

Epidemiologic Approaches for Studying Assisted Reproductive Technologies: Design, Methods, Analysis, and Interpretation

Carmen Messerlian¹ · Audrey J. Gaskins^{2,3}

Published online: 17 April 2017
© Springer International Publishing AG 2017

Abstract

Purpose While considerable progress has been made since the advent of assisted reproductive technology (ART), the field remains a complex and challenging one for clinicians and researchers alike. This review discusses some of the most salient issues pertaining to the study of ART and whenever possible suggestions on how to address them.

Recent Findings More than 5 million babies have been born through ART to date, representing up to 4% of all births worldwide. While technologies continue to evolve and demand for treatment grows, it is more important than ever to conduct rigorous and timely research to help guide clinical practice that is safe and effective, and that minimizes potential short- and long-term adverse outcomes to mother and child.

Summary ART research will require exceedingly more sophisticated research methods, designs, and analyses that are rooted in a reproductive epidemiological framework in order to improve future research and ultimately promote better outcomes for all subfertile couples and their children.

Keywords Assisted reproductive technology · Infertility · Research methods · Design · Analysis

This article is part of the Topical Collection on *Reproductive and Perinatal Epidemiology*

✉ Audrey J. Gaskins
agaskins@hsph.harvard.edu

¹ Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA

² Department of Nutrition, Harvard T.H. Chan School of Public Health, 655 Huntington Ave, Building II 3rd Floor, Boston, MA 02115, USA

³ Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

Introduction

Approximately 5 million babies have been born through assisted reproductive technology (ART) since 1978, representing between 1 and 4% of all births among countries worldwide [1–3]. During this period, the number of ART cycles steadily increased as demand for treatment surged [4]. Globally, among countries contributing data, more than 4 million initiated cycles were reported in 2008, 2009, and 2010, resulting in the birth of over 1,144,800 ART-conceived babies during this 3-year period alone [5].

Over the last several decades, ART research has largely focused on understanding patient and treatment factors that increase the likelihood of a live birth and improving outcomes for mothers—through reduced complications—and their ART-conceived children. Great strides have been made to improve clinical protocols and develop new technologies that maximize the success of ART and minimize the risk to the mother and child, including the adoption of single-embryo transfer policies and the improvement of embryo culture media [6, 7•]. Indeed ART success rates have improved over time as witnessed by the rising pregnancy and live birth rates, and the overall declining number of multiple births [8]. While the advent of ART has had a tremendous impact on subfertile couples globally, an unintended consequence associated with its use has been a higher risk of complications in pregnancy and greater risk of preterm birth and other adverse outcomes in both mothers and children [9•].

The study of ART remains a complex and challenging area of epidemiologic research. The purpose of this review is to discuss the most relevant issues pertaining to this field of study in the broader context of understanding the factors that predict clinical success, and investigating the health impacts of these technologies on a potentially high-risk population and their offspring. When possible, we also aim to provide

suggestions on how to minimize or circumvent these issues moving forward in order to improve the study of ART.

Definitions, Study Populations, and Data Sources

Definitions The study of ART is hindered by a lack of an accepted and uniform definition of infertility itself [10, 11]. Infertility is a heterogeneous condition with potentially diverse underlying pathologies that occurs across a continuum of severity [12]. It is also couple specific and the involvement of two individuals rather than one further complicates the definition. Variability in the use of terms such as “subfertility,” “infertility,” and “subfecundity” has led to confusion and discrepancies in the study of infertility and as a consequence studies on ART [10]. The International Committee for Monitoring Assisted Reproductive Technology and the World Health Organization (WHO) characterize infertility as “a disease of the reproductive system defined by the failure to achieve a clinical pregnancy after 12 months or more of regular unprotected sexual intercourse” [13]. Definitions of infertility have centered on a couples’ duration of trying time to pregnancy and maternal age, both considered important prognostic factors [7••]. Severity, on the other hand, is determined by the underlying conditions associated with a long time to pregnancy, and the probability of conceiving with different forms of treatment depends on the cause of infertility, maternal age, and other risk factors [14].

One particular concern is that a lack of consensus on a definition and diagnostic criteria for infertility influences who seeks evaluation and obtains treatment for infertility [7••, 10, 15]. This issue, along with barriers to accessing treatment including income, education, and race, as well as patient dropouts impact who makes up the study population [16••]. The resultant heterogeneity across clinics in patient populations and risk factors may limit valid comparisons across studies. Consistency in defining infertility and setting more uniform criteria for the purpose of clinical management strengthens our ability to identify couples that benefit from treatment [15]. Doing so also helps advance research that relies on such clinical settings for participants and remedies a significant obstacle to meaningful comparison. Thus, a shared and unified definition of infertility remains a primary and necessary goal of studying ART (Table 1).

Study Populations A related concern has been the restricted framework by which ART is studied. ART refers to all treatments and procedures whereby *both* gametes, or embryos, are manipulated in vitro. These procedures include, among others, in vitro fertilization (IVF), intracytoplasmic sperm injection, frozen embryo transfers, gamete and embryo cryopreservation, gamete or zygote intrafallopian transfers, and tubal transfers (see WHO glossary) [13]. It is estimated that ART-based

procedures account for only about 20% of all pregnancies achieved in the subfertile population [7••]. However, ART does *not* include intrauterine insemination or ovulation induction/ovarian stimulation, procedures which are much more common than ART. In fact, it has been estimated that ovulation induction alone accounts for two to six times more births than ART in the USA (~191,000 per year) [17], making the broader term, *medically assisted reproduction* (MAR), a more relevant area of study. There is a growing body of literature pointing to other non-ART treatments also being associated with adverse pregnancy outcomes [18]. Moreover, up to 40% of couples who seek infertility treatment end up conceiving naturally either before treatment begins, in between cycles, or after withdrawing from treatment [11, 19]. As such, limiting the study population to only those who conceive via ART censors a large proportion of infertile patients and limits our ability to accurately and validly quantify the contribution of ART treatment and underlying infertility to various clinical, pregnancy, and child health outcomes of interest.

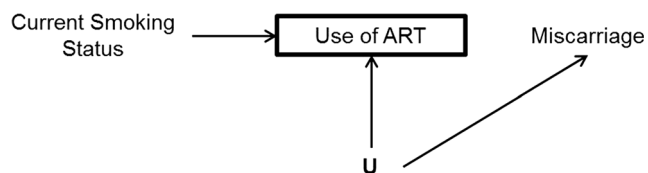
Access to and use of infertility treatments is also determined by a multitude of socio-economic factors including education, household income, and private health insurance among women [20] and men [16••, 21]. It has also been shown that women adhering to healthy lifestyle habits, including those with a lower BMI, never smokers, multivitamin users, and who are physically active, are more likely to have an infertility evaluation compared to their peers who adopted less healthy behaviors [16••]. Thus, associations found between specific risk factors or lifestyle/environmental exposures and reproductive outcomes in ART studies should be interpreted carefully as results may not be generalizable to the broader population of couples experiencing infertility. In addition, depending on the causal question of interest, using an ART population could result in selection bias. For example, if current smokers are less likely to seek infertility treatment and there is another factor (U) associated with ART utilization and the likelihood of miscarriage, then studying the effect of smoking on miscarriage in an ART population may result in a biased effect estimate for the causal question of interest (Fig. 1). In general, causal diagrams, like the one illustrated in Fig. 1, can help identify scenarios in which ART analyses may be biased, and if necessary, inverse probability weights can be used to remove such bias, as described by Ahrens and colleagues [22•]. Heterogeneity in patient populations, along with known differences in treatment protocols across clinics, further hampers the ability to compare ART study results across studies. Thus, great care should be taken when comparing results across ART publications and conducting meta-analyses.

Data Sources Many studies from the USA, Canada, and Europe rely on national-level surveillance systems that monitor and report ART cycles performed by clinics. In the USA, the Centers for Disease Control (CDC) obtains data on all

Table 1 Key methodological issues in the design and analysis of ART studies

Area of concern	Key methodological issues
Definitions and study population	<ul style="list-style-type: none"> • Inconsistent definitions of infertility and criteria for treatment. • Narrow framework on ART that excludes other forms of medically assisted reproduction. • Women/men/couples who participate in ART studies may be different from those who do not. • Heterogeneity across clinics and risk factors in patient populations.
Data sources	<ul style="list-style-type: none"> • Data sources influenced by patient selection factors, treatment policies, and reporting practices. • Comparing across ART studies is limited by clinic specific populations and protocols.
Exposure assessment	<ul style="list-style-type: none"> • Complexity in ART as an exposure: varying treatment types, clinical guidelines, governmental/insurance policies, and differences in risk of adverse outcomes. • Choice of comparison group on studies of adverse outcomes.
Confounding	<ul style="list-style-type: none"> • How to account for underlying infertility and its severity. • Measuring and accounting for shared common causes of environmental factors and ART outcomes.
Study outcome(s)	<ul style="list-style-type: none"> • Focusing on surrogate endpoints instead of live birth as the key outcome. • Using a rate as an outcome metric.
Statistical analysis	<ul style="list-style-type: none"> • Restricting analysis to only cycles with embryo transfer. • Non-independence of oocytes, embryos, and ART outcomes within a woman undergoing multiple cycles. • Using odds ratios with common outcomes. • Relying on <i>p</i> values.

ART procedures as part of the National ART Surveillance System and can be linked with other state-level health outcomes [23]. Similarly, the Canadian ART Register (CARTR), established in 1999, has achieved 97% participation by voluntary fertility clinic reporting and can be linked with additional data elements extending its capacity to track individual outcomes [24]. The European Society of Human Reproduction and Embryology (ESHRE) has monitored ART since 1997, and currently, more than 33 countries provide data to this well-established source [25]. Denmark, however, is the only country in Europe with an established national registry of their own, and has the significant advantage of including all MAR procedures [26]. Other such surveillance programs are limited to monitoring and reporting solely on ART cycle procedures, a significant hindrance to evaluating the risk of adverse outcomes by different treatment modalities of conception. Furthermore, while surveillance reports provide ART researchers with data on cycles in the tens of thousands—enhancing power and allowing one to investigate a multitude of questions—these data sources are not without limitations. Differences in patient selection factors, treatment policies (e.g., single- vs. multiple-embryo transfer), cycle reporting practices, and undocumented clinical practices, may partially explain the different outcomes observed across clinics, states, and countries rather than true differences.

**Fig. 1** The potential for selection bias to affect studies on environmental exposures and ART outcomes

Evaluating ART as the Exposure

Within the subfertile population, the evaluation of various ART procedures as interventions by measuring the proportion of treated cycles ending in favorable outcomes has helped advance clinical protocols, reduce risks, and improve success rates [27]. However, success of ART has been differentially defined over time and has depended on many factors, including availability of data, reporting practices, and clinical guidelines [7••]. While the ultimate goal of couples seeking fertility treatment is the birth of a healthy, live born infant [28], the preponderance of studies have historically examined pregnancy rates as a method of assessing successes and failures [7••]. Yet, multiple vs. single-embryo transfer policies [29], for example, can impact implantation and pregnancy rates while doing little to achieve the desired outcome of a healthy live birth through a low-risk pregnancy [30]. Nevertheless, the complexities in different ART treatment practices make assessment of interventions and their outcomes difficult. General consensus by the ART community on what constitutes a successful and healthy pregnancy and birth along with adherence to clinical guidelines and more rigorous research assessing standardized treatment would help improve the evaluation of ART and generate a stronger evidence base for clinical practice.

In 1980, the pioneer of IVF—Dr. Robert Edwards—was the first to report that IVF-conceived pregnancies could result in problematic outcomes for the mother or infant [31]. Ever since, there has been an explosion of ART studies investigating the short- and long-term risks and consequences associated with IVF-conceived pregnancies and births compared to those naturally conceived [9•]. While these studies have helped shape our understanding of the impact of ART on

adverse pregnancy and child health outcomes, they have a number of limitations. First, whether we measure ART exposure as current (i.e., cycle specific to conception ART procedures), recent (i.e., all ART procedures in the last several cycles prior to outcome) or lifetime (i.e., any ART procedures used prior to outcome) has important implications with respect to assessing the risk of adverse pregnancy, child, and long-term health outcomes. Second, when designing studies to quantify the risk of ART on poor pregnancy and birth outcomes, care must be taken when choosing the appropriate comparison group to ensure that every level of exposure (including the unexposed group) is represented across the study cohort in order to minimize bias and confounding. For instance, while sampling from the general clinical obstetric population from the hospital from which the infertility patients arose is often the most readily available and straightforward comparison group, it can also lead to a type of Berkson's bias if the hospital is a tertiary-care center that overselects high-risk pregnancies from the population. Furthermore, use of this comparison group does not allow researchers to directly separate the effects of ART treatment from underlying infertility and its severity. As such, many researchers have designed studies to examine the effects of assisted reproduction by using non-medically assisted/naturally conceived cycles within a fertility clinic as the comparison group [9, 18]. This approach has the advantage of being able to compare differences in risk of outcomes by treatment mode of conception among couples with some degree of subfertility; however, similar to above, the differences in outcomes observed could be confounded by type and severity of infertility [9]. Another method used to address the potential issues involved in choosing a comparison group has centered on the use of siblings conceived with discordant use of ART treatment [32]. This method has the unique advantage of accounting for all the within-woman risk factors that may have remained constant between pregnancies (e.g., BMI, socioeconomic status, education underlying infertility, general health); however, it can only partially control for time-varying factors such as age, parity, potential differences in partners, and changes in habits and lifestyle over time. While there are advantages and disadvantages to each design, expanding data sources to include all MAR, and possibly even the causes of infertility, would allow for improved comparison between treatment types, conception modes, and underlying infertility, and potential outcomes.

Measured and Unmeasured Confounding

That underlying infertility, and not just ART treatment, could contribute to complications during pregnancy was first reported by Saunders and colleagues in 1988 [33]. Using Australian IVF registry data, the authors found that couples who conceived naturally while on a waiting list for assessment and

treatment of infertility had a higher risk of preterm birth compared with the general population. Moreover, among singleton births, the risk of preterm birth was comparable to babies conceived through IVF, suggesting an elevated baseline risk among infertile couples, irrespective of treatment. Mechanisms leading to infertility and its severity are thought to play a role in the etiology of adverse pregnancy and child health outcomes, making it a strong confounder and further complicating ART research [9, 19]. Without accounting for underlying infertility and its potential inherent risks, it is possible to misestimate the effect of ART or other procedures on the success rates of treatments as well as on the risk of adverse outcomes. Furthermore, not adjusting for other known confounding factors such as age and parity, the two variables most strongly related to fertility, is common among studies using data sources and linkages with limited individual-level variables [34, 35]. Unmeasured confounding is also a concern as often there are shared common causes of infertility (and therefore ART) with outcomes such as preterm birth, low birth weight, and other unintended birth outcomes. One example is paternal factors, which are rarely measured in naturally conceived pregnancies, but have been associated with reduced fecundability (and thereby ART) and adverse outcomes [36]. One invaluable benefit of the ART clinical setting is that it is uniquely designed to examine new and emerging risk factors including the role that maternal and paternal environmental chemicals and lifestyle factors play in ART treatment success, and pregnancy and child health outcomes [37–41].

Studying the environmental, dietary, paternal, and clinical factors that predict ART success is of great interest to current and future infertility patients and health care providers [42]. Studying the effects of environmental exposures such as phthalates and other endocrine disruptors on fertility and pregnancy outcomes in this unique setting allows for the direct observation of many previously unobservable outcomes, including oocyte quality, fertilization rate, and embryo development. ART populations are also distinct in that couples presenting for treatment suffer from some degree of subfertility, and represent a potentially more susceptible high-risk population where effects of certain environmental exposures and diets may be stronger. Despite the advantages of using this setting to answer new research questions, such studies are not without challenges including recruiting a representative study population, defining the appropriate outcomes, properly analyzing the data, and interpreting the results in light of potential limitations to generalizability.

ART Study Outcome(s)

The choice of a primary outcome in studies investigating predictors of ART success has implications for both the relevance and methodological validity of a study's results.

In the ART setting, the majority of studies fail to report on the most relevant outcome, live birth. The main reason for this is due to the natural time lag between infertility treatment and the birth of an infant, which may result in loss to follow-up, because obstetrical and infant care are delivered by other providers. Thus, instead of focusing on live birth, studies in ART patients tend to focus on surrogate outcomes of varying clinical importance, such as oocyte retrieval parameters, fertilization rate, embryo quality measures, and ongoing pregnancy. While the intermediate outcomes of IVF are more often used as surrogates of a woman's fertility potential, ongoing pregnancy is generally considered an acceptable surrogate for live birth, despite its inherent flaws. The major potential limitation of using ongoing pregnancies as the primary endpoint of ART studies is the significant proportion of pregnancies that are lost between the pregnancy confirmation (generally ~6–8-week gestation) and live birth. Therefore, any exposures or treatments that reduce or increase the risk of pregnancy loss would be shown to have no association with ongoing pregnancy despite having important clinical relevance to patients and health care providers. The further along in gestation ongoing pregnancy can be measured, the better, as the incidence of pregnancy loss declines substantially after 20-week gestation [43]. Therefore, future studies should strive to focus on the most relevant outcome, live birth, as the primary endpoint and when this is not feasible to focus on ongoing pregnancies as far along in gestation as possible.

Another concern is the reliance on rates as the measured outcome evaluated in ART studies, with implantation, clinical pregnancy, and live birth rates—estimated as the number of embryos that resulted in implantation, clinical pregnancy, or live birth divided by the number of embryos transferred—being the most common. Most often rates are used as a way to control or take into account the effect of another factor. For instance, with implantation rates, researchers are often trying to account for the fact that some women may have had more than one embryo transferred and thus have a higher likelihood of implantation. However, division by number of embryos transferred can also create unwanted variation, particularly when the numerator has low variability. Dividing by a denominator with larger variability results in an outcome that is highly related to the factor whose effect we wished to remove, in this case, number of embryos transferred. Furthermore, if the exposure is associated with the denominator of the rate, the exposure will be associated with the rate in the direction opposite to that of the denominator even when the exposure has no association with the numerator. Therefore, caution must be taken when utilizing rates in ART studies as oftentimes the methodological limitations vastly outweigh their potential clinical importance.

Statistical Analysis

The most egregious error in the design or analysis of ART studies is focusing on clinical outcomes only among cycles with embryo transfer. This can occur by design (e.g., only including cycles undergoing embryo transfer in the study) or a post hoc decision to exclude these cycles in the analysis. This restriction also occurs, although not as transparently, when studies utilize a rate with number of embryos transferred as the denominator. Limiting the analysis to only cycles with embryos transferred can induce a type of survival bias, where only those cycles that have succeeded to the point of embryo transfer are included in the analysis. If the exposure of interest is associated with early ART failure such as poor response to stimulation, no oocytes fertilized, and poor embryo development then the exclusion of these cycles will likely lead to bias. Due to the large number of cycles that fail prior to embryo transfer (10–20%), the potential effect of this type of bias is not negligible. In the most extreme example, imagine that exposure to a certain chemical X has such a strong effect on oocyte quality that 60% of cycles for women exposed to chemical X fail during fertilization and there are no embryos available for transfer (Table 2). If these cycles are excluded and there is no further effect of chemical X on later ART outcomes, the analysis would show that there is no association between chemical X and percent of embryo transfer cycles resulting in live birth (scenario A). However, if there is a depletion of susceptibles and the surviving women/cycles who were exposed to chemical X actually have better success after embryo transfer compared to unexposed women, then the effect of chemical X and live birth per embryo transfer could potentially appear beneficial (scenario B). In either scenario, if all initiated cycles were included, and the data were analyzed as intention to treat (similar to the approach used in the analysis of clinical trials), the results would then demonstrate that chemical X had a negative impact on the outcome. The main criticism of the intention to treat analysis, however, is that only cases at risk of the outcome should be included in the denominator when addressing questions of the biologic or mechanistic effects of an exposure [44]. Thus, moving forward, more sophisticated statistical techniques such as discrete survival models where the probability of failing at each ART stage is modeled conditional on succeeding in the previous stage should be more widely considered in ART studies [45].

Another area of statistical concern in ART studies is “unit of analysis” errors. Simple group comparison tests, such as the *t* test for continuous data or χ^2 for categorical data, require that observations are statistically independent. Similarly, in linear or logistic regression models, each line of data is considered statistically independent. In the context of ART studies, to fulfill these independence assumptions, only one observation per patient could be included in any such analysis. However, the hierarchical nature of ART data which often contains

Table 2 Example of survival bias in ART studies by restricting to cycles with embryo transfer (ET)

	Exposure to chemical X (<i>n</i> = 100 cycles)	No exposure (<i>n</i> = 100 cycles)
Scenario A		
Cycles with ET	40	90
Cycles with live birth	20	45
Percent of ET cycles resulting in live birth	50%	50%
Percent of initiated cycles resulting in live birth	20%	45%
Scenario B		
Cycles with ET	40	90
Cycles with live birth	30	45
Percent of ET cycles resulting in live birth	75%	50%
Percent of initiated cycles resulting in live birth	30%	45%

multiple oocytes, embryos, and implantations per treatment cycle and multiple treatment cycles per woman provides extended scope for “unit of analysis” errors [46]. Not accounting for multiple observations per woman leads to unpredictable bias in the estimate of the effect, but exaggerates the apparent sample size. This exaggeration leads to spuriously narrow confidence intervals and low *p* values. The most common methods to account for multiple observations per cycle or per woman are generalized estimating equations or mixed models. While the first option is generally easier to fit, it requires the assumption that any imbalance in the number of observations per cycle or per woman is not related to the outcome of interest. This is oftentimes hard to assume in ART studies since the number of embryos a woman produces or the number of ART cycles a woman undergoes is directly related to her underlying fertility, the outcome being studied [47]. Mixed models are preferred over generalized estimating equations because this imbalance in observations is actually modeled, and thus, as long as the entire joint distribution is correctly specified (including the model for the mean response and the within-subject association), no bias should result [48].

Other important statistical issues that plague ART studies are the use of odds ratios for common outcomes and reliance on *p* values which can lead to inappropriate conclusions regarding effect size and significance. In cohort studies and RCTs, odds ratios are often interpreted as risk ratios. This is problematic because an odds ratio always overestimates the risk ratio, and this overestimation becomes larger with increasing incidence of the outcome [49]. In ART studies, most of the outcomes being studied are common. For example, in 2014, 33% of initiated ART cycles using fresh non-donor eggs or embryos in the USA resulted in clinical pregnancy and even more resulted in implantation. This issue also applies to studies of adverse pregnancy outcomes, such as preterm birth, where the percentage of ART singletons born before 37-week completed gestations can exceed 10% and can reach as high as 25%, making it also a common outcome [50].

Fortunately, there are many ways to avoid this overestimation (as summarized by Austin and Laupacis [51]), the simplest being to use log–binomial or Poisson regression with robust standard errors [52]. However, another simple way would be to present predicted probabilities obtained from a logistic regression model (rather than an odds ratio) and, if desired, use these predicted probabilities to estimate risk ratios [53]. Finally, while the utility of the *p* value in biomedical research has been argued for many years [54, 55], a thorough discussion of their use in reproductive epidemiology has recently been described by Farland et al. [56]. The pitfalls of relying on *p* values remain a concern for ART studies. *P* values combine two important pieces of information: the magnitude of the effect size and the sample size. Therefore, scientific conclusions should not be solely based on whether a *p* value passes a specific threshold (e.g., <0.05). Important differences can easily be missed in small ART studies by using a threshold of 0.05 due to low power. On the other hand, statistically significant results based on a threshold of 0.05 may be unimportant if the difference at hand is not clinically meaningful or biologically plausible. A better alternative to merely providing *p* values is using confidence intervals, which provide information on the size and precision of the effect and contribute to the calculation of the *p* value. Moving forward, ART researchers should carefully consider whether using a *p* value adds anything meaningful to the presentation of their results and keep in mind the importance of effect size and biological plausibility when attributing clinical and public health importance to the finding.

Conclusions

Considerable progress has been made over the last 40 years of ART research; however, the field remains a complex and challenging one for clinicians and researchers alike. While technologies continue to evolve and demand for treatment grows,

fertility specialists will increasingly rely on rigorous and timely research to help guide clinical practice that is safe and effective, and that minimizes potential short- and long-term adverse outcomes to the mother and baby. There is also a burgeoning need to better understand how parental environmental and lifestyle factors impact ART success and pregnancy and child health. This will require exceedingly stronger research methods, designs, and analyses that are rooted in a reproductive epidemiological framework. This review discussed some of the most salient issues pertaining to the field and how to address them in order to improve future research and ultimately promote better outcomes for all subfertile couples.

Acknowledgements We thank Dr. Russ Hauser, Harvard T.H. Chan School of Public Health, for his invaluable feedback of this review.

Compliance with Ethical Standards

Conflict of Interest Carmen Messerlian and Audrey J. Gaskins each declare no potential conflicts of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance
- Of major importance

1. Ishihara O, Adamson GD, Dyer S, de Mouzon J, Nygren KG, Sullivan EA, et al. International committee for monitoring assisted reproductive technologies: world report on assisted reproductive technologies, 2007. *Fertil Steril*. 2015;103(2):402–13 e11. doi:10.1016/j.fertnstert.2014.11.004.
2. Zegers-Hochschild F, Mansour R, Ishihara O, Adamson GD, de Mouzon J, Nygren KG, et al. International Committee for Monitoring Assisted Reproductive Technology: world report on assisted reproductive technology, 2005. *Fertil Steril*. 2014;101(2):366–78. doi:10.1016/j.fertnstert.2013.10.005.
3. ESHRE. The world's number of IVF and ICSI babies has now reached a calculated total of 5 million. Istanbul, Turkey: European society of human reproduction and embryology; 2012 [cited 2012 October 16th]; Available from: <http://www.eshre.eu/ESHRE/English/Press-Room/Press-Releases/Press-releases-2012/5-million-babies/page.aspx/1606>. 2012.
4. Connolly MP, Hoorens S, Chambers GM. The costs and consequences of assisted reproductive technology: an economic perspective. *Hum Reprod Update*. 2010;16(6):603–13. doi:10.1093/humupd/dmq013.
5. Dyer S, Chambers GM, de Mouzon J, Nygren KG, Zegers-Hochschild F, Mansour R, et al. International Committee for Monitoring Assisted Reproductive Technologies world report: assisted reproductive technology 2008, 2009 and 2010. *Hum Reprod*. 2016;31(7):1588–609. doi:10.1093/humrep/dew082.
6. Chronopoulou E, Harper JC. IVF culture media: past, present and future. *Hum Reprod Update*. 2015;21(1):39–55. doi:10.1093/humupd/dmu040.
7. ESHRE. Failures (with some successes) of assisted reproduction and gamete donation programs. *Hum Reprod Update*. 2013;19(4):354–65. doi:10.1093/humupd/dmt007. **A comprehensive overview by ESHRE working group on timely issues pertaining to infertility and assisted reproductive technology with important key messages and next steps included in the conclusion.**
8. Toner JP. Progress we can be proud of: U.S. trends in assisted reproduction over the first 20 years. *Fertil Steril*. 2002;78(5):943–50.
9. Pinborg A, Wennerholm UB, Romundstad LB, Loft A, Aittomaki K, Soderstrom-Anttila V, et al. Why do singletons conceived after assisted reproduction technology have adverse perinatal outcome? Systematic review and meta-analysis. *Hum Reprod Update*. 2013;19(2):87–104. doi:10.1093/humupd/dms044. **A complete review and thorough discussion on understanding the complexities of why singletons conceived after assisted reproductive technology have a higher risk of adverse outcomes.**
10. Gurunath S, Pandian Z, Anderson RA, Bhattacharya S. Defining infertility—a systematic review of prevalence studies. *Hum Reprod Update*. 2011;17(5):575–88. doi:10.1093/humupd/dmr015.
11. Gnoth C, Godehardt E, Frank-Herrmann P, Friol K, Tigges J, Freundl G. Definition and prevalence of subfertility and infertility. *Hum Reprod*. 2005;20(5):1144–7.
12. Messerlian C, Maclagan L, Basso O. Infertility and the risk of adverse pregnancy outcomes: a systematic review and meta-analysis. *Hum Reprod*. 2013;28(1):125–37. doi:10.1093/humrep/des347.
13. Zegers-Hochschild F, Adamson GD, de Mouzon J, Ishihara O, Mansour R, Nygren K, et al. The International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) revised glossary on ART terminology, 2009. *Hum Reprod*. 2009;24(11):2683–7. doi:10.1093/humrep/dep343.
14. Allen VM, Wilson RD, Cheung A. Pregnancy outcomes after assisted reproductive technology. *J Obstet Gynaecol Can*. 2006;28(3):220–50.
15. Donckers J, Evers JL, Land JA. The long-term outcome of 946 consecutive couples visiting a fertility clinic in 2001–2003. *Fertil Steril*. 2011;96(1):160–4. doi:10.1016/j.fertnstert.2011.04.019.
16. Farland LV, Collier AR, Correia KF, Grodstein F, Chavarro JE, Rich-Edwards J, et al. Who receives a medical evaluation for infertility in the United States? *Fertil Steril*. 2016;105(5):1274–80. doi:10.1016/j.fertnstert.2015.12.132. **This paper details the demographic, lifestyle, and access barriers associated with seeking infertility treatment in the US and gives readers an idea of how representative infertility treatment cohorts are of women experiencing infertility.**
17. Schieve LA, Devine O, Boyle CA, Petrini JR, Warner L. Estimation of the contribution of non-assisted reproductive technology ovulation stimulation fertility treatments to US singleton and multiple births. *Am J Epidemiol*. 2009;170(11):1396–407. doi:10.1093/aje/kwp281.
18. Messerlian C, Platt RW, Tan SL, Gagnon R, Basso O. Low-technology assisted reproduction and the risk of preterm birth in a hospital-based cohort. *Fertil Steril*. 2015;103(1):81–8 e2. doi:10.1016/j.fertnstert.2014.10.006.
19. Messerlian C, Platt RW, Ata B, Tan SL, Basso O. Do the causes of infertility play a direct role in the aetiology of preterm birth? *Paediatr Perinat Epidemiol*. 2015;29(2):101–12. doi:10.1111/ppe.12174.

20. Chandra A, Copen CE, Stephen EH. Infertility service use in the United States: data from the National Survey of Family Growth, 1982–2010. *Natl Health Stat Rep*. 2014;22(73):1–21.
21. Hotaling JM, Davenport MT, Eisenberg ML, VanDenEeden SK, Walsh TJ. Men who seek infertility care may not represent the general U.S. population: data from the National Survey of Family Growth. *Urology*. 2012;79(1):123–7. doi:10.1016/j.urology.2011.09.021.
22. Ahrens KA, Cole SR, Westreich D, Platt RW, Schisterman EF. A cautionary note about estimating effects of secondary exposures in cohort studies. *Am J Epidemiol*. 2015;181(3):198–203. doi:10.1093/aje/kwu276. **This paper details the potential bias that can arise when studying secondary exposures in cohorts enriched for a primary exposure. While the example used throughout the paper is maternal smoking (exposure), study population (fetal growth restriction), and outcome (preterm birth), the same logic applies using ART as the study population and miscarriage as the outcome.**
23. Sunderam S, Kissin DM, Crawford SB, Folger SG, Jamieson DJ, Warner L, et al. Assisted reproductive technology surveillance—United States, 2014. *MMWR Surveill Summ*. 2017;66(6):1–24. doi:10.15585/mmwr.ss6606a1.
24. Gunby J, Bissonnette F, Librach C, Cowan L. Assisted reproductive technologies (ART) in Canada: 2007 results from the Canadian ART Register. *Fertil Steril*. 2011;95(2):542–7 e1–10. doi:10.1016/j.fertnstert.2010.05.057.
25. Kupka MS, D’hooghe T, Ferraretti AP, de Mouzon J, Erb K, Castilla JA, et al. Assisted reproductive technology in Europe, 2011: results generated from European registers by ESHRE. *Hum Reprod*. 2016;31(2):233–48. doi:10.1093/humrep/dev319.
26. Nyboe Andersen A, Erb K. Register data on assisted reproductive technology (ART) in Europe including a detailed description of ART in Denmark. *Int J Androl*. 2006;29(1):12–6. doi:10.1111/j.1365-2605.2005.00577.x.
27. Khamsi F, Lacanna I, Endman M, Wong J. Recent advances in assisted reproductive technologies. *Endocrine*. 1998;9(1):15–25. doi:10.1385/ENDO:9:1:15.
28. Land JA, Evers JL. Risks and complications in assisted reproduction techniques: report of an ESHRE consensus meeting. *Hum Reprod*. 2003;18(2):455–7.
29. Guidelines on number of embryos transferred. *Fertil Steril*. 2009;92(5):1518–9. doi:10.1016/j.fertnstert.2009.08.059.
30. Land JA, Evers JL. What is the most relevant standard of success in assisted reproduction? Defining outcome in ART: a Gordian knot of safety, efficacy and quality. *Hum Reprod*. 2004;19(5):1046–8. doi:10.1093/humrep/deh215.
31. Steptoe PC, Edwards RG, Purdy JM. Clinical aspects of pregnancies established with cleaving embryos grown in vitro. *Br J Obstet Gynaecol*. 1980;87(9):757–68.
32. Romundstad LB, Romundstad PR, Sunde A, von Düring V, Skjærven R, Gunnell D, et al. Effects of technology or maternal factors on perinatal outcome after assisted fertilisation: a population-based cohort study. *Lancet*. 2008;372(9640):737–43. doi:10.1016/S0140-6736(08)61041-7.
33. Saunders DM, Mathews M, Lancaster PA. The Australian Register: current research and future role. A preliminary report. *Ann N Y Acad Sci*. 1988;541:7–21.
34. McElrath TF, Wise PH. Fertility therapy and the risk of very low birth weight. *Obstet Gynecol*. 1997;90(4 Pt 1):600–5.
35. Hill GA, Bryan S, Herbert 3rd CM, Shah DM, Wentz AC. Complications of pregnancy in infertile couples: routine treatment versus assisted reproduction. *Obstet Gynecol*. 1990;75(5):790–4.
36. Braun JM, Messerlian C, Hauser R. Fathers matter: why it’s time to consider the impact of paternal environmental exposures on children’s health. *Curr Epidemiol Rep*. 2017.
37. Messerlian C, Wylie BJ, Minguez-Alarcon L, Williams PL, Ford JB, Souter IC, et al. Urinary concentrations of phthalate metabolites and pregnancy loss among women conceiving with medically assisted reproduction. *Epidemiology*. 2016;27(6):879–88. doi:10.1097/EDE.0000000000000525.
38. Minguez-Alarcon L, Chiu YH, Messerlian C, Williams PL, Sabatini ME, Toth TL, et al. Urinary paraben concentrations and in vitro fertilization outcomes among women from a fertility clinic. *Fertil Steril*. 2016;105(3):714–21. doi:10.1016/j.fertnstert.2015.11.021.
39. Hauser R, Gaskins AJ, Souter I, Smith KW, Dodge LE, Ehrlich S, et al. Urinary phthalate metabolite concentrations and reproductive outcomes among women undergoing fertilization: results from the EARTH study. *Environ Health Perspect*. 2015; doi:10.1289/ehp.1509760.
40. Minguez-Alarcon L, Afeiche MC, Chiu YH, Vanegas JC, Williams PL, Tanrikut C, et al. Male soy food intake was not associated with in vitro fertilization outcomes among couples attending a fertility center. *Andrology*. 2015;3(4):702–8. doi:10.1111/andr.12046.
41. Gaskins AJ, Afeiche MC, Hauser R, Williams PL, Gillman MW, Tanrikut C, et al. Paternal physical and sedentary activities in relation to semen quality and reproductive outcomes among couples from a fertility center. *Hum Reprod*. 2014;29(11):2575–82. doi:10.1093/humrep/deu212.
42. McLernon DJ, Steyerberg EW, Te Velde ER, Lee AJ, Bhattacharya S. Predicting the chances of a live birth after one or more complete cycles of in vitro fertilisation: population based study of linked cycle data from 113,873 women. *BMJ*. 2016;355:i5735. doi:10.1136/bmj.i5735.
43. Ammon Avalos L, Galindo C, Li DK. A systematic review to calculate background miscarriage rates using life table analysis. *Birth Defects Res A Clin Mol Teratol*. 2012;94(6):417–23. doi:10.1002/bdra.23014.
44. Mumford SL, Schisterman EF, Cole SR, Westreich D, Platt RW. Time at risk and intention-to-treat analyses: parallels and implications for inference. *Epidemiology*. 2015;26(1):112–8. doi:10.1097/EDE.0000000000000188.
45. Maity A, Williams PL, Ryan L, Missmer SA, Coull BA, Hauser R. Analysis of in vitro fertilization data with multiple outcomes using discrete time-to-event analysis. *Stat Med*. 2014;33(10):1738–49. doi:10.1002/sim.6050.
46. Missmer SA, Pearson KR, Ryan LM, Meeker JD, Cramer DW, Hauser R. Analysis of multiple-cycle data from couples undergoing in vitro fertilization: methodologic issues and statistical approaches. *Epidemiology*. 2011;22(4):497–504. doi:10.1097/EDE.0b013e31821b5351.
47. Pearson KR, Hauser R, Cramer DW, Missmer SA. Point of failure as a predictor of in vitro fertilization treatment discontinuation. *Fertil Steril*. 2009;91(4 Suppl):1483–5. doi:10.1016/j.fertnstert.2008.07.1732.
48. Fitzmaurice GM, Laird NM, Ware JH. Missing data and dropout. Applied longitudinal analysis. Hoboken: John Wiley and Sons, Inc.; 2004.
49. Zhang J, Yu KF. What’s the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA*. 1998;280(19):1690–1.
50. Sunderam S, Kissin DM, Flowers L, Anderson JE, Folger SG, Jamieson DJ, et al. Assisted reproductive technology surveillance—United States, 2009. *MMWR Surveill Summ*. 2012;61(7):1–23.
51. Austin PC, Laupacis A. A tutorial on methods to estimating clinically and policy-meaningful measures of treatment effects in prospective observational studies: a review. *Int J Biostat*. 2011;7(1):6. doi:10.2202/1557-4679.1285.
52. Zou G. A modified Poisson regression approach to prospective studies with binary data. *Am J Epidemiol*. 2004;159(7):702–6.

53. Austin PC. Absolute risk reductions and numbers needed to treat can be obtained from adjusted survival models for time-to-event outcomes. *J Clin Epidemiol*. 2010;63(1):46–55. doi:[10.1016/j.jclinepi.2009.03.012](https://doi.org/10.1016/j.jclinepi.2009.03.012).
54. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol*. 2016;31(4):337–50. doi:[10.1007/s10654-016-0149-3](https://doi.org/10.1007/s10654-016-0149-3).
55. Wasserstein R, Lazar N. The ASA's statement on *p*-values: context, process, and purpose. *Am Stat*. 2016;70(2):129–33. doi:[10.1080/00031305.2016.1154108](https://doi.org/10.1080/00031305.2016.1154108).
56. Farland LV, Correia KF, Wise LA, Williams PL, Ginsburg ES, Missmer SA. *P*-values and reproductive health: what can clinical researchers learn from the American Statistical Association? *Hum Reprod*. 2016;31(11):2406–10. doi:[10.1093/humrep/dew192](https://doi.org/10.1093/humrep/dew192). **A thorough discussion of the utility of *p*-values in reproductive epidemiology research complete with recommendations for presenting results moving forward.**