COMMENTARY

Variability in Contraceptive Clinical Trial Design and the Challenges in Making Comparisons Across Trials

David Portman · David L. Eisenberg · Patty Cason · Brandon Howard ⓘ

ABSTRACT

Clinical trials of contraceptives have often differed in their study designs, making cross-trial comparisons difficult. This brief report outlines some of the technical design features that can vary from trial to trial. For example, the overall number of menstrual cycles in a study has substantial impact on the final efficacy determination; however, the rules related to qualifying cycles can differ based on the length of the study and the statistical analysis plan. In two commonly used methods of calculating efficacy, the Pearl Index and the time-to-event analysis, inclusion of fewer menstrual cycles results in higher calculated failure rates. Statistical analysis plans for contraceptive trials have sometimes excluded menstrual cycles because of an absence of documented vaginal intercourse and the concomitant use of another birth control method. Other design features that have varied between contraceptive trials relate to body mass index inclusion/exclusion criteria and the definition of "on-treatment" pregnancy. In addition, study designs of non-hormonal products can differ from those of hormonal products in their length and rules for qualifying cycles. The Draft Guidance for Hormonal Contraception, published in 2019 by the US Food and Drug Administration (FDA), will hopefully lead to more uniform trial design in the future, particularly for hormonal products. In the meantime, health care providers and patients should be aware of the nuances in trial design that make direct comparisons about relative efficacy challenging.

D. Portman
Department of Obstetrics and Gynecology, Ohio State University College of Medicine, Columbus, OH, USA

D. L. Eisenberg
Department of Obstetrics and Gynecology, Washington University in St. Louis School of Medicine, St. Louis, MO, USA

P. Cason
Envision Sexual and Reproductive Health, Los Angeles, CA, USA

B. Howard (✉)
Evofem Biosciences, Inc., 12400 High Bluff Drive, Suite 600, San Diego, CA 92130, USA
e-mail: bhoward@evofem.com

Key Summary Points

The study designs of contraceptive clinical trials have historically differed in technical but impactful ways, making it difficult to directly compare efficacy results across trials.

One such design feature that has often differed between contraceptive trials is the set of rules related to qualifying menstrual cycles; in some trials, the absence of documented vaginal intercourse and/or the concomitant use of another birth control method can exclude a menstrual cycle from the efficacy determination.

For two of the most commonly used methods of calculating efficacy, the Pearl Index and the time-to-event analysis, inclusion of fewer menstrual cycles results in higher calculated failure rates.

Non-hormonal contraceptive study designs can differ from hormonal study designs in length and in rules for qualifying menstrual cycles.

Recent FDA guidance will hopefully increase uniformity in trial design, leading to more accurate comparisons between trials and conceptive products.

COMMENTARY

One of the primary characteristics that people consider when selecting a contraceptive method is how well it works to prevent pregnancy [1, 2]. However, there is limited awareness of the variable ways that efficacy is measured in clinical trials or of how "efficacy" differs from "effectiveness" as determined via non-interventional studies. Many family planning providers and patients are also unaware that a commonly used counseling tool that groups contraceptives into three tiers conveys effectiveness rather than efficacy and is not based on head-to-head clinical trial data [3].

Efficacy determinations from clinical trials of contraceptives are meant to convey, as closely as possible, an estimate of risk of pregnancy based on the method, while controlling for other possible factors that might impact the efficacy rate. Although contraceptive clinical trials share this common goal, their clinical study designs have historically differed in ways that can impact the ultimate efficacy determination. Older clinical trials of hormonal contraceptives often enrolled less socioeconomically and ethnically diverse populations and excluded those with a high body mass index (BMI), which research indicates likely led to better reported efficacy than would have been seen if the enrolled population had been more diverse and representative of today's heavier population [4–8]. For example, a US Food and Drug Administration (FDA) meta-analysis suggests that women with obesity have a 44% higher failure rate during combined oral conceptive (COC) use compared to non-obese women after adjusting for age and race [4].

The overall number of menstrual cycles included in the analysis is another, perhaps lesser known, clinical trial design feature that can impact the efficacy calculation. This number can vary based on the length of the study and the specific statistical analysis plan (SAP) of a given clinical trial. Two of the most commonly used approaches to calculate efficacy in contraceptive clinical trials, the Pearl Index and the time-to-event analysis, both take into account the number of qualifying menstrual cycles versus the number of on-treatment pregnancies.

The Pearl Index, defined as the number of pregnancies per 100 woman-years, is commonly used for hormonal contraceptives and assumes a constant failure rate over time. The Pearl Index is calculated using the number of qualifying menstrual cycles in the denominator, as follows: [(number of pregnancies × 13 cycles/total number of 28-day cycles) × 100].

In contrast, the time-to-event analysis, also called life table analysis or cumulative failure rate, reports failure rates at set time points. The time-to-event approach is based on the Kaplan–Meier calculation, which is a survival analysis. The survival analysis depends on the

duration of time (including only qualifying cycles) from enrollment until exiting the trial or becoming pregnant, with the number of participants still active in the trial at the time of each pregnancy used to calculate the cumulative failure rate. Although their calculations differ, for both the Pearl Index and time-to-event analysis, inclusion of fewer qualifying menstrual cycles results in higher calculated failure rates. This is an important consideration when comparing findings from studies of varying duration, e.g., 7 cycles vs 13 cycles.

For current FDA approval, the two most significant reasons that cycles must be omitted are no documented vaginal intercourse and/or concomitant birth control use, such as when condoms are used for prevention of sexually transmitted infection. However, on the basis of our knowledge of phase 3, US-based contraceptive clinical trials initiated during or after 2007, seven of the 11 trials that assessed efficacy excluded cycles in which concomitant contraception was used [9–15], and four trials did not [16–19]. Furthermore, four of these 11 trials documented the occurrence of regular vaginal intercourse during the trial and excluded cycles reporting no intercourse from the efficacy analysis [9–12], while seven trials did not [13–19]. Although this literature search was not systematic and was limited by the fact that not all details of the statistical methods used in clinical trials are always published or readily accessible online, we believe our findings underscore the variability in how contraceptive clinical trials have historically defined qualifying cycles. Given the substantial impact of qualifying cycles on the final efficacy determination, this variability in trial design and efficacy results makes it very challenging to make comparisons across trials.

Another possible reason for exclusion of cycles from the efficacy analysis is a menstrual cycle length that is too long or too short (typically < 21 days and > 35 days). This is applicable primarily for non-hormonal contraceptives. One recently completed clinical trial for the newly approved vaginal pH modulator (VPM) gel excluded cycles where no vaginal intercourse was documented and cycles in which concomitant contraception was used, as well as

on-trial menstrual cycles that were not 21–35 days in length [9]. The impact of the number of cycles on the efficacy calculation was recently highlighted by Chappell et al. [20]. In this analysis, the perfect-use cumulative failure rate of VPM ranged from 9.99% to 6.67% depending on whether non-standard-length menstrual cycles were included and whether women who had incomplete washout of previous hormonal contraception were removed, among other factors.

Clinical studies may also differ in the criteria used to determine the number of pregnancies in the calculation. Newer clinical trials differ from older trials in that they require more frequent, sensitive, and mandatory pregnancy testing and more frequent utilization of high-resolution transvaginal ultrasound, which results in the identification of more and earlier pregnancies [21]. Adding to the difficulties in comparing clinical trials, definitions of "on-treatment" pregnancies have not been uniform, with some hormonal trials including pregnancies occurring up to 7 days after the last treatment cycle versus others that have included pregnancies occurring up to 14 days after the last treatment cycle [21, 22]. Each of these factors has contributed in varying degrees to the clearly decreasing efficacy results seen in more recent clinical trials [21].

In the USA, contraceptive clinical trials are designed by sponsors to maximize the likelihood of product approval from the FDA. However, FDA guidance has changed over time and historically has taken the form of one-on-one discussions between the FDA and the trial sponsor that ultimately guided choices in study design and inclusion criteria. Lack of clearly articulated and consistent regulatory guidance for hormonal or non-hormonal contraceptive trials has led to inconsistent clinical trial designs. The FDA's 2019 Draft Guidance for hormonal contraception provides an important starting point for more uniform trial design that will ultimately help family planning providers and patients more accurately evaluate findings across future studies [23]. Among other features, the FDA recommendations include placing no restrictions on BMI in the enrollment criteria, counting pregnancies that occur within 7 days

of the last treatment, and using the Pearl Index as the primary endpoint, calculated over the first year of use and including only cycles during which vaginal intercourse occurred and no concomitant contraception was used. The FDA should also consider providing similar published guidelines for non-hormonal and device clinical trials.

As contraceptive clinical trials in the USA become increasingly uniform in design and inclusive in nature, making cross-trial comparisons will become more feasible and the findings will become more relevant to the USA population as a whole. However, features of clinical trial design still differ between hormonal contraceptives and non-hormonal contraceptives, suggesting a need to educate family planning providers and patients. For example, trials of longer duration (e.g., 13 cycles as is common for hormonal contraceptives) will likely result in lower failure rates compared to trials of shorter duration (e.g., 7 cycles as is common for non-hormonal contraceptives) given the reduced number of cycles contributing to the analysis. Also, family planning providers and patients should be aware that, in some aspects, adherence to the FDA's 2019 Draft Guidance will widen the gap between efficacy rates in the clinical trial setting and effectiveness rates in the real world. For example, exclusion of cycles in which no intercourse is documented and/or another method of contraception is used is not reflective of how contraceptives are used in the real world [24].

Achieving inclusive and uniform trial design will take time and will not address the inconsistencies in past trials. In the meantime, family planning providers should understand some of the important nuances in contraceptive study designs in order to put efficacy data into context for the patient. This in turn will support reproductive autonomy by enabling people to make informed contraceptive choices. Future studies of real-world, prospective use of different contraceptive methods across a diverse population of people would provide more relevant and consistent data that allow patients to contextualize and understand their pregnancy risk and control their own reproductive life.

ACKNOWLEDGEMENTS

Employee and Shareholder of Evofem Biosciences, Inc.

Compliance with Ethics Guidelines. This article is based on previously conducted studies and does not contain any new studies with human participants or animals performed by any of the authors.

Data Availability. Data sharing is not applicable to this article as no datasets were generated or analyzed.

Open Access. This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc/4.0/.

REFERENCES

1. Mansour D. International survey to assess women's attitudes regarding choice of daily versus nondaily female hormonal contraception. Int J Womens Health. 2014;6:367–75.

2. Madden T, Secura GM, Nease RF, Politi MC, Peipert JF. The role of contraceptive attributes in women's contraceptive decision making. Am J Obstet Gynecol. 2015;213(46):e1-6.

3. Centers for Disease Control and Prevention. Effectiveness of family planning methods. 2014. https://www.cdc.gov/reproductivehealth/contraception/unintendedpregnancy/pdf/family-planning-methods-2014.pdf. Accessed 4 Feb 2018.

4. Yamazaki M, Dwyer K, Sobhan M, et al. Effect of obesity on the effectiveness of hormonal contraceptives: an individual participant data meta-analysis. Contraception. 2015;92:445–52.

5. Edelman A, Trussell J, Aiken ARA, Portman DJ, Chiodo JA, Garner EIO. The emerging role of obesity in short-acting hormonal contraceptive effectiveness. Contraception. 2018;97:371–7.

6. Gerlinger C, Trussell J, Mellinger U, et al. Different Pearl Indices in studies of hormonal contraceptives in the United States: impact of study population. Contraception. 2014;90:142–6.

7. Westhoff CL, Torgal AT, Mayeda ER, Shimoni N, Stanczyk FZ, Pike MC. Predictors of noncompliance in an oral contraceptive clinical trial. Contraception. 2012;85:465–9.

8. Sundaram A, Vaughan B, Kost K, 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.

9. Thomas MA, Chappell BT, Maximos B, Culwell KR, Dart C, Howard B. A novel vaginal pH regulator: results from the phase 3 AMPOWER contraception clinical trial. Contracept X. 2020;2:100031.

10. ClinicalTrials.gov. E4 FREEDOM (female response concerning efficacy and safety of estetrol/drospirenone as oral contraceptive in a multicentric study)—United States/Canada study. 2020. https://clinicaltrials.gov/ct2/show/NCT02817841?term=NCT02817841&draw=2&rank=1. Accessed 9 Apr 2020.

11. ClinicalTrials.gov. Safety, efficacy, tolerability and pharmacokinetics of LF111 (drospirenone 4.0 mg) during 13 cycles. 2019. https://clinicaltrials.gov/ct2/show/NCT02269241?term=NCT02269241&draw=2&rank=1. Accessed 9 Apr 2020.

12. Agile Therapeutics. Agile Patch (AG200-15). 2019. https://www.fda.gov/advisory-committees/advisory-committee-calendar/october-30-2019-meeting-bone-reproductive-and-urologic-drugs-advisory-committee-meeting-announcement#event-materials. Accessed 26 Feb 2020.

13. Portman DJ, Kaunitz AM, Howard B, Weiss H, Hsieh J, Ricciotti N. Efficacy and safety of an ascending-dose, extended-regimen levonorgestrel/ethinyl estradiol combined oral contraceptive. Contraception. 2014;89:299–306.

14. Schwartz JL, Weiner DH, Lai JJ, et al. Contraceptive efficacy, safety, fit, and acceptability of a single-size diaphragm developed with end-user input. Obstet Gynecol. 2015;125:895–903.

15. Kroll R, Ackerman R, Feldman R, et al. Efficacy and safety of a 21/7-active combined oral contraceptive with continuous low-dose ethinyl estradiol. Contraception. 2016;93:249–56.

16. ClinicalTrials.gov. Transdermal contraceptive delivery system (TCDS), also known as AG200-15 patch arm (ATI-CL13). 2017. https://clinicaltrials.gov/ct2/show/NCT01236768?term=NCT01236768&draw=2&rank=1. Accessed 9 Apr 2020.

17. Kaunitz AM, Portman D, Westhoff CL, et al. Low-dose levonorgestrel and ethinyl estradiol patch and pill: a randomized controlled trial. Obstet Gynecol. 2014;123:295–303.

18. Jensen JT, Garie SG, Trummer D, Elliesen J. Bleeding profile of a flexible extended regimen of ethinylestradiol/drospirenone in US women: an open-label, three-arm, active-controlled, multicenter study. Contraception. 2012;86:110–8.

19. ClinicalTrials.gov. Study of safety and efficacy of an oral contraceptive. 2013. https://clinicaltrials.gov/ct2/show/NCT00477633?term=NCT00477633&draw=2&rank=1. Accessed 9 Apr 2020.

20. Chappell BT, Culwell K, Dart C, Howard B. Perfect-use pregnancy rates with the vaginal pH regulator: efficacy results from AMPOWER [abstract]. Obstet Gynecol. 2020;135:995.

21. Trussell J, Portman D. The creeping pearl: why has the rate of contraceptive failure increased in clinical trials of combined hormonal contraceptive pills? Contraception. 2013;88:604–10.

22. Lobo Abascal P, Luzar-Stiffler V, Giljanovic S, Howard B, Weiss H, Trussell J. Differences in reporting Pearl Indices in the United States and Europe: focus on a 91-day extended-regimen combined oral contraceptive with low-dose ethinyl estradiol supplementation. Eur J Contracept Reprod Health Care. 2016;21:88–91.

23. Food and Drug Administration. Establishing effectiveness and safety for hormonal drug products intended to prevent pregnancy: guidance for industry. 2019. https://www.fda.gov/media/128792/download. Accessed 18 Nov 2019.

24. Frohwirth L, Blades N, Moore AM, Wurtz H. The complexity of multiple contraceptive method use and the anxiety that informs it: implications for theory and practice. Arch Sex Behav. 2016;45:2123–35.