perspective, between the retrospective
case-control studies and the randomized clinical trials. The
only randomized clinical trials on methods of NFP (Wade
et al., 1981; Medina et al., 1980) are of limited use: they
showed huge recruitment problems and retention as well as
having a very strong selection bias (participants had to agree
to expect quite high failure rates while attracted by free
medical care at study entry), their results are therefore very
questionable (Grimes et al. 2004). Randomized controlled
trials are rarely used to investigate other family planning
methods either because most couples have a preference for a
certain method or do not wish to be randomized. In addition
with most family planning methods, it is impossible to blind
the couples from the allocated method unless comparing
certain different hormonal contraceptives or intrauterine
devices.

A unique feature of this study was that we have applied
advanced analytical methods to our data. We used non-
parametric models to estimate pregnancy rates: we applied
the Kaplan–Meier survival curve to estimate the total preg-
nancy rate and the Trussell approach to calculate the pregnancy
rates according to the proper risk set. Nevertheless, we recog-
nize that these approaches are all influenced to some extent by
the differing risk of conception per cycle (Ecochard, 2006;
Dunson et al., 1999).

The analysis of the efficacy has demonstrated that a STM
that uses two indicators of fertility—temperature and cervical
secretion observations to determine the end of the fertile time
and cervical secretions plus a calculation to identify the onset
of the fertile time—is an effective and acceptable method of
family planning. The overall rates of unintended pregnancies
were 1.8% and the drop out rate for dissatisfaction with the
method was only 9.2 per 100 women after 13 cycles of
method use.

We have demonstrated that the STM is significantly more
effective to avoid pregnancies if used consistently and per-
fected with couples abstaining from intercourse during the
fertile time: 0.4% pregnancy rate per year. We found similar
pregnancy rates for couples who occasionally use barrier

<table>
<thead>
<tr>
<th>Ordinal cycle number</th>
<th>Women exposed</th>
<th>Cumulative number of cycles</th>
<th>Cumulative number of unintended pregnancies</th>
<th>Rate (SE) of unintended pregnancies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>900</td>
<td>900</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>846</td>
<td>2624</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>740</td>
<td>4945</td>
<td>4</td>
<td>0.52 (0.26)</td>
</tr>
<tr>
<td>9</td>
<td>618</td>
<td>6933</td>
<td>10</td>
<td>1.4 (0.44)</td>
</tr>
<tr>
<td>12</td>
<td>509</td>
<td>8571</td>
<td>11</td>
<td>1.57 (0.47)</td>
</tr>
<tr>
<td>13</td>
<td>434</td>
<td>9005</td>
<td>12</td>
<td>1.79 (0.42)</td>
</tr>
<tr>
<td>18</td>
<td>318</td>
<td>10 815</td>
<td>15</td>
<td>2.61 (0.7)</td>
</tr>
<tr>
<td>24</td>
<td>229</td>
<td>12 386</td>
<td>15</td>
<td>2.61 (0.7)</td>
</tr>
</tbody>
</table>

Table V. Overall unintended pregnancy rates per 100 women according to the Kaplan–Meier approach for the whole cohort (n = 900) cut at 24 cycles

Table VI. Overall unintended pregnancy rates per 100 women according to the Kaplan–Meier approach within subcohorts ‘STM only’ (1, bold) and ‘STM mix’ (2, italic) cut at 24 cycles
methods, mainly condoms, during the fertile time as compared to couples who were abstinent. However, there were very few pregnancies in the two samples. We therefore barely had enough statistical power to evaluate the multivariate and adjusted effect of barrier methods for avoiding pregnancy.

The use-effectiveness rates (\(=\) total unintended pregnancy rates) compare very well with the results of the European study and with the interim results of the German database, and with the symptothermal subgroup of the Italian study (Frank-Herrmann et al., 1991; Barbato and Bertolotti, 1988; The European Natural Family Planning Study Groups, 1999). However, the two latter studies were still using the Pearl formula as statistical method. The previous International Rice Fairfield study, that used a NFP-method that preceded the STM, showed reasonable results for developed countries (Rice et al., 1981). The markedly high use-effectiveness rates of our data may partly be explained by the motivation of those couples and their teachers who agreed to participate in the study.

When comparing different methods of family planning, method effectiveness rates are more frequently quoted than the use-effectiveness rates which are strongly dependent on the selection of the study population. The fact that variations in study population make it more difficult to interpret the overall effectiveness of the FAB methods has been discussed in depth (Kambic, 1999).

For a contraceptive method to be rated highly efficient as the hormonal pill, it requires a method failure rate of less than one pregnancy per 100 women per year. Our method-effectiveness of 0.4% can be interpreted as one pregnancy occurring per 3250 cycles (assuming a 13 cycle year). We therefore maintain that the method effectiveness of the STM investigated in this study is comparable to the method effectiveness of modern contraceptive methods like oral contraceptives.

The authors were surprised by the high efficacy during additional barrier method use. We did not find any differences in pregnancy rates between STM only users and STM mix users. Obviously, couples with fertility awareness knowledge are more likely to use condoms more consistently in the fertile time. Most cited NFP studies do not report the quantity of additional barrier method use, yet we have learnt from the European study that it exists to a certain extent within all communities that use NFP methods.

### Table VII. Rates of unintended pregnancies per 100 women and year according to sexual behaviour during the fertile time

<table>
<thead>
<tr>
<th>Cycles</th>
<th>n</th>
<th>%</th>
<th>n per year(^a)</th>
<th>95% CI lower limit(^b)</th>
<th>95% CI upper limit(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstinence in the fertile time</td>
<td>6022</td>
<td>34.14</td>
<td>2  0.43</td>
<td>0.05</td>
<td>1.55</td>
</tr>
<tr>
<td>Protected intercourse in the fertile time</td>
<td>4375</td>
<td>24.80</td>
<td>2  0.59</td>
<td>0.07</td>
<td>2.13</td>
</tr>
<tr>
<td>Unprotected intercourse in the fertile time</td>
<td>2353</td>
<td>13.34</td>
<td>14  7.46</td>
<td>4.15</td>
<td>1.23</td>
</tr>
<tr>
<td>Unprotected and protected intercourse in the fertile time</td>
<td>1183</td>
<td>6.71</td>
<td>2  2.18</td>
<td>0.27</td>
<td>7.65</td>
</tr>
<tr>
<td>Genital contact or coitus interruptus in the fertile time</td>
<td>1080</td>
<td>6.12</td>
<td>1  1.20</td>
<td>0.03</td>
<td>6.50</td>
</tr>
<tr>
<td>No documented sexual behaviour</td>
<td>2625</td>
<td>14.88</td>
<td>1  0.49</td>
<td>0.01</td>
<td>2.72</td>
</tr>
<tr>
<td>Total</td>
<td>17 638</td>
<td>100</td>
<td>22  1.61</td>
<td>1.01</td>
<td>2.43</td>
</tr>
</tbody>
</table>

\(^a\)Rate per 13 cycles and 100 women, calculated according to the formula \(100 \times (1 - (1 - P)^{13})\), \(P\) = probability per cycle (not displayed).

\(^b\)CI = confidence interval, calculated out of the confidence intervals of the pregnancy rates per cycle (not displayed).
For the risk taking couples who had unprotected intercourse during the fertile time, the pregnancy rate increases up to 7.5% per year. We acknowledge that this is surprisingly low. However, one has to realize that the median fertile time determined by the STM is 13 days a cycle (less days after the first year). The potential fertile time is in fact longer than the actual physiological fertile time. Therefore, we recognize that some of the couples were practising conscious intelligent risk taking, i.e. no unprotected intercourse during the few highly fertile days and intercourse only occurred on days at the margins of the beginning and end of the fertile time that would be considered to be a relatively low fertile time. The pregnancy rates during the different days of the identified fertile time varied a lot according to their interval between the day of sexual intercourse and their distance to the estimated day of ovulation (Gnoth et al., 2003). To summarize, one cannot compare the pregnancy rate on possibly fertile days—derived from couples who explicitly wanted to avoid a pregnancy and therefore practised ‘intelligent risk taking’—with pregnancy rates derived from intention to get pregnant.

In addition, we acknowledge that the pregnancy rates during the teaching phase which included the first three cycles were shown to be as low as in the subsequent cycles. We suggest this may be attributed to the high quality training and supervision of the local STM teachers as well as the standardized teaching materials (Arbeitsgruppe NFP, 2004).

Our findings suggest that compared to other FAB methods, e.g. the Billings method, the Creighton model or other cervical secretion methods, a method such as the German STM that uses two indicators of fertility—cervical secretions plus a calculation to identify the onset of the fertile time and basal body temperature and cervical secretions observation to determine the end of the fertile time—is an effective and acceptable method of family planning (World Health Organization, 1981a,b and 1983). In the last two decades, several effectiveness studies have been carried out in developing countries (Thapa et al., 1990, Xu et al., 1994; Arevalo et al., 2004). We believe that the social setting and infrastructure to deliver the FAB methods in these countries is very different from that of the European countries and it is difficult to compare our results directly with these groups. We have therefore only considered studies carried out within the last 25 years within the developed world to compare our results with. There is only a small number of European effectiveness studies based on cervical secretion as a single indicator method, due to the fact that these single indicator methods are not used very frequently in Europe. It is interesting to note that in the WHO five-country study, the pregnancy rates of the ‘Billings method’ (cervical secretion = single indicator) was much higher in the two industrial countries compared to the developing countries: in Ireland with 5.1 pregnancies per 100 women years and 9.4 in New Zealand (World Health Organization 1981a,b).

Pregnancy rates can also be biased if the studies include participants who are likely to have very low fertility, for example, if they include participants who are fully breastfeeding, or they can be biased downward if they include

### Table VIII. Dropout rates with their SE according to the Kaplan–Meier analysis, cut at 24 cycles (n = 900 women)

<table>
<thead>
<tr>
<th>Ordinal cycle number</th>
<th>Rate of dropout for dissatisfaction</th>
<th>Rate of lost to follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.22(0.16)</td>
<td>0.22(0.31)</td>
</tr>
<tr>
<td>3</td>
<td>1.02(0.34)</td>
<td>0.80(0.59)</td>
</tr>
<tr>
<td>6</td>
<td>2.40(0.53)</td>
<td>2.47(1.07)</td>
</tr>
<tr>
<td>9</td>
<td>4.39(0.74)</td>
<td>3.95(1.39)</td>
</tr>
<tr>
<td>12</td>
<td>8.78(1.11)</td>
<td>6.30(1.86)</td>
</tr>
<tr>
<td>13</td>
<td>9.20(1.15)</td>
<td>6.73(1.94)</td>
</tr>
<tr>
<td>18</td>
<td>12.18(1.4)</td>
<td>8.57(2.33)</td>
</tr>
<tr>
<td>24</td>
<td>16.12(1.74)</td>
<td>10.87(2.92)</td>
</tr>
</tbody>
</table>
those cycles from women who are trying to achieve a pregnancy, while excluding all the planned pregnancies from the unintended pregnancy data. (Hilgers and Stanford, 1998; Howard and Stanford, 1999). The Italian STM study had two groups of participants; one group used two parameters to determine the onset of the fertile time and the other group only observation of cervical secretions. The study found that all of the 12 method-related unintended pregnancies occurred in the group that used only cervical secretion rules to determine the onset of the fertile time (460 women, 8140 cycles). Studies that have investigated the efficacy of the hormonal kit Persona and the Computer thermometers have found that the effectiveness is not as high as double-check variation of the STM (Freundl et al., 2003).

Conclusions
This is the first time that a large STM database has been established with sufficient detailed information on sexual behaviour. It enables the true method effectiveness for the STM to be calculated. Our results show that 0.4 unintended pregnancies occurred per 100 women years, if there was abstinence during the fertile time. In addition, our results showed that when barrier methods were used during the fertile time the rate of unintended pregnancies was not significantly different. The use-effectiveness of the method, i.e. the overall pregnancy rate was 1.8% after 13 cycles of use and the discontinuation rate due to dissatisfaction with the STM was only 9.2 per 100 women after 13 cycles; this demonstrates a fairly good acceptability of a FAB method that uses two indicators of fertility—cervical secretions plus a calculation to identify the onset of the fertile time and temperature and cervical secretion observations to determine the end of the fertile time.

Acknowledgements
The authors cordially thank Silvia Heil-Schlehuber, NFP study centre, for her assistance in the follow-up procedures and her valuable comments.

References


Submitted on September 4, 2006; resubmitted on November 8, 2006; resubmitted on December 7, 2006; accepted on January 2, 2007


Submitted on September 4, 2006; resubmitted on November 8, 2006; resubmitted on December 7, 2006; accepted on January 2, 2007