Big data approach to evaluation of birth defects and assisted reproductive technology: the Chinese linkage cohort

In the article by Hui-ting et al. (1) a regional registry is utilized to define a cohort of over 2 million births. By using China's national identifier of individual parents, this regional registry links to detailed clinical information from a single large in vitro fertilization (IVF) center, providing conception and birth outcomes on more than 6,000 assisted reproductive technology (ART) births. We commend the authors for managing an extensive database and analyzing a formidable amount of patient information. This work illustrates a growing ability to use big data to interrogate important research hypotheses from secondary data sources.

Compared to primary data collection, secondary data analyses using linked resources offer several advantages. In addition to being economically efficient, the magnitude of these datasets facilitates the study of rare exposures and outcomes. In this case, the use of comprehensive hospital admission and discharge data ensured that all population births were accounted for when examining the association between IVF/intracytoplasmic sperm injection (ICSI) exposures and rare birth defect outcomes and whether these exposures impact single versus multiple births differentially. By pooling data spanning over a decade, Hui-ting and colleagues (1) amass a large repository of data with the ability to identify multiple children within a family both conceived naturally and from ART. Despite the inviting power and efficiency advantages of working with secondary data, there are several issues that must be considered when evaluating or conducting research based upon linked information.

With the expanded use of linked datasets, transparency regarding the matching of records from multiple sources is essential. Particular attention should be paid to the accuracy of matching as well as the assumptions made about unmatched records. For example, Davies et al (2) linked data from a regional birth defects registry to data from 2 clinics in South Australia whose work accounts for 99.99% of all ART births. In contrast, the present study defines the exposed population as those 6,372 births resulting from IVF/ICSI at the Shanghai JIAI Genetics and IVF Institute. While this institution is the preeminent fertility center in the region, it accounts for only 40% of the IVF cycles in Shanghai during the study period. The resulting un-matched births in the cohort are used as the comparator total population. This heterogeneous group includes not only spontaneous conceptions but also an unknown number of conceptions from ovulation induction, intrauterine insemination, and IVF/ICSI cycles from other centers. As a result, there is inherent misclassification within the data such that some IVF-conceived infants are inadvertently labeled as unexposed. By comparing IVF births to an IVF-contaminated comparison population, associations of interest are likely biased towards the null.

As with any retrospective study, administrative and registry data studies are limited by the quality of the collected data. In the present study, birth defects were defined as those detected within the first 7 days of life and reported by a physician using ICD-10 codes. This method is advantageous in that birth defect data could be efficiently captured with limited resources. However, without multiyear follow up similar to that provided by Davies et al., it is possible that birth defects are underreported or misclassified. Additionally, the present study only evaluates the association between IVF/ICSI and birth defects, without examining the impact of infertility history or other treatments such as ovulation induction or intrauterine insemination. Finally, as with many linked datasets, there are potentially unmeasured confounders which may bias estimates of the associations of interest.

Hui-ting and colleagues (1) make innovative use of a robust, linked, regional dataset to report small but significant associations between ART and specific birth defects. Their results are similar to the growing international literature on this topic using linked data sources. Despite their limitations, these studies provide valuable information about associations that would otherwise be difficult to obtain. China's national identifier of individual parents enables the linkage used by the current study, a convenience that is often unavailable in other countries including U.S. In the U.S., issues of patient privacy can present a barrier to linkage studies. Often, data arising from different sources require individual negotiations and contracts to gain access. When no unique identifier exists, a probabilistic linkage can be applied which relies on information shared across data sources, such as demographics, and uses an algorithm to compute a match probability. Procedures for matching should be carefully documented including a validation of the matching procedures. Details such as the percent of records ultimately matched and results of the validation should be included in study methods (3).

A greater understanding of the impact of infertility and its treatment on the long-term health of the child and mother is critical to the field of reproductive endocrinology and infertility. The use of high fidelity big data for research, as presented here, has been used successfully to explore these questions. In the U.S., IVF data are reported annually to the Society for Assisted Reproductive Technologies, and significant effort is being put forth to obtain and validate the addition of a clinical outcomes reporting system. However, a recent study (4) compared the accuracy of these data on birth outcomes with the Massachusetts birth registry and noted disappointingly low sensitivity (18.4%–50.0%) for specific birth defects. This study highlights the importance of reliable prospectively collected data, particularly with regard to birth outcomes, to draw accurate conclusions from similar large database studies. An improvement upon prior methods, the States Monitoring ART collaborative (5) has demonstrated the ability to link about 90% of Society for Assisted Reproductive Technologies data from the Centers for Disease Control and Prevention to vital records in three states while

maintaining patient privacy. The States Monitoring ART collaborative represents a small but important step forward in infertility health services research. In an era of extensive electronic medical records and data digitalization, we urge governmental agencies and investigators to work together to remove barriers, develop standards, and capitalize upon state systems to collect and integrate high quality data on IVF outcomes (3). While we in the U.S. work to improve our collection of big data, we are fortunate to learn from the experience of our international colleagues and welcome their advances.

Mary D. Sammel, Sc.D.[a]
Natalie Stentz, M.D.[b]
Divya K. Shah, M.D., M.M.E.[b]

[a] Department of Biostatistics, Epidemiology and Informatics, Center for Clinical Epidemiology and Biostatistics, Women's Health Clinical Research Center; and [b] Division of Reproductive Endocrinology and Infertility, Department of Obstetrics & Gynecology, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania

https://doi.org/10.1016/j.fertnstert.2018.02.115

You can discuss this article with its authors and other readers at
https://www.fertstertdialog.com/users/16110-fertility-and-sterility/posts/30219-25768

REFERENCES

1. Hui-ting Y, Yang Q, Xiao-xi S, Guo-wu C, Nai-si Q, Ren-zhi C, et al. Associations of birth defects with the Mode of assisted reproductive technology in a Chinese data linkage cohort. Fertil Steril 2018;109:849–56.
2. Davies MJ, Moore VM, Willson KJ, Van Essen P, Priest K, Scott H, et al. Reproductive technologies and the risk of birth defects. N Engl J Med 2012;366:1803–13.
3. Bradley CJ, Penberthy L, Devers KJ, Holden DJ. Health services research and data linkages: issues, methods, and directions for the future. Health Serv Res 2010;45:1468–88.
4. Stern J, Gopal D, Liberman RF, Anderka M, Kotelchuck M, Luke B. Validation birth outcomes from the SART CORS: population-based analysis from the Massachusetts Outcome Study of Assisted Reproductive Technology (MOSART). Fertil Steril 2016;106:717–22.
5. Mneimneh AS, Boulet SL, Sunderam S, Zhang Y, Jamieson DJ, Crawford S, et al. States Monitoring Assisted Reproductive Technology (SMART) Collaborative: data collection, linkage, dissemination, and use. J Womens Health 2013;22:571–7.