Published in final edited form as:

Fertil Steril. 2014 May ; 101(5): 1205–1208. doi:10.1016/j.fertnstert.2014.03.026.

Live Birth is the Correct Outcome for Clinical Trials Evaluating Therapy for the Infertile Couple

Kurt T. Barnhart, MD, MSCE

Department of Obstetrics and Gynecology, Division of Reproductive Endocrinology and Infertility, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA

Abstract

Well-designed and conducted clinical trials are needed to further advance the field for reproductive medicine. However current reporting of outcomes of trials is ambiguous and disparate. In this manuscript it is offered that the preferred outcome for clinical trials in reproductive medicine should be live birth. Multiple births should be listed and it should be specified whether this is multiple births per couple or multiple births per conception. The unit of measure should be women (or couples) and not cycles. The duration of exposure should also be clearly identified (i.e., treatment was one cycle, a pre-specified number of cycles, or a period of time). Pregnancy loss should be specified and the denominator should be those who conceived. While live birth is the primary outcome, complications should be defined and reported including multiple births and other objective markers such as preterm delivery, small-for-gestational age, or stillbirth.

The goal of clinical research is to inform the science and to aid clinical medicine. Reproductive medicine has enjoyed tremendous growth both scientifically and clinically in the past decades. To further inform clinical care and to practice evidence based medicine, there have been appropriate and widespread calls for well conducted clinical trials in infertility (1, 2). However clinical trials in infertility are challenging to conduct and reporting has been incomplete and inconsistent (1–3). In the attempt to improve transparency and impact of clinical trials, the Consolidated Standards of Reporting Trials (CONSORT) (4) statement has been reviewed by experts in reproductive medicine. Consensus was sought and obtained to provide guidance for the specificity needed for reporting outcomes of clinical. One of the main deficiencies noted by this group was the lack of consistency in the reporting of the primary outcome trials designed to improve fertility.

The goal of therapy in clinical reproductive medicine is to assist couples in starting or extending their families. How could this outcome be ambiguous? It should be relatively easy

© 2014 American Society for Reproductive Medicine. Published by Elsevier Inc. All rights reserved.

Kurt T. Barnhart, MD, MSCE, Department of Obstetrics and Gynecology, Division of Reproductive Endocrinology and Infertility, 3701 Market Street, Suite 800, Philadelphia, PA 19104, 215-662-2974, kbarnhart@obgyn.upenn.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Kurt T. Barnhart has nothing to disclose.

to just count our successes and report them. Shouldn't this problem be as simple as counting the number of children in a family? It is obvious if a childless family now has a child. It should also be obvious if a family has grown from one child to two, or more. The first stumbling block in this simple strategy is that not every pregnancy results in a single birth. Of course infertility trials must account for multiple births. However, the problem goes beyond multiple births, as our literature is replete with disparate and confusing outcomes. Reported outcomes include stimulation parameters, number and quality of gametes (eggs and sperm), embryo survival rate, implantation rate, chemical pregnancy, clinical pregnancy, ongoing pregnancy, and live birth. At times, pregnancy is further subdivided into twins, triplets, higher order multiple gestations, vanishing twins, and vanishing triplets.

Confounding this difficulty is that some of these commonly reported terms are not uniformly defined. For example, is a chemical pregnancy the earliest form of pregnancy (and thus a positive outcome), or is it the earliest form of a miscarriage (and thus a poor outcome)? Success with IVF is often reported as the number of women with positive pregnancy test. However, no couple is happy with the outcome of a chemical pregnancy (a pregnancy loss). Clinical pregnancy is often defined as a gestational sac identified with ultrasound, but when that sac is identified is not uniform. It can be as early as six weeks or perhaps as late as 12 weeks. Thus, a clinical pregnancy in one trial may be a chemical pregnancy in another. The term "ongoing pregnancy" which is meant to suggest that there is a very high likelihood of a pregnancy continuing to term is equally ill defined. At times, ongoing pregnancy is classified as pregnancy that has fetal cardiac activity at 8 weeks, 10 weeks, 12 weeks, or often it is not specified.

Live birth, the preferred primary outcome, is used in some fertility trials. However, even then the definition of a live birth is not without controversy. The SART database defines live-birth delivery as birth of one or more live-born infants (with no specify of gestational age), with delivery of multiple infants counted as one live-birth delivery. A multiple birth is defined as a birth of two or more infants, at least one of whom was live-born. By contrast, CDC's National Center for Health Statistics (NCHS), which bases its statistics on live birth records rather than delivery records, classifies the delivery of a single live-born infant with one or more stillbirths as a singleton birth (5). How far does a gestation need to progress in order to be considered a live birth? For example, if a fetus is born at 19 weeks, with cardiac activity and respiratory effort, is it a live birth? Some have suggested that the definition of a live birth should only be in what is considered a viable gestational age such as 23 or 24 weeks. We propose a definition of a live birth to include a gestational age of 20 weeks, reflecting the World Health Organization definition (6).

Why is live birth not universally accepted as the primary outcome in infertility trials? Many excuses have been proffered including that data are not easy to get because of the fragmentation of reproductive and obstetrical care. Once a clinical or ongoing pregnancy is identified, the mother is referred to another practice—an obstetrician, a midwife, or at times even a completely different institution possibly in a different state or region. Therefore, collection of these data to a clinical investigator in reproductive medicine is a "burden." The excuse should be dismissed simply as lazy. It is understood that the conduct of a clinical trial is expensive and burdensome. However, the small incremental cost is necessary to obtain the

appropriate outcome. If one is to conduct appropriate high quality clinical research, then the cost to obtain information on the circumstance of birth after intervention is a necessity and not a luxury. Randomized clinical research should not be performed on the cheap.

Some have suggested that information gleaned from the clinical trial is so time-sensitive that one cannot wait the additional seven months necessary to find out if a pregnancy conceived results in a live birth. Clearly, if the results are meaningful they are worth the wait. It is also important to understand and report the perinatal outcomes experienced by mother and child (7).

It is possible that pregnancy and live birth as endpoints in a clinical trial are “comparable.” This was objectively assessed by Clarke et al (8) who noted that only 20% of 654 randomized clinical trials reported both live birth and clinical pregnancy as an outcome. The loss from clinical pregnancy to live birth was approximately 19%. The differential absolute loss in those with an active therapy compared to those who conceived without medical intervention (controls) was similar at approximately a 5.4% and 5.5% loss rate (30.3% versus 24.9% in those with treatment, and 27.9% versus 22.4% in controls). Therefore, one possible conclusion is that the 19% pregnancy loss can simply be extrapolated from the clinical pregnancy to achieve a reasonable proximity of the live birth (8). Thus, clinical pregnancy can be a surrogate marker for live birth. However, it is possible that an intervention may have a differential effect on miscarriage and/or survival of a pregnancy compared to an unassisted pregnancy. This can only be noted if only clinical pregnancy and live births are reported. One needs a very strong rationale to accept a surrogate endpoint when the true clinical endpoint can easily be obtained (9, 10).

There are many examples in medicine where surrogate endpoints have misled. For example, some medications decreased arrhythmias but paradoxically increased the risk of death from other causes (11). Fluorides increases bone mineral density in women with osteoporosis but lead to more fractures (12). This important phenomenon is pertinent in reproductive medicine as well. For example, live birth is higher with fresh compared to frozen transfer (13), but perinatal outcomes for children appear to be worse when a child is conceived with fresh transfer (compared to a frozen thawed transfer) (14–16).

The reporting of standardized secondary outcomes is also important in infertility trials. If a conception does not result in a live birth, the outcome or that pregnancy should be reported. Timing of the loss is important and should be reported. Is the pregnancy loss in the first trimester, the second trimester, or did it result in a stillbirth? If pregnancy is never visualized, there are consensus documents which can be utilized to classify the ultimate clinical outcome which may include a visualized or non-visualized ectopic pregnancy, a histologic intrauterine pregnancy, or a resolved or treated pregnancy of unknown location (17). The outcome of all gestations should be reported including spontaneous and active fetal reduction.

Another potential difficulty in reporting outcomes in fertility trials is the unit of measurement. If the unit randomized is something other than the woman (or the couple), results can be misleading. Randomization of eggs, embryos, or cycles can result in a unit of

analysis error (18). Eggs or embryos from the same women are interrelated and when combined with those randomized from other women, the premise of independence necessary for statistical analysis can be violated. Second, multiple observations per woman can lead to an unpredictable bias in the estimate of a treatment difference. It will also exaggerate the apparent sample size giving more precision to an outcome that may be inappropriate. Therefore, papers will have a spurious narrowing of confidence intervals and lower p values, potentially resulting in a Type I statistical error (where an apparent statistical association is noted when in truth none occurs). Many reported trials had a “unit of analysis” error (18). Sometimes the actual endpoint is unclear, and at times pregnancy is not even reported (1). This can result in misinformation, and worse, conclusions that some therapies are objectively better than others when this is simply not the case.

It has been suggested that the desired outcome should be healthy child. Of course the health of the child conceived is important.. There is also great interest in understanding any potential association between any morbidity associated with assisted reproductive technologies or IVF (7). However, using the definition of a healthy pregnancy as a primary outcome can be problematic for a number of reasons. First, because morbidity can only occur if a child is conceived, and (hopefully) only in a subset of those children, a “healthy” child will be less frequent than live birth. A less frequent outcome will require a large sample size. Therefore, most randomized clinical trials appropriately designed to assess a defined prior difference in live birth would be underpowered for assessment of a healthy birth as an outcome. One way to overcome the loss of power is to increase the frequency of outcome by using a composite outcome. Some have reported healthy live birth versus non-healthy combining small-for-gestational age, preterm, or stillbirth (19). However, composite outcomes have their own limitations especially when the biological pathways may not be similar resulting in arbitrary lumping or splitting.

A second issue with an outcome of a healthy child is the definition of healthy. The WHO defines health as “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity” (20). How do we define health in a newborn? Is it the absence of demonstrable medical diagnosis, or do children have to be the top percentage in terms of childhood development, intellect, and accomplishment? To be considered healthy, does the live birth have to be a singleton? Does health include the absence of a cesarean section, not admitted to the NICU, and/or free of congenital anomaly (or only major congenital anomaly)? A restrictive definition would suggest that any multiple birth or child delivered preterm can never be considered healthy.

While accepting that we cannot come to consensus regarding the definition of delivery of a healthy child, evaluating outcomes beyond live birth is very important. For example, comparing “good perinatal outcome” (defined as live birth of a singleton neonate born at 37 or more completed weeks of gestation weighing 2500 grams or more) for children conceived with IVF in 2000 and 2008 is revealing (21). This analysis demonstrated that the proportion of good perinatal outcomes among live born neonates conceived with IVF in 2008 was 42.5%. This is not a typo; fewer than half of the live births conceived with IVF are considered of “good perinatal outcome” based on this definition. This is driven mostly by

multiple births. The proportion of good perinatal outcomes in singletons is 83.4 in 2008 (21).

The preferred way to report results for a trial in infertility is evaluate time to pregnancy (live birth) with the use of Life Table Analysis. This allows objective demonstration of absolute differences in pregnancy (live birth) rates and how those differences evolve over time using the correct unit of analysis of women (or couple). This will inform the clinical question of not only the efficacy, but the time needed to achieve a live birth with different therapies. Such examples of proper reporting include demonstration of clomiphene citrate compared to metformin in the treatment of women with Polycystic Ovary Syndrome (22). Time to pregnancy may not be applicable to a study comparing two techniques for women undergoing a single IVF cycle. However, it should be understood that a comparison of single IVF cycles does not address the question of whether IVF is the appropriate therapy for infertility. Such a study will not determine the optimal way to achieve live birth because it does not take into account therapy after the single cycle. When reporting success with a single IVF cycle, cumulative live birth (including fresh and all frozen cycles) should be reported.

In summary, the outcome of a clinical trial should be specified in advance. Live birth is the preferred outcome. Multiple births should be listed and it should be specified whether this is multiple births per couple or multiple births per conception. The unit of measure should be women (or couples) and not cycles, gametes or embryos. The duration of exposure should also be clearly identified (i.e., treatment was one cycle, a pre-specified number of cycles, or a period of time). Pregnancy loss should be specified and the denominator should be those who conceived. While live birth is the primary outcome, complications should be defined and reported including multiple births and other objective markers such as preterm delivery, small-for-gestational age, or stillbirth. Maternal complications including gestational diabetes, preeclampsia, abruption and C-section should also be defined and reported. Reproductive medicine, especially IVF, is a complex field that requires special consideration in study design. Consultation with a well versed methodologist or statistician, before the trial is initiated, will reduce ambiguity and greatly improve the quality of our research findings.

References

1. Dapuzzo L, Seitz FE, Dodson WC, Stetter C, Kunselman AR, Legro RS. Incomplete and inconsistent reporting of maternal and fetal outcomes in infertility treatment trials. *Fertil Steril.* 2011; 95:2527–2530. [PubMed: 21435640]
2. Johnson NP, Proctor M, Farquhar CM. Gaps in the evidence for fertility treatment-an analysis of the Cochrane Menstrual Disorders and Subfertility Groups database. *Hum Reprod.* 2003; 18:947–954. [PubMed: 12721167]
3. Schlaff WD. Barriers to conducting clinical research in reproductive medicine around the world. *Fertil Steril.* 2011; 96:801. [PubMed: 21961911]
4. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomized trials. *Obstet Gynecol.* 2010; 115:1063–1070. [PubMed: 20410783]
5. Martin JA, Hamilton BE, Sutton PD, et al. Births: final data for 2009. *National Vital Stat Rep.* 2009; 57:1–104.

6. World Health Organization. International statistical classification of diseases and related health problems. Geneva, Switzerland: World Health Organization; 1993. p. 129
7. Barnhart KT. Assisted reproductive technologies and perinatal morbidity: Interrogating the association. *Fertil Steril*. 2013; 99(2):299–302. [PubMed: 23375143]
8. Clarke JF, van Rumste MM, Farquhar CM, Johnson NP, Mol BW, Herbison P. Measuring outcomes in fertility trials: Can we rely on clinical pregnancy rates? *Fertil Steril*. 2010; 94:1647–1651. [PubMed: 20056216]
9. Grimes DA, Schulz KF, Raymond EG. Surrogate end points in women's health research: science, protoscience, and pseudoscience. *Fertil Steril*. 2010; 93(6):1731–1734. [PubMed: 20153470]
10. Grimes DA, Schulz KF. Surrogate end points in clinical research: hazardous to your health. *Obstet Gynecol*. 2005; 105:1114–1118. [PubMed: 15863552]
11. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med*. 1991; 324:781–788. [PubMed: 1900101]
12. Riggs BL, Hodgson SF, O'Fallon WM, Chao EY, Wahner HW, Muhs JM, et al. Effect of fluoride treatment on the fracture rate in postmenopausal women with osteoporosis. *N Engl J Med*. 1990; 322:8.
13. Luke B, Brown M, Wantman E, Liderman A, Gibbons W, Schattman G, Lobo R, Leach R, Stern JE. Cumulative Birth Rates with Linked Assisted Reproductive Technology Cycles. *N Engl J Med*. 2012; 366:2483–2491. [PubMed: 22738098]
14. Maheshwari A, Pandey S, Shetty A, Hamilton M, Bhattacharya S. Obstetric and perinatal outcomes in singleton pregnancies resulting from the transfer of frozen thawed versus fresh embryos generated through in vitro fertilization treatment: a systematic review and meta-analysis. *Fertil Steril*. 2012; 98(2):368–377. [PubMed: 22698643]
15. Kansal Kalra S, Ratcliffe SJ, Coutifaris C, Molinaro T, Barnhart KT. Ovarian stimulation and low birth weight in newborns conceived through in vitro fertilization. *Obstet Gynecol*. 2011; 118(4): 863–871. [PubMed: 21934450]
16. Kansal Kalra S, Ratcliffe SJ, Milman L, Gracia CR, Coutifaris C, Barnhart KT. Perinatal morbidity after in vitro fertilization is lower with frozen embryo transfer. *Fertil Steril*. 2011; 95:548–553. [PubMed: 20663500]
17. Barnhart K, van Mello NM, Bourne T, Kirk E, Van Calster B, Bottomley C, Chung K, Condous G, Goldstein S, Hajenius PJ, Mol BW, Molinaro T, O'Flynn O'Brien KL, Husicka R, Sammel M, Timmerman D. Pregnancy of unknown location: A consensus statement of nomenclature, definitions, and outcome. *Fertil Steril*. 2011; 95(3):857–866. [PubMed: 20947073]
18. Vale A, Gardner E. Common statistical errors in the design and analysis of subfertility trials. *Hum Reprod*. 2003; 18:1000–1004. [PubMed: 12721176]
19. Chung K, Coutifaris C, Chalian R, Lin K, Ratcliffe SJ, Castelbaum AJ, Freedman MF, Barnhart KT. Factors influencing adverse perinatal outcomes in pregnancies achieved through use of in vitro fertilization. *Fertil Steril*. 2006; 86(6):1634–1641. [PubMed: 17074345]
20. Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference; Official Records of the World Health Organization, no. 2; 19–22 June; New York. 1946. p. 100
21. Joshi N, Kissin D, Anderson JE, Session D, Macaluso M, Jamieson D. Trends in correlative good perinatal outcomes in assisted reproductive technology. *Obstet Gynecol*. 2012; 120:843–851. [PubMed: 22996102]
22. Legro RS, Barnhart HX, Schlaff WD, Carr BR, Diamond MP, Carson SA, Steinkampf MP, Coutifaris C, McGovern PG, Cataldo NA, Gosman GG, Nestler JE, Giudice LC, Leppert PC, Myers ER. Cooperative Multicenter Reproductive Medicine Network. Clomiphene, metformin, or both for infertility in the polycystic ovary syndrome. *New England Journal of Medicine*. 2007; 356(6):551–566. [PubMed: 17287476]