Contraception: *Review*

Effectiveness of Fertility Awareness–Based Methods for Pregnancy Prevention

A Systematic Review

Rachel Peragallo Urrutia, MD, MS, Chelsea B. Polis, PhD, Elizabeth T. Jensen, PhD, Margaret E. Greene, PhD, Emily Kennedy, MA, and Joseph B. Stanford, MD, MSPH

OBJECTIVE: To summarize best available prospective data on typical and perfect use effectiveness of fertility awareness–based methods for avoiding pregnancy.

DATA SOURCES: We conducted a systematic review of studies published in English, Spanish, French, or German by June 2017 in MEDLINE, EMBASE, CINAHL, Web of Science, and ClinicalTrials.gov.

METHODS OF STUDY SELECTION: We reviewed 8,755 unique citations and included 53 studies that contained 50 or greater women using a specific fertility awareness–based method to avoid pregnancy, calculated life table pregnancy probabilities or Pearl rates, and prospectively measured pregnancy intentions and outcomes. We systematically evaluated study quality.

TABULATION, INTEGRATION, AND RESULTS: Of 53 included studies, we ranked 0 high quality, 21 moderate quality, and 32 low quality for our question of interest. Among moderate-quality studies, first-year typical use pregnancy rates or probabilities per 100 woman-years varied widely: 11.2–14.1 for the Standard Days Method, 13.7 for the TwoDay Method, 10.5–33.6 for the Billings Ovulation Method, 4–18.5 for the Marquette Mucus-only Method, 9.0–9.8 for basal body temperature methods, 13.2 for single-check symptothermal methods, 11.2–33.0 for Thyma double-check symptothermal methods, 1.8 for Sensiplan, 25.6 for Persona, 2–6.8 for the Marquette Monitor-only Method, and 6–7 for the Marquette Monitor and Mucus Method. First-year perfect use pregnancy rates or probabilities among moderate-quality studies were 4.8 for the Standard Days Method, 3.5 for the TwoDay Method, 1.1–3.4 for the Billings Ovulation Method, 2.7 for the Marquette Mucus Method, 0.4 for Sensiplan, 12.1 for Persona, and 0 for the Marquette Monitor.

CONCLUSION: Studies on the effectiveness of each fertility awareness–based method are few and of low to moderate quality. Pregnancy rates or probabilities varied widely across different fertility awareness–based

From the Department of Obstetrics and Gynecology, University of North Carolina, Chapel Hill, North Carolina; Reply OB/Gyn & Fertility, Cary, North Carolina; the Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; the Guttmacher Institute, New York, New York; the Department of Epidemiology and Prevention, Wake Forest University School of Medicine, Winston-Salem, North Carolina; GreeneWorks, Washington, DC; and the Department of Family and Preventive Medicine, University of Utah, Salt Lake City, Utah.

Funding for the review was provided through a USAID contract. KNDR Healthcare, LLC provided in-kind donation (EK effort). KNDR had no role in the design of the project or review of the data or manuscript. USAID provided funding starting after the design of the project, the literature search, the abstract review, and the full-text review. They had no role in design of the project or data analysis plan. USAID read the manuscript before publication and provided comments that the authors were not obligated to incorporate.

Presented at the North American Forum on Family Planning, October 14–16, 2017, Atlanta, Georgia; the National Title X Reproductive Health Conference, July 15–18, 2018, Kansas City, Missouri; and the Annual Scientific Assembly Meeting of the International Institute of Restorative Reproductive Medicine, July 25, 2018, Ontario, California.

Chelsea B. Polis is affiliated with the Guttmacher Institute. Her affiliation is included for informational purposes only; this work was not conducted under the auspices of the Guttmacher Institute. The views expressed herein are those of the authors and do not necessarily reflect the views of the Guttmacher Institute.

The authors thank Ammu Vijay and Nancy DeMaria for their help during the abstract and full-text screening portions of the study; Jennifer Walker, MLS, for her assistance with developing and conducting the literature search; and Eugene Urrutia, PhD, for his help with preparation of Figures 2 and 3.

Each author has indicated that he or she has met the journal's requirements for authorship.

Received March 27, 2018. Received in revised form May 10, 2018. Accepted May 24, 2018.

Corresponding author: Rachel Peragallo Urrutia, MD, MS, Department of Obstetrics and Gynecology, University of North Carolina, CB#7570, Chapel Hill, NC 27599-7570; email: rachel_peragallo@med.unc.edu.

Financial Disclosure

Dr. Peragallo Urrutia's salary to the University of North Carolina is paid in part by KNDR Healthcare, LLC, for her work at Reply Ob/Gyn & Fertility. KNDR's mission is to promote fertility awareness–based methods of family planning. The other authors did not report any potential conflicts of interest.

ISSN: 0029-7844/18

methods (and in some cases, within method types), even after excluding low-quality studies. Variability across populations studied precludes comparisons across methods.

(Obstet Gynecol 2018;132:591–604)
DOI: 10.1097/AOG.0000000000002784

Fertility awareness–based methods of family planning help users identify the days of the menstrual cycle when intercourse is most likely to result in pregnancy.[1–3] Users track changes in one or more biomarkers (menstrual dates, basal body temperature, cervical mucus or position, and urinary hormone metabolites) to estimate days of highest fecundity (the fertile window) (Table 1). People who want to avoid pregnancy can avoid intercourse or use additional family planning methods (eg, barriers) during their fertile window.

Users of any method of pregnancy prevention need reliable evidence on perfect and typical use effectiveness to support informed choice. Different study designs for effectiveness estimation have different advantages and disadvantages. Because participants in prospective studies may be highly selected and less representative of typical users than participants in population-based surveys, estimates from clinical studies likely represent "best-case scenarios."[4] Retrospective surveys depend on patients accurately recalling details about reproductive behavior over several years and cannot be used to calculate perfect use estimates. Importantly, low use of fertility awareness–based methods has necessitated lumping users of disparate fertility awareness–based methods together to calculate an overall typical use estimate in retrospective surveys.[5,6] However, this may obscure the effectiveness of individual methods. We conducted a comprehensive search to identify prospective studies estimating typical and perfect use effectiveness of individual fertility awareness–based methods and systematically evaluated their quality.

SOURCES

We registered the study protocol in PROSPERO (CRD42015017760). Our comprehensive search strategy (Appendix 1, available online at http://links.lww.com/AOG/B132) identified studies published in English, Spanish, French, or German (languages spoken by the reviewers) from database inception through June 6, 2017, in MEDLINE, EMBASE, CINAHL, or Web of Science. We manually searched reference lists of included studies and pertinent review articles. We also searched clinicaltrials.gov with key words for ongoing trials published through June 27, 2017. We used Covidence available at www.covidence.org for title–abstract and full-text screening.

STUDY SELECTION

We assembled a multidisciplinary team, including experts in clinical reproductive health care (R.P.U. and J.B.S.), epidemiology (C.B.P. and E.T.J.), demography (M.E.G.), fertility awareness–based method instruction (R.P.U. and E.K.), systematic reviews (C.B.P. and R.P.U.), and conduct of fertility awareness–based method effectiveness studies (J.B.S.). We included studies that prospectively collected information on pregnancy intentions and outcomes and which followed at least 50 participants using a specific fertility awareness–based method to avoid pregnancy. Fertility awareness–based methods were defined as any method using specific rules to approximate the fertile window based on tracking one or more of the following biomarkers: menstrual dates, basal body temperature, cervical mucus, cervical position or consistency, or urinary hormones. Appendix 2, available online at http://links.lww.com/AOG/B132, provides more detailed study inclusion criteria.

Trained investigator dyads independently screened titles and abstracts for inclusion and then screened relevant full-text articles. We reconciled conflicts through discussion. We included studies that did not calculate a pregnancy rate if enough data were provided for the review team to calculate a 12-month or 13-cycle Pearl rate.[7]

Starting with the key questions outlined in our protocol, we drafted an abstraction form and each investigator piloted the form on several studies. We met in person to discuss and finalize the form (Appendix 3, available online at http://links.lww.com/AOG/B132). Investigator dyads independently extracted data for included studies in the Systematic Review Data Repository (https://srdr.ahrq.gov). We extracted data on study design, exclusion and inclusion criteria, sample size, demographics, type of fertility awareness–based method, attrition, pregnancy rate or probabilities, and methodologic approach. We attempted to contact authors of included studies as necessary to obtain clarification on statistical approaches, overlap of published reports, or other aspects of study design, but we did not include any unpublished effectiveness data.

To develop quality evaluation criteria, we reviewed established quality frameworks including the U.S. Preventive Services Task Force and Cochrane.[8–10] Given specific quality issues unique to our topic,[11–16] we modified the Task Force framework to enhance relevance to our specific question of interest. The final

Table 1. Overview of Commonly Used Fertility Awareness–Based Methods for Avoiding Pregnancy

Method Type	Biomarkers Used*	Examples of Specific Methods (List May Not be Exhaustive)	Cycle Length Restrictions	Determination of Potentially Fertile Days*
Calendar methods	Cycle length	*Rhythm*	Rhythm: less than 10-d cycle length variation	Rhythm: 1st fertile day is the shortest prior cycle length minus 19 d; last fertile day is the longest prior cycle length minus 11 d
		Standard Days Method	Standard days: less than 26- or greater than 32-d cycles	Standard Days: days 8–19 are fertile
		Dynamic Optimal Timing	Dynamic optimal timing: 20–40 d	*Dynamic Optimal Timing:* Computer algorithm bases fertile days on cycle length history
Mucus-only methods	Cervical mucus, observed externally at vulva	TwoDay	No restriction	TwoDay: any cervical mucus observed that day or the prior day indicates a fertile day
		Billings Ovulation *Creighton Model Fertility Care System*		Billings, *Creighton*: fertile from change of mucus from baseline infertile pattern to 3rd or 4th day after peak mucus day[†]
		Modified Mucus Method		Modified Mucus: fertile on days of highly fertile-type mucus and then after 2 d
BBT plus	BBT cycle length[§]	Bioself Natural cycles *Ladycomp-Babycomp* *Daysy* *Sophia*	No formal restriction[‡]	Fertile days begin on day 1 of the cycle or based on previous cycle history; the last fertile day is the 3rd day of a temperature shift[§] sometimes, calculated with a computer algorithm based on previous cycles and day of temperature shift
Symptothermal single check	Cervical mucus, BBT	Centre de liaison des Equipes de Recherches *Taking Charge of Your Fertility*	No formal restriction[‡]	Fertile days begin on 1st day of cervical mucus or based on a calendar calculation taking into account previous cycles; the last fertile day is the 3rd day of a BBT shift[§]
Symptothermal double-check	Cervical mucus, BBT, cycle length, cervical position[‖]	Sensiplan Thyma *Couple to Couple League* *Justisse* *Symptotherm*	No formal restriction[‡]	Fertile days begin on day 6 or prior earliest temperature shift minus 6 d or shortest cycle length minus 19 d or 1st day of cervical mucus, whichever is first The last fertile day is 3rd day or more of temperature shift[§] and 3rd day or more after peak mucus day[†] (whichever comes last)
Hormone monitoring (urine)	Urinary hormones, cycle length	Persona contraception monitor	23–35 d	Computer algorithm calculates fertile days based on prior cycle length, rise of urinary estrogen in the current cycle, and luteinizing hormone surges
Symptohormonal	Urinary hormones, cycle length, cervical mucus[‖]	Marquette model,[¶] incorporates Clearblue fertility monitor	No formal restriction[‡]	Marquette: fertile days begin 6 d before earliest peak mucus or luteinizing hormone surge day[†] in prior six cycles or 1st day of cervical mucus or 1st day of rise urinary estrogen metabolite, whichever is first; the last fertile day is the 4th day after the luteinizing hormone surge or the 3rd day after the peak mucus day,[†‖] whichever comes last

(*continued*)

Table 1. Overview of Commonly Used Fertility Awareness–Based Methods for Avoiding Pregnancy (*continued*)

Method Type	Biomarkers Used*	Examples of Specific Methods (List May Not be Exhaustive)	Cycle Length Restrictions	Determination of Potentially Fertile Days*
		Ovarian Monitor		*Ovarian monitor*: fertile days begin based on a urinary estrogen threshold or appearance of cervical mucus and end based on a urinary progesterone threshold

BBT, basal body temperature.
Methods in italics do not have standard effectiveness estimates derived from prospective studies of at least moderate quality in this review; inclusion of any method in this table does not indicate endorsement of the method by the authors.
* The major determination of fertile days is described. There is some variation for specific methods.
† Peak mucus day is generally indicating the last day of highly fertile characteristics of cervical mucus (slippery, clear, stretchy).
‡ Although there are no formal cycle length restrictions, women with long cycles would have prolonged periods of abstinence or alternate method use.
§ Basal body temperature shift generally can be identified by 3 consecutive days with temperature readings reaching at least 0.2°C (0.5°F) above the readings of the previous 6 days.
‖ Optional marker.
¶ There are also mucus-only and monitor-only versions of the Marquette Method.

quality assessment framework contained 13 items, explained in detail in Appendix 4, available online at http://links.lww.com/AOG/B132. To be considered a "high-quality" study for our research question (effectiveness for prevention of unintended pregnancy), a study had to be ranked a "1" on all 13 criteria. Any study with a "3" ranking on any of 13 quality criteria was considered "low quality" for our primary question of interest. We considered all other studies "moderate quality." We rated overall study quality using all peer-reviewed manuscripts pertaining to that study.

RESULTS

We briefly describe all studies meeting our inclusion criteria (Table 2). Next, we describe studies ranked moderate or high quality in greater detail along with the number of criteria (out of 13) that received the highest rating (see Appendix 5, available online at http://links.lww.com/AOG/B132, for full details on effectiveness estimates and populations for the moderate-quality studies). In summarizing effectiveness estimates, we excluded information from studies ranked low quality as a result of concerns about the validity of their estimates. In reporting the results, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidance.[17]

We extracted first-year life table probabilities where possible. If these were not reported, we included 12-month (or 13-cycle) Pearl rates. Because Pearl rate estimates can be significantly affected by use duration, we limited our reporting to include Pearl rates of approximately 1-year duration, thus enhancing cross-study comparability.[4,14,16] We required that perfect use pregnancy rates or probabilities be calculated using a denominator of only perfect use cycles to avoid reporting of underestimated effectiveness estimates.[4,15]

From 8,755 unique records, we excluded 8,312 in title–abstract review and 369 in full-text review (Fig. 1). We included 74 published articles[11,16,18–58,59–88] in English (n=67), German (n=2), French (n=3), and Spanish (n=2). We identified 53 unique studies that describe 65 cohorts of specific fertility awareness–based method users (Table 2). The studies were conducted in a variety of geographic locations, languages, and settings in North and South America, Europe, Africa, and Asia. We ranked 0 as "high quality," 21 as "moderate quality," and 32 as "low quality" for our primary review question. Detailed quality ranking tables for each included study are provided in Appendix 6, available online at http://links.lww.com/AOG/B132, which also provides details on any recalculated specific effectiveness estimates (if multiple estimates were presented in the article). Appendix 5 (available online at http://links.lww.com/AOG/B132) provides detailed information about the population and effectiveness estimates at various time points for each study rated moderate or higher quality. Figures 2 and 3 present the 12-month or 13-cycle typical and perfect use pregnancy rates (ie, Pearl Index) and probabilities (ie, life table, Kaplan-Meier) for new users. Given the small number of moderate or higher quality studies for each method and the heterogeneity of study populations, we determined meta-analysis to be inappropriate.

Six studies assessed the calendar-based method with the Standard Days Method or variants. Of these,

Table 2. Included Studies and Quality Rankings

Method Variant	Country or Countries	No. Enrolled	Quality Ranking (Reasons for Low-Quality Rank)
Calendar-based method: Standard Days Method and variants			
Arevalo (2002)[18,76,78]	Bolivia, Peru, Philippines	478	Moderate
Burkhart (2000),[26] variant	Guatemala	301	Moderate
Dicker (1989),[28,61] variant	Israel	64	Low (description of study population, detection of pregnancy, study duration, and statistical methods)
Gribble (2008)[53]	Benin, Ecuador, El Salvador, Honduras, India, Philippines	1,646	Moderate
Kursun (2014)[65]	Turkey	993	Low (detection of pregnancy, statistical methods)
Sinai (2012),[77] Postpartum bridge to Standard Days	Peru, Guatemala	157	Moderate
Calendar-based method: Rhythm			
Guerrero (1970)[55]	Colombia	83	Low (inclusion–exclusion criteria, populations excluded, pregnancy detection, study duration, statistical methods)
Tietze (1951)[45,61,80]	United States	409	Low (study duration, statistical methods)
Kambic (1996)[61]	NR	2,718	Low (inclusion–exclusion criteria, study population, FABM method, detection of pregnancy, classification of pregnancy, study duration, other)
Mucus-only Method: Billings Ovulation Method and variants			
Ball (1976)[20]	Australia	124	Low (description of study population, detection of pregnancy, study duration, statistical methods)
Bhargava (1996)[23]	India	2,059	Moderate
Gomes (1988)[54]	Bangladesh	416	Low (inclusion–exclusion criteria, description of study population, detection of pregnancy, classification of pregnancy)
Johnston (1979),[59] includes variants	Australia	586	Low (inclusion–exclusion criteria, description of study population, detection of pregnancy, classification of pregnancy, statistical methods)
Klaus (1979)[62,63]	United States	1,139	Low (description of study population, detection of pregnancy, statistical methods, other)
Labbok (1991)[64]	Kenya	368	Low (inclusion–exclusion criteria, detection of pregnancy, statistical methods, other)
Mascarenhas (1979)[69]	India	3,530	Low (detection of pregnancy)
Medina (1980)[70]	Colombia	277	Moderate
Perez (1983)[72,73]	Chile	660	Low (detection of pregnancy, classification of pregnancy)
Thapa (1990)[79]	Indonesia	453	Moderate
Trussell (reanalysis of WHO study) (1991)[11,16,85,86]	New Zealand, India, Ireland, Philippines, El Salvador	869	Moderate
Wade (1981)[81,82]	United States	573	Moderate
Weissmann (1972)[84]	Tonga	282	Low (inclusion–exclusion criteria, detection of pregnancy, classification of pregnancy, study duration, statistical methods)
Xu (1994)[87]	China	688	Low (inclusion–exclusion criteria, detection of pregnancy)

(*continued*)

Table 2. Included Studies and Quality Rankings (*continued*)

Method Variant	Country or Countries	No. Enrolled	Quality Ranking (Reasons for Low-Quality Rank)
Mucus-only Method: Creighton Model Fertility Care System			
Doud (1985)[32,56]	United States	376	Low (detection of pregnancy, classification of pregnancy, statistical methods)
Howard (1999)[56,57]	United States	701	Low (classification of pregnancy)
Fehring (2009),[37,41,56] includes variant	United States	315	Low (description of study population)
Hilgers (1998)[56]	United States	1,876	Low (detection of pregnancy, classification of pregnancy)
Mucus-only Method: TwoDay Method			
Arevalo (2004)[19,78]	Peru, Guatemala, Philippines	450	Moderate
Jennings (2011)[58]	Peru	167	Moderate
Mucus-only methods: Modified Mucus Method and variants			
Dorairaj (1984)[30]	India	5,752	Low (other)
Dorairaj (1991)[29]	India	3,003	Low (inclusion–exclusion criteria, populations excluded, description of study population, detection of pregnancy, study duration, statistical methods)
Kambic (1994)[60]	Liberia	553	Low (detection of pregnancy)
Thapa (1990)[79]	Indonesia	209	Moderate
Thapa (1990),[79] Variant	Indonesia	188	Moderate
Mucus-only methods: Marquette Mucus-only			
Fehring (2013)[39]	United States	160	Moderate
Fehring (2014)[38]	United States	73	Moderate
Fehring (2017)[44]	United States	118	Moderate
Basal body temperature–based methods			
Bartzen (1967),[21] Ogino-Knaus	United States	441	Low (inclusion–exclusion criteria, description of study population, classification of pregnancy, study duration, statistical methods)
Berglund-Scherwitzl (2016),[22] natural cycles	Sweden	4,054	Moderate
Döring (1967),[31] Strenge Kombinierte	Germany	2,276	Low (description of study population, study duration)
Drouin (1994),[35] Bioself	Canada	83	Moderate
Flynn (1991),[46] Bioself	England	131	Low (detection of pregnancy, study duration, other)
Guerrero (1970),[55] Rhythm+basal body temperature	Colombia	208	Low (inclusion–exclusion criteria, populations excluded, detection of pregnancy, study duration, statistical methods)
Marshall (1968),[66,67] Rhythm+basal body temperature	England	502	Low (study duration)
Symptothermal methods: Single-Check			
Ecochard (1996)[36]	Belgium, France, Switzerland	626	Moderate

(*continued*)

Table 2. Included Studies and Quality Rankings (*continued*)

Method Variant	Country or Countries	No. Enrolled	Quality Ranking (Reasons for Low-Quality Rank)
Fehring (2008),[43] Marquette+basal body temperature	United States	76	Low (detection of pregnancy)
Freundl (1999)[51,88]	France, Great Britain, Spain	214	Moderate
Johnston (1979),[59] multiple variants	Australia	460	Low (inclusion–exclusion criteria, description of study population, detection of pregnancy, classification of pregnancy, statistical methods)
Marshall (1976)[68]	United Kingdom	84	Low (study duration)
Rice (1981)[71,74,99]	Canada, Colombia, France, Mauritius, United States	1,022	Low (description of study population, statistical methods)
Weeks (1982),[83] Billings Ovulation +basal body temperature	United States	148	Moderate
Symptothermal methods: double-check			
Frank-Herrmann (2007),[47–49,52] Sensiplan*	Germany	900	Moderate
Freundl (1998),[27,33,34,51,88] Sensiplan and variants*	Austria, Belgium, Czech Republic, France, Germany, Great Britain, Ireland, Italy, Spain, Switzerland	1,046*	Moderate
Medina (1980),[70] Thyma	Colombia	286	Moderate
Wade (1981),[81] Thyma	United States	590	Moderate
Urinary hormone methods: Persona			
Bonnar (1999),[25,50] Persona	England, Ireland, Germany	710	Moderate
Urinary hormone methods: Marquette Monitor			
Bouchard (2012),[24] variant	United States	198	Low (statistical methods)
Fehring (2009)[41,100]	United States	313	Low (description of study population, threats to internal validity)
Fehring (2013)[39]	United States	212	Moderate
Fehring (2014)[38,†]	United States	35	Moderate
Fehring (2017)[44]	United States	160	Moderate
Symptohormonal methods: Marquette Mucus and Monitor			
Fehring (2008)[43]	United States	69	Low (detection of pregnancy)
Fehring (2011)[40]	United States	468	Low (other)
Fehring (2014)[38,†]	United States	160	Moderate
Fehring (2017)[44]	United States	816	Moderate

NR, not reported.

* Freundl 1999 includes incomplete overlap with data from Frank-Herrmann 2007. In the Freundl 1999 study, 339 women with 7,362 cycles were included from the Frank-Herrmann site up to 1995. In Frank-Herrmann 2007, data from 900 women and 17,368 cycles were reported up to 2005.

† Fehring 2014 includes data only for perimenopausal women and were drawn from data sets from the 2009, 2011, and 2017 data sets. It was included separately because the quality of the estimate was higher than any from 2009 or 2011.

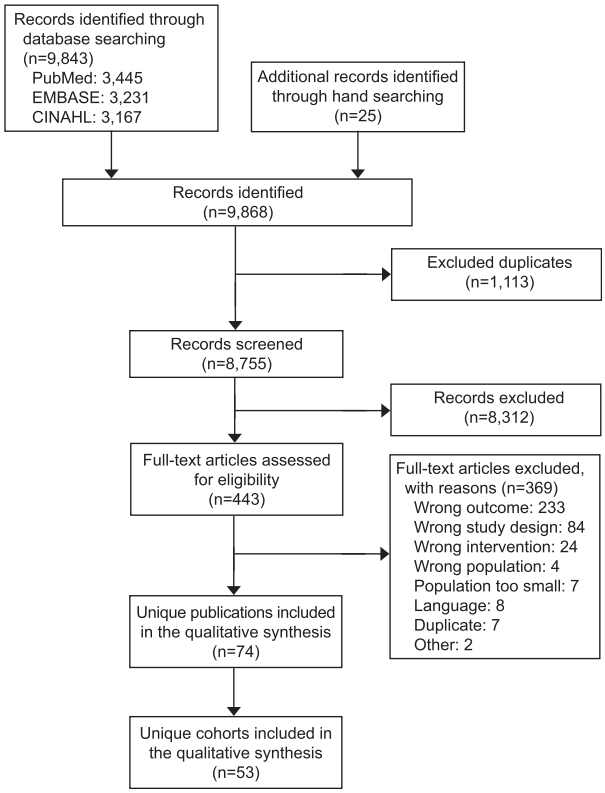

Fig. 1. Study flow diagram.

Urrutia. Fertility Awareness Method Effectiveness. Obstet Gynecol 2018.

two were low quality[28,65] and four were moderate quality.[18,26,53,77] All were conducted in low- to middle-income countries (largely South America) among populations with lower literacy and socioeconomic status. In moderate-quality studies, first-year typical use pregnancy probabilities were 11.2 (95% CI 7.6–14.9), 12.9 (95% CI 8.5–15.3), and 14.1 (95% CI 11.8–16.4).[18,26,53] One moderate-quality study reported a first-year perfect use pregnancy probability of 4.8 (95% CI 2.3–7.1).[18] Two calendar-based rhythm method studies and one meta-analysis were classified as low quality.[55,61,80]

We identified three mucus-based Creighton Model studies[32,37,41,57] and one meta-analysis,[56] all conducted in the United States and Canada. Two studies and the meta-analysis did not provide standard typical use pregnancy estimates. Instead, in these three analyses, pregnancies occurring as a result of intercourse on a day identified by the woman or the couple as fertile were classified by investigators as caused by "achieving [pregnancy]-related behavior." Achieving-related pregnancies thus included both intended or planned and unintended or unplanned pregnancies, and all these pregnancies were excluded from the effectiveness estimates. By excluding most unintended or unplanned pregnancies from effectiveness calculations, these studies underestimate pregnancy probabilities relative to standard typical use calculations, potentially quite substantially.[89] We considered this type of calculation low-quality evidence related to the question of typical use effectiveness. One additional Creighton study provided a standard typical use pregnancy estimate in addition to the described approach but was ranked low for other reasons.[37,41] In addition, perfect use pregnancy probabilities were incorrectly calculated in these studies (eg, using all cycles in the denominator rather than only perfect use cycles).

Two studies provided pregnancy probabilities for the mucus-based TwoDay method[19,58] and we ranked both moderate quality. These studies were conducted in Guatemala, Peru, and the Philippines among populations characterized by lower literacy and socioeconomic status. One study reported first-year typical use (13.7, 95% CI 9.9–17.3) and perfect use (3.5, 95% CI 1.4–5.5) pregnancy probabilities.[19]

Fourteen studies evaluated the mucus-based Billings Ovulation Method or variants. We ranked nine low quality[20,54,59,62,64,72,84,87,90] and five moderate quality.[11,23,70,79,81] Among the moderate-quality studies, one was conducted in the United States, three in low- or middle-income countries, and one was multicountry (New Zealand, India, Ireland, Philippines, El Salvador). The typical use pregnancy probabilities for new users in the first year of use were 10.5 (95% CI 9.1–11.9), 22.4, 22.8, and 33.6.[11,23,70,81] Another moderate study reported a pregnancy probability of 2.5 (95% CI 0.9–4.1)[79] among experienced users. First-year perfect use pregnancy probabilities included 1.1 (95% CI 0.5–1.7)[23] and 3.4.[11] A study providing a reanalysis of the multicountry study data (typical use pregnancy probability 22.8 and perfect use probability 3.4) had the most high-quality indicators of any study in our review (11/13).[11]

We ranked three studies of the mucus-based Modified Mucus Method or variants low quality.[29,30,60,79] We ranked one moderate quality; in this study, both the Modified Mucus Method and a simplified local variant of the Modified Mucus Method were assessed,[79] but pregnancy probabilities among new users were not provided. Among experienced users, the pregnancy probabilities were 10.3 (95% CI 6.0–14.6) for the Modified Mucus Method and 11.5 (95% CI 5.8–17.2) for the users of the simplified variant. Correctly calculated perfect use estimates were not available.

Women who learn the Marquette Method have the option of using only a cervical mucus monitoring

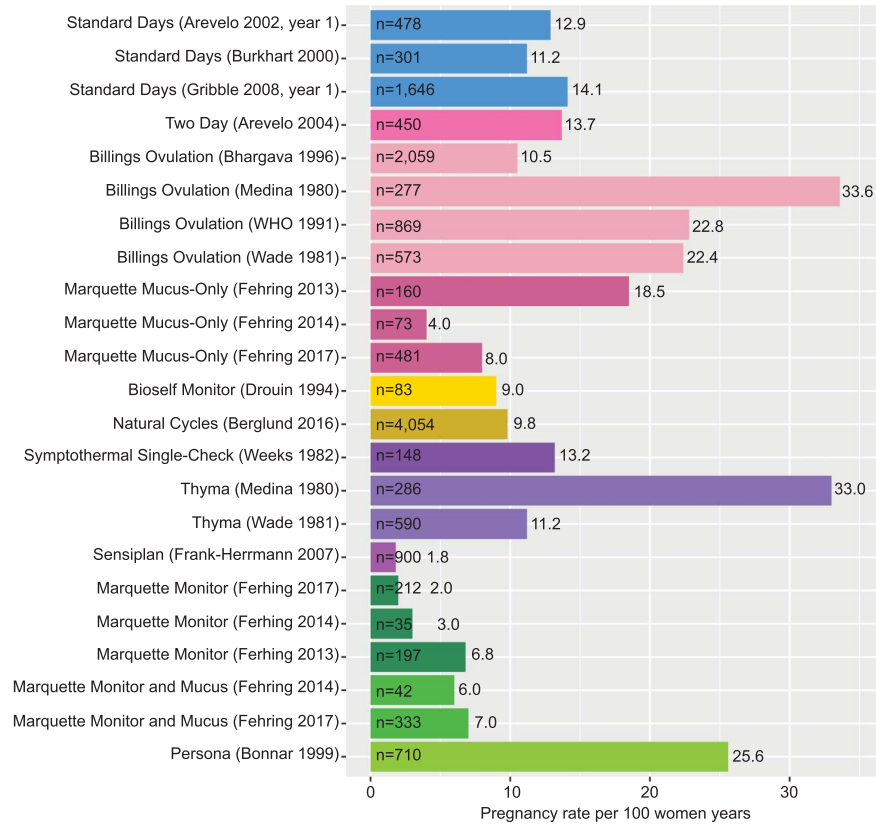

Fig. 2. Overview of typical use pregnancy rates and probabilities among new users during the first year of use. Data included for moderate- or higher quality studies only.

Urrutia. Fertility Awareness Method Effectiveness. Obstet Gynecol 2018.

component. We ranked three studies on the Marquette Mucus-only Method, all in the United States, as moderate quality.[38,39,44] First-year typical use pregnancy probabilities were 8 and 18.5 for all users[39,44] and 4 in a perimenopausal-aged population.[38] The first-year perfect use pregnancy probability was 2.7.[39]

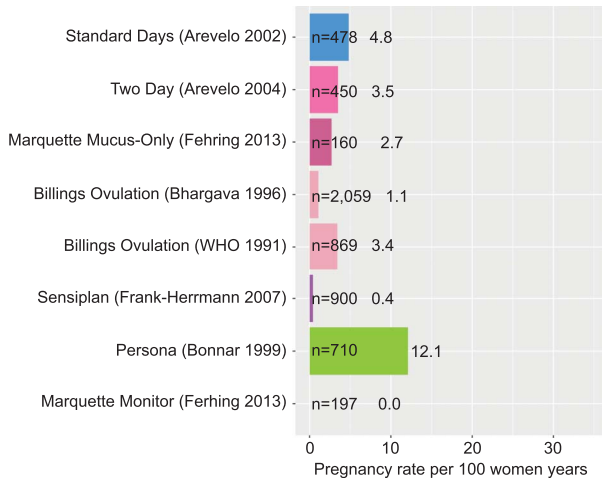

Fig. 3. Overview of perfect use pregnancy rates and probabilities among new users during the first year of use. Data included for moderate- or higher quality studies only.

Urrutia. Fertility Awareness Method Effectiveness. Obstet Gynecol 2018.

We ranked five basal body temperature–based method studies low quality[21,31,46,55,67] and two moderate quality.[22,35] A small, moderate-quality study of the Bioself method (a handheld monitor with an internal computer algorithm) reported a first-year pregnancy probability of 9.0 for new users.[35] A large, moderate-quality study of an algorithm-based internet application called Natural Cycles generated a first-year typical use pregnancy rate of 9.8 among all new users of the application.[22] This estimate was obtained by assuming that the subset of women for whom pregnancy could not be ascertained from prospectively collected pregnancy test data all became pregnant, thus generating a conservative pregnancy rate estimate. Correctly calculated perfect use pregnancy rates were unavailable at the time of our literature search.

Among seven studies investigating single-check symptothermal methods, we ranked four low quality[43,59,68,74] and three moderate quality.[36,51,83] Only one study reported a first-year typical use pregnancy probability of 13.2.[83] Among experienced users in

Europe, typical use effectiveness estimates were 17.6 and 8.5 (95% CI 3.6–13.4).[36,51] Correctly calculated perfect use pregnancy probabilities were unavailable.

Four studies evaluated double-check symptothermal methods (Thyma and Sensiplan); we ranked all four moderate quality.[47,51,70,81] Two evaluated the Thyma method[70,81] and reported typical use pregnancy probabilities of 11.2 and 33. Two studies evaluated Sensiplan: one conducted in Germany[47] and one multicountry European study[51]; these two studies had some participant overlap (less than 50%). The pregnancy probability among new users of this method in Germany was 1.8 (95% CI 1.0–2.6).[47] In the European multicenter study, the first-year pregnancy probability in new and experienced (40%) users combined was 2.6 (95% CI 1.4–3.8).[51] Perfect use pregnancy rates among new users were 0.4 (95% CI 0.1–1.6) among those who used abstinence and 0.6 (95% CI 0.1–2.6) among those who consistently used barrier methods during the fertile time.

One moderate-quality study assessed the symptohormonal Persona urinary hormone monitor method[25] and reported a typical use pregnancy probability of 25.6 and a perfect use probability of 12.1 among new users.[25] Four studies assessed the symptohormonal Marquette Monitor-only Method (Clearblue Fertility Monitor): we ranked three moderate[38,39,44] and one low quality.[41] For new users of the Marquette Monitor-only Method, the typical use pregnancy probabilities in moderate-quality studies were 2 and 6.8.[39,44] One study reported a perfect use pregnancy probability of 0.[39]

Four studies (two ranked moderate[38,44] and two ranked low quality[40,43]) assessed users who chose to use the Monitor Plus Mucus form of the symptohormonal Marquette Method. In the moderate-quality studies, the typical use pregnancy probability for new users was 6 in a group of women 40–54 years old[38] and 7 in a larger group of women whose mean age was 30 years.[44] Correctly calculated perfect use pregnancy probabilities for the Marquette Monitor and Mucus Method were unavailable. The two moderate-quality studies reported perfect use pregnancy probabilities of 1.5–1.6 for users of all three Marquette Methods combined. These were not reported in the table as a result of the combination of methods but are in a similar range of the 0 and 2.7 probabilities reported in the randomized controlled trial for Marquette Monitor Only and Marquette Mucus Only.

Five studies in our review systematically assessed whether participants reported effectiveness estimates for users of barrier methods separately from users of periodic abstinence only. Among these studies, the percent of cycles in which other methods were used ranged from 3% to 55% and the percent of women who ever used barriers ranged from 30% to 46%; however, there were generally no substantive differences in pregnancy estimates for typical and perfect use.[18,19,47,51,83] One exception was from a single-check symptothermal method in Europe, in which the typical use pregnancy rate for abstinence users was higher (19.2) than for barrier method users (3.9).[51]

Seven moderate-quality studies compared users of different fertility awareness–based methods with each other.[38,39,44,51,70,79,81] The Marquette Monitor-only method had lower pregnancy rates than the Marquette Mucus-only method in a randomized trial (6.8 vs 18.5), and the Thyma method had lower pregnancy rates than the Billings Ovulation Method in one randomized control trial in the United States (11.2 vs 22.4, multiple decrement life tables, $P<.01$) but no differently in a second randomized trial in Colombia (33.6 vs 33). Given space limitations, a full description of comparative studies is provided in Appendix 7, available online at http://links.lww.com/AOG/B132.

In clinicaltrials.gov, we identified two ongoing (unpublished) studies; one assessing typical and perfect use effectiveness estimates for Creighton Model users and another assessing effectiveness of a new calendar-based method called Dynamic Optimal Timing.[91–93]

DISCUSSION

Our comprehensive systematic review of the effectiveness of fertility awareness–based methods for avoiding pregnancy reveals that the current evidence base for each method is small and of low to moderate quality. Typical use pregnancy probabilities varied between and among methods; correctly calculated perfect use estimates were less common.

One prior review of fertility awareness–based methods (2005) assessed only randomized controlled trials (n=2).[13] Randomized trials are ideal for minimizing threats to internal validity; however, randomizing women to various contraceptive methods can raise logistic and, sometimes, ethical concerns.[94] A 2011 review included nonrandomized prospective studies but did not comprehensively describe the evidence base.[12]

We identified several important limitations of the existing literature. No study achieved a high-quality rating across all 13 criteria (Table 2). Key limitations included failure to prospectively and regularly collect pregnancy intentions and inappropriate inclusion or

exclusion of pregnancies in effectiveness estimates. An overall limitation of the evidence base is the heterogeneity of populations and settings among the studies. This obscures whether differences in effectiveness estimates are attributable to differences in the methods or in the populations studied. Many studies had high attrition. We were unable to interpret the potential effect of discontinuation on effectiveness estimates as a result of a lack of life table discontinuation rates stratified by discontinuation reasons.

Many people believe they are using a fertility awareness–based method to avoid pregnancy when in fact the method they are using has not undergone a standard prospective effectiveness assessment. Several methods shown in italics in Table 1 (including many increasingly popular internet applications) have no prospective trial data available to support their effectiveness yet are nonetheless being marketed as effective ways to avoid pregnancy.[95] We strongly encourage investigation of the effectiveness of any new adaptation of existing fertility awareness–based methods, including internet-based versions. Likewise, users who make their own modifications should be counseled that effectiveness estimates from a specific fertility awareness–based method may not apply to their use of the method.

Strengths of our approach include a comprehensive search strategy in several languages, dyadic screening and abstraction, and careful assessment of quality. The study was further strengthened by assembling multidisciplinary team encompassing a variety of viewpoints. This fostered an environment of "oppositional collaboration," which enhanced quality and transparency.[96] Our approach excluded retrospective population-based surveys, which limits the generalizability of reported effectiveness estimates in our review. However, it provided the advantages of using prospectively collected information and assessing effectiveness estimations for individual fertility awareness–based methods. Systematic assessment of the risk of bias in individual observational studies is necessarily subjective[17,97,98]; we used an individual component process[17] in which quality criteria were modified to be specific to our topic of study. Other investigators might have chosen different criteria or ranked studies differently for specific quality criteria. However, of the studies ranked low quality, all but four[24,40,41,57] received more than one "low" ranking (out of 13). Of those ranked moderate, all had more than one reason for not achieving the highest quality ranking. Thus, we think it unlikely that the overall rank of the study would have changed with different reviewer teams. We encourage continued discussion of our quality framework (Appendix 4, http://links.lww.com/AOG/B132).

Clinicians advising patients interested in fertility awareness–based methods can share the effectiveness estimates identified in this review. However, they should note that these estimates are not necessarily applicable to all women or populations nor can they be used to definitively compare effectiveness between methods. Fertility awareness–based methods will continue to be relevant and important for many people for a variety of reasons; obtaining the best possible data, and documenting use in varying populations, is critical.

REFERENCES

1. Wilcox AJ, Weinberg CR, Baird DD. Timing of sexual intercourse in relation to ovulation. Effects on the probability of conception, survival of the pregnancy, and sex of the baby. N Engl J Med 1995;333:1517–21.
2. Lynch CD, Jackson LW, Buck Louis GM. Estimation of the day-specific probabilities of conception: current state of the knowledge and the relevance for epidemiological research. Paediatr Perinat Epidemiol 2006;20(suppl 1):3–12.
3. Dunson DB, Baird DD, Wilcox AJ, Weinberg CR. Day-specific probabilities of clinical pregnancy based on two studies with imperfect measures of ovulation. Hum Reprod 1999;14:1835–9.
4. Steiner M, Dominik R, Trussell J, Hertz-Picciotto I. Measuring contraceptive effectiveness: a conceptual framework. Obstet Gynecol 1996;88(suppl):24–30S.
5. Kost K, Singh S, Vaughan B, Trussell J, Bankole A. Estimates of contraceptive failure from the 2002 National Survey of Family Growth. Contraception 2008;77:10–21.
6. Sundaram A, Vaughan B, Kost K, Bankole A, Finer L, Singh S, et al. Contraceptive failure in the United States: estimates from the 2006–2010 National Survey of Family Growth. Perspect Sex Reprod Health 2017;49:7–16.
7. Trussell J. Methodological pitfalls in the analysis of contraceptive failure. Stat Med 1991;10:201–20.
8. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, et al. Current methods of the US Preventive Services Task Force: a review of the process. Am J Prev Med 2001;20(suppl):21–35.
9. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch S, Song F, et al. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7:iii–x, 1–173.
10. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in healthcare. York (UK): University of York; 2009.
11. Trussell J, Grummer-Strawn L. Further analysis of contraceptive failure of the ovulation method. Am J Obstet Gynecol 1991;165:2054–9.
12. Manhart MD, Duane M, Lind A, Sinai I, Golden-Tevald J. Fertility awareness-based methods of family planning: a review of effectiveness for avoiding pregnancy using SORT. Osteopath Fam Physician 2013;5:2–8.
13. Grimes DA, Gallo MF, Grigorieva V, Nanda K, Schulz KF. Fertility awareness-based methods for contraception: system-

atic review of randomized controlled trials. Contraception 2005;72:85–90.

14. Lamprecht V, Trussell J. Natural family planning effectiveness: evaluating published reports. Adv Contracept 1997;13:155–65.
15. Trussell J, Hatcher RA, Cates W Jr, Stewart FH, Kost K. A guide to interpreting contraceptive efficacy studies. Obstet Gynecol 1990;76:558–67.
16. Trussell J, Grummer-Strawn L. Contraceptive failure of the ovulation method of periodic abstinence. Fam Plann Perspect 1990;22:65–75.
17. Liberati M, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analysis of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009;6:e1000100.
18. Arevalo M, Jennings V, Sinai I. Efficacy of a new method of family planning: the Standard Days Method. Contraception 2002;65:333–8.
19. Arevalo M, Jennings V, Nikula M, Sinai I. Efficacy of the new TwoDay Method of family planning. Fertil Steril 2004;82:885–92.
20. Ball M. A prospective field trial of the 'ovulation method' of avoiding conception. Eur J Obstet Gynecol Reprod Biol 1976;6:63–6.
21. Bartzen PJ. Effectiveness of the temperature rhythm system of contraception. Fertil Steril 1967;18:694–706.
22. Berglund Scherwitzl E, Gemzell Danielsson K, Sellberg JA, Scherwitzl R. Fertility awareness-based mobile application for contraception. Eur J Contracept Reprod Health Care 2016;21:234–41.
23. Bhargava H, Bhatia JC, Ramachandran L, Rohatgi P, Sinha A. Field trial of billings ovulation method of natural family planning. Contraception 1996;53:69–74.
24. Bouchard T, Fehring RJ, Schneider M. Efficacy of a new postpartum transition protocol for avoiding pregnancy. J Am Board Fam Med 2013;26:35–44.
25. Bonnar J, Flynn A, Freundl G, Kirkman R, Royston R, Snowden R. Personal hormone monitoring for contraception. Br J Fam Plann 1999;24:128–34.
26. Burkhart MC, de Mazariegos L, Salazar S, Lamprecht VM. Effectiveness of a standard-rule method of calendar rhythm among Mayan couples in Guatemala. Int Fam Plan Perspect 2000;26:131–6.
27. De Leizaola-Cordonnier A. Natural family planning effectiveness in Belgium. Adv Contracept 1995;11:165–72.
28. Dicker D, Wachsman Y, Feldberg D, Ashkenazi J, Yeshaya A, Goldman JA. The vaginal contraceptive diaphragm and the condom–a reevaluation and comparison of two barrier methods with the rhythm method. Contraception 1989;40:497–504.
29. Dorairaj K. The modified mucus method in India. Am J Obstet Gynecol 1991;165:2066–7.
30. Dorairaj K. Use-effectiveness of fertility awareness among the urban poor. Soc Action 1984;34:286–306.
31. Döring GK. On the reliability of temperature method for contraception [in German]. Dtsch Med Wochenschr 1967;92:1055–61.
32. Doud J. Use-effectiveness of the Creighton model of NFP. Int Rev Nat Fam Plann 1985;9:54–72.
33. de Leizaola MA. Prospective study of the effectiveness of a new symptothermal method of natural family planning [in French]. J Gynecol Obstet Biol Reprod (Paris) 1998;27:174–80.
34. De Leizaola MA. The 1st phase of a prospective effectiveness study of natural family planning conducted in the French-speaking part of Belgium [in French]. J Gynecol Obstet Biol Reprod (Paris) 1994;23:359–64.
35. Drouin J, Guilbert EE, Désaulniers G. Contraceptive application of the Bioself fertility indicator. Contraception 1994;50:229–38.
36. Ecochard R, Pinguet F, Ecochard I, De Gouvello R, Guy M, Huy F. Analysis of natural family planning failures. In 7007 cycles of use [in French]. Contracept Fertil Sex 1998;26:291–6.
37. Fehring RJ, Lawrence D, Philpot C. Use effectiveness of the Creighton model ovulation method of natural family planning. J Obstet Gynecol Neonatal Nurs 1994;23:303–9.
38. Fehring RJ, Mu Q. Cohort efficacy study of natural family planning among perimenopause age women. J Obstet Gynecol Neonatal Nurs 2014;43:351–8.
39. Fehring RJ, Schneider M, Raviele K, Rodriguez D, Pruszynski J. Randomized comparison of two Internet-supported fertility-awareness-based methods of family planning. Contraception 2013;88:24–30.
40. Fehring RJ, Schneider M, Raviele K. Pilot evaluation of an internet-based natural family planning education and service program. J Obstet Gynecol Neonatal Nurs 2011;40:281–91.
41. Fehring RJ, Schneider M, Barron ML, Raviele K. Cohort comparison of two fertility awareness methods of family planning. J Reprod Med 2009;54:165–70.
42. Fehring RJ, Schneider M, Barron ML, Pruszynski J. Influence of motivation on the efficacy of natural family planning. MCN Am J Matern Child Nurs 2013;38:352–8.
43. Fehring RJ, Schneider M, Barron ML. Efficacy of the Marquette Method of natural family planning. MCN Am J Matern Child Nurs 2008;33:348–54.
44. Fehring RJ, Schneider M. Effectiveness of a natural family planning service program. MCN Am J Matern Child Nurs 2017;42:43–9.
45. Fleck S, Snedeker EF, Rock J. The contraceptive safe period. N Engl J Med 1940;223:1005–9.
46. Flynn A, Pulcrano J, Royston P, Spieler J. An evaluation of the Bioself 110 electronic fertility indicator as a contraceptive aid. Contraception 1991;44:125–39.
47. Frank-Herrmann P, Heil J, Gnoth C, Toledo E, Baur S, Pyper C, et al. 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. Hum Reprod 2007;22:1310–9.
48. Frank-Herrmann P, Freundl G, Gnoth C, Godehardt E, Kunert J, Baur S, et al. Natural family planning with and without barrier method use in the fertile phase: efficacy in relation to sexual behavior: a German prospective long-term study. Adv Contracept 1997;13:179–89.
49. Frank-Herrmann P, Freundl G, Baur S, Bremme M, Döring GK, Godehardt EAJ, et al. Effectiveness and acceptability of the symptothermal method of natural family planning in Germany. Am J Obstet Gynecol 1991;165:2052–4.
50. Freundl G, Bonnar J, Flynn A, Frank-Herrmann P, Kirkmann R, Snowden R. Contraception per computer. Hormone system persona–results of studies in Germany [in German]. Fortschr Med 1998;116:47–8.

51. European multicenter study of natural family planning (1989–1995): efficacy and drop-out. The European Natural Family Planning Study Groups. Adv Contracept 1999;15:69–83.
52. Gnoth C, Frank-Herrmann P, Schmoll A, Godehardt E, Freundl G. Cycle characteristics after discontinuation of oral contraceptives. Gynecol Endocrinol 2002;16:307–17.
53. Gribble JN, Lundgren RI, Velasquez C, Anastasi EE. Being strategic about contraceptive introduction: the experience of the Standard Days Method. Contraception 2008;77:147–54.
54. Gomes I, Congdon D. The natural family planning programme in Bangladesh. Int J Fertil 1988;33(suppl):52–3.
55. Guerrero V, Lores H. Evaluation of periodic continence as a method of family planning [in Spanish]. Rev Colomb Obstet Ginecol 1970;21:545–52.
56. Hilgers TW, Stanford JB. Creighton Model NaProEducation Technology for avoiding pregnancy. Use effectiveness. J Reprod Med 1998;43:495–502.
57. Howard MP, Stanford JB. Pregnancy probabilities during use of the Creighton model fertility care system. Arch Fam Med 1999;8:391–402.
58. Jennings V, Sinai I, Sacieta L, Lundgren R. TwoDay Method: a quick-start approach. Contraception 2011;84:144–9.
59. Johnston JA, Roberts DB, Spencer RB. NFP services and methods in Australia: a survey evaluation. Int Rev Nat Fam Plann 1979;3:20–53.
60. Kambic RT, Lanctot CA, Wesley R. Trial of a new method of natural family planning in Liberia. Adv Contracept 1994;10:111–9.
61. Kambic RT, Lamprecht V. Calendar rhythm efficacy: a review. Adv Contracept 1996;12:123–8.
62. Klaus H, Goebel JM, Muraski B, Egizio MT, Weitzel D, Taylor RS, et al. Use-effectiveness and client satisfaction in six centers teaching the Billings Ovulation Method. Contraception 1979;19:613–29.
63. Klaus H, Goebel J, Woods RE, Castles M, Zimny G. Use-effectiveness and analysis of satisfaction levels with the Billings Ovulation Method: two-year pilot study. Fertil Steril 1977;28:1038–43.
64. Labbok MH, Stallings RY, Shah F, Perez A, Klaus H, Jacobson M, et al. Ovulation method use during breastfeeding: is there increased risk of unplanned pregnancy? Am J Obstet Gynecol 1991;165:2031–6.
65. Kursun Z, Cali S, Sakarya S. The Standard Days Method(®): efficacy, satisfaction and demand at regular family planning service delivery settings in Turkey. Eur J Contracept Reprod Health Care 2014;19:203–10.
66. Marshall J, Rowe B. The effect of personal factors on the use of the basal body temperature method of regulating births. Fertil Steril 1972;23:417–21.
67. Marshall J. A field trial of the basal-body-temperature method of regulating births. Lancet 1968;2:8–10.
68. Marshall J. Cervical-mucus and basal body-temperature method of regulating births. Field trial. Lancet 1976;2:282–3.
69. Mascarenhas MM, Lobo A, Ramesh AS. Contraception and the effectiveness of the ovulation method in India. Trop Doct 1979;9:209–11.
70. Medina JE, Cifuentes A, Abernathy JR, Spieler JM, Wade ME. Comparative evaluation of two methods of natural family planning in Columbia. Am J Obstet Gynecol 1980; 138:1142–7.
71. Parenteau-Carreau S, Lanctot CA, Rice FJ. International Fairfield Study on the effectiveness of the sympto-thermal method of family planning. Canadian results compared to world results [in French]. Vie Med Can Fr 1976;4:145–53.
72. Pérez A, Zabala A, Larrain A, Widmer S, Núñez M, Baranda B, et al. Clinical effectiveness of the ovulation method (Billings) [in Spanish]. Rev Chil Obstet Ginecol 1983;48:97–107.
73. Perez A, Labbok M, Barker D, Gray R. Use-effectiveness of the ovulation method initiated during postpartum breastfeeding. Contraception 1988;38:499–508.
74. Rice FJ, Lanctôt CA, Garcia-Devesa C. Effectiveness of the sympto-thermal method of natural family planning: an international study. Int J Fertil 1981;26:222–30.
75. Rice FJ, Lanctot CA, Garcia-Devasa C. Biological effectiveness of the sympto thermal rhythm method. An international study. A preliminary report. Acta Med Rom 1978;16:349–56.
76. Sinai I, Lundgren RI, Gribble JN. Continued use of the Standard Days Method. J Fam Plann Reprod Health Care 2012;38: 150–6.
77. Sinai I, Cachan J. A Bridge for postpartum women to Standard Days Method®: II. Efficacy study. Contraception 2012;86:16–21.
78. Sinai I, Arévalo M. It's all in the timing: coital frequency and fertility awareness-based methods of family planning. J Biosoc Sci 2006;38:763–77.
79. Thapa S, Wonga MV, Lampe PG, Pietojo H, Soejoenoes A. Efficacy of three variations of periodic abstinence for family planning in Indonesia. Stud Fam Plann 1990;21:327–34.
80. Tietze C, Poliakoff SR, Rock J. The clinical effectiveness of the rhythm method of contraception. Fertil Steril 1951;2:444–50.
81. Wade ME, McCarthy P, Braunstein GD, Abernathy JR, Suchindran CM, Harris GS, et al. A randomized prospective study of the use-effectiveness of two methods of natural family planning. Am J Obstet Gynecol 1981;141:368–76.
82. Wade ME, McCarthy P, Abernathy JR, Harris GS, Danzer HC, Uricchio WA. A randomized prospective study of the use-effectiveness of two methods of natural family planning: an interim report. Am J Obstet Gynecol 1979;134:628–31.
83. Weeks JR. An evaluation of the use-effectiveness of fertility awareness methods of family planning. J Biosoc Sci 1982;14:25–32.
84. Weissmann MC, Foliaki L, Billings EL, Billings JJ. A trial of the ovulation method of family planning in Tonga. Lancet 1972;2:813–6.
85. A prospective multicentre trial of the ovulation method of natural family planning. II. The effectiveness phase. Fertil Steril 1981;36:591–8.
86. A prospective multicentre trial of the ovulation method of natural family planning. I. The teaching phase. Fertil Steril 1981;36:152–8.
87. Xu JX, Yan JH, Fan DZ, Zhang DW. Billings natural family planning in Shanghai, China. Adv Contracept 1994; 10:195–204.
88. Prospective European multi-center study of natural family planning (1989–1992): interim results. The European Natural Family Planning Study Groups. Adv Contracept 1993;9:269–83.
89. Trussell J. Analyzing contraceptive failure. J Reprod Med 1999;44:478–9.
90. Mascarenhas MJ, Lobo A, Ramesh AS. The use effectiveness of the ovulation method in India. Bangalore (India): Center for Research, Education, Service and Training for Family Life Promotion; 1979.

91. Simmons RG, Shattuck DC, Jennings VH. Assessing the efficacy of an app-based method of family planning: the dot study protocol [published erratum appears in JMIR Res Protoc 2018;7:e9]. JMIR Res Protoc 2017;6:e5.
92. ClinicalTrials.gov. Creighton model effectiveness, intentions and behaviors assessment (CEIBA). Available at: https://clinicaltrials.gov/ct2/show/NCT01012596. Retrieved July 9, 2018.
93. ClinicalTrials.gov. Effectiveness and efficacy of dynamic optimal timing, a smart phone app for avoiding pregnancy: an observational study. Available at: https://clinicaltrials.gov/ct2/show/NCT02833922. Retrieved July 9, 2018.
94. Trussell J. Understanding contraceptive failure. Best Pract Res Clin Obstet Gynaecol 2009;23:199–209.
95. Polis CB. Published analysis of contraceptive effectiveness of Daysy and DaysyView app is fatally flawed. Reproductive Health 2018;15:113.
96. Stanford JB. Population, reproductive, and sexual health: data are essential where disciplines meet and ideologies conflict. Front Public Health 2016;4:27.
97. Hodges JS. Are quality assessment methods any good? J Evid Base Dent Pract 2004;4:24–31.
98. Jüni P, Wischi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;282:1054–60.
99. Rice FJ, Lanctôt CA, Garcia-Devesa C. Effectiveness of the sympto-thermal method of natural family planning: an international study. Int J Fertil 1981;26:222–30.
100. Fehring RJ, Schneider M, Raviele K, Barron ML. Efficacy of cervical mucus observations plus electronic hormonal fertility monitoring as a method of natural family planning. J Obstet Gynecol Neonatal Nurs 2007;36:152–60.