

# Limitations and opportunities in male fertility databases

Brent M. Hanson<sup>1</sup>, James M. Hotaling<sup>2</sup>

<sup>1</sup>Department of Obstetrics and Gynecology, University of Utah, Salt Lake City, UT, USA; <sup>2</sup>Department of Surgery-Urology, University of Utah Center for Reconstructive Urology and Men's Health, Salt Lake City, UT, USA

*Contributions:* (I) Conception and design: All authors; (II) Administrative support: None; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

*Correspondence to:* Brent M. Hanson, MD. Department of Obstetrics and Gynecology, University of Utah, 50 N Medical Drive, Salt Lake City, UT 84132, USA. Email: Brent.hanson@hsc.utah.edu; James M. Hotaling, MD, MS, FECSM. Department of Surgery-Urology, University of Utah Center for Reconstructive Urology and Men's Health, 675 Arapleen Drive, Suite 205, Salt Lake City, UT 84108, USA. Email: Jim.hotaling@hsc.utah.edu.

**Abstract:** Over the last several years, the male component of reproduction has begun to gain clinical momentum. The medical literature has traditionally focused on infertility from the female perspective, but recent publications have demonstrated that male infertility is an important marker of overall health for infertile men as well as their family members. In order to perform large-scale, quality research related to male infertility, comprehensive databases are necessary. Currently, research in male infertility is limited by the fact that there is not a centralized, comprehensive database specifically designed to collect patient information related to male fertility. A database of this nature exists for female infertility research in the form of the Society for Assisted Reproductive Technology (SART) clinical summary report and the National ART Surveillance System (NASS) published by the Centers for Disease Control (CDC). This review outlines the strengths and weaknesses of several male fertility data sources, including the National Survey of Family Growth, the Reproductive Medicine Network, the Andrology Research Consortium (ARC), the Truven Health MarketScan<sup>®</sup> databases, the Utah Population Database, and data available from the Ober Lab related to the Hutterites. While each of these sources has been instrumental in the creation of meaningful research within the field of male fertility, a need remains for the creation of a centralized database for use in future male fertility research. The ideal database would consist of vast amounts of patient data which link individuals and couples to biologic specimens as well as data from family members, designed with parameters specifically purposed for male fertility research. The use of electronic medical records (EMR) systems such as Epic may play a role in the development of such a database going forward. At present, although some information is available through current databases, researchers must utilize suboptimal data sources to perform studies.

**Keywords:** Database; male infertility; research strategy

Submitted Jan 01, 2018. Accepted for publication Apr 11, 2018.

doi: 10.21037/tau.2018.04.18

**View this article at:** <http://dx.doi.org/10.21037/tau.2018.04.18>

## Introduction

Historically, research within the field of infertility has placed a disproportionate emphasis on the female component of reproduction. As a result, the field of male infertility is severely lacking in high quality, large-scale studies evaluating important questions as they relate to

male fertility. Many issues which have been investigated from a female perspective have never been evaluated using data from male subjects. In recent years, however, publications regarding the male component of reproduction and possible associations between male infertility and other health comorbidities have gained increasing attention (1). Male factor infertility appears to be associated with somatic

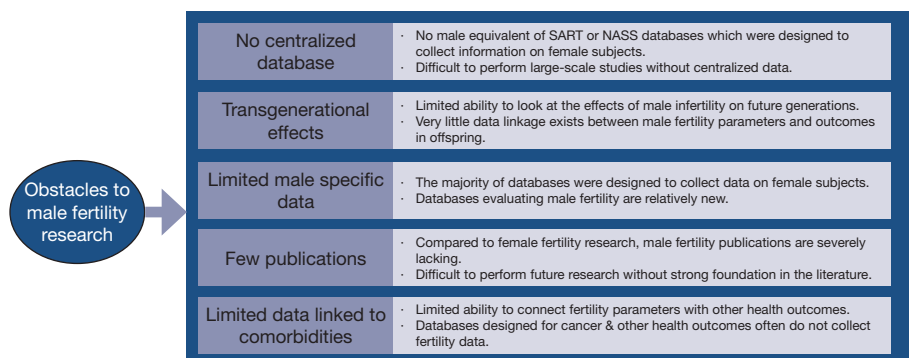


Figure 1 Obstacles to male fertility research.

health, and men with infertility appear to be at an increased risk of developing various somatic health problems as well as certain types of malignancy (2-5). The use of male infertility as a biomarker of the overall health status of infertile men as well as their family members is becoming increasingly accepted (6-8). Ongoing research is crucial to identify specific associations between male infertility and various health outcomes, which are likely the result of a complex interplay of molecular, environmental, and genetic factors. Recent evidence demonstrating decreasing sperm quality, specifically in Western nations, is troublesome (9). Investigating this trend further requires dedication to the field of male infertility as well as deep and granular repositories of patient information. Databases containing patient health outcomes which are linked to fertility parameters are an invaluable source of novel information. With access to ideal repositories of male fertility data, researchers could more adequately investigate unanswered questions regarding how to maximize male fertility. While the recent trend toward a greater focus on male fertility is encouraging, researchers who attempt to utilize large databases to extract information about male fertility patients often face significant obstacles (Figure 1).

One of the primary drawbacks to performing male fertility research using available databases is the severe lack of centralized data specifically related to male infertility. Many of the large databases such as the Society for Assisted Reproductive Technology (SART) clinical summary report and the National ART Surveillance System (NASS) published by the Centers for Disease Control (CDC) cannot be directly translated to research regarding infertile men because they are not specifically designed to include information about male factor infertility. The SART database reports outcomes related to egg retrievals, frozen

cycles, use of donor oocytes, live births, preterm deliveries, multiple gestations, and the number of procedures performed annually on a national scale (10). NASS contains data on nearly all ART procedures performed in the United States and details characteristics of patients, specific reproductive procedures, and pregnancy outcomes (11). Both SART and NASS have been used for male fertility research since a diagnosis of male factor infertility is included in these data sets, but the vast majority of data in both of these sources relates to the female component of fertility (12).

When evaluating associations between male infertility and malignancy, cancer databases such as the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute are not ideal because the information contained in the SEER database is not tied to fertility parameters. Databases such as the National Survey of Family Growth (NSFG) and the Reproductive Medicine Network have shown promise in the realm of male infertility. However, each of these data repositories has significant limitations since they were originally designed to collect data regarding women. One of the few data sources specifically designed to collect data about male factor infertility is the Andrology Research Consortium (ARC) database, but currently there is relatively limited published data available from this source. Many retrospective cohort studies evaluating male factor infertility and associated health outcomes have been performed using large, population-level databases such as the Truven Health MarketScan® and the Utah Population Database. Specific populations such as the Hutterites have also proven valuable as sources of information related to male infertility.

While each data source has specific advantages and limitations, the overarching theme within the field of male

infertility is that most data sources were not specifically designed to collect data related to male infertility and therefore are frequently unable to answer many important research questions. Additionally, researchers have a limited ability to look at transgenerational effects of male infertility due to poor longitudinal data. This review discusses several individual data sources and highlights the specific limitations and opportunities for each as they relate to male infertility research (*Table 1*).

## Male fertility data sources

### *National survey of family growth*

The National Survey of Family Growth (NSFG) began in 1973 as a survey designed by the Centers for Disease Control and Prevention (CDC) to be nationally representative of noninstitutionalized, civilian women 15 to 44 years of age. The survey gathers information on family life, marriage and divorce, pregnancy, infertility, use of contraception, and general and reproductive health (13). The first five cycles of the survey were conducted periodically between 1973 and 1995, providing a wealth of information regarding somatic health and infertility related specifically to women. Hundreds of publications on female fertility utilized this data set, but due to the lack of male data, the early years of the NSFG were unable to provide information to further the field of male fertility. In 2002, the sixth cycle of the survey included a study population of men aged 15 to 44 who were not connected to female survey participants. All cycles of the survey since 2002 have included both male and female survey participants, obtaining useful information related to male fertility. Questionnaires specifically address attitudes about childbearing among men, relationship status, number of biologic children fathered, the father's age at the time of fathered birth, history of infertility or difficulty achieving conception, use of adoption services, sexually transmitted infections, use of infertility services or medications, number of sexual partners, sexual practices, vasectomy history and reasons for past vasectomy, smoking status, and utilization of contraception (13). Since 2006, the NSFG has obtained data through a continuous interviewing protocol, conducting interviews 48 weeks out of every year, rather than through periodic surveys. Selected participants engage in a single, 60 to 80 minutes interview. Longitudinal data on individual participants is not available due to the design of the survey (14). In September 2015, the NSFG expanded

the age range for survey participants to include both men and women between the ages of 15 and 49, with similar survey content and scope. It is anticipated that during each calendar year, approximately 5,000 in-person interviews will continue to be completed by the NSFG (13). With the addition of male subjects and the expanded age range of survey respondents, the NSFG has become a worthwhile tool to further male fertility research.

One of the clear strengths of the NSFG is the high number of participants who have undergone well-validated surveys. Thousands of men and women participate in the NSFG survey each year, adding valuable data which has benefitted the fields of behavioral science, medicine, government, and health policy. Between the years of 1973 and 2015, over 600 journal articles were published using NSFG data (15). However, since the relatively recent addition of men to NSFG data in 2002, the vast majority of publications from the NSFG have not focused on male factor infertility. Between the years of 2002 and 2015, only a handful of studies have utilized NSFG data to evaluate issues related to male infertility. Despite lower numbers of publications related to male fertility, a distinct advantage of the NSFG is the ability to perform a propensity score analysis to reduce confounding from data extracted from this survey. A propensity score analysis is possible since the NSFG collects a wide range of demographic and health information in addition to fertility data. Multiple variables from the NSFG can be utilized by researchers to stratify groups and compare effects when evaluating questions related to male fertility, which is a noteworthy strength of this data set.

In spite of some limitations, there are remarkable opportunities for male fertility research within the NSFG. Publications which have utilized the NSFG have demonstrated its ability to provide meaningful data which can address a wide variety of male fertility topics. One of the earliest studies to utilize NSFG data to evaluate male infertility was published by Anderson *et al.* in 2009. This publication reported that 7.5% of all sexually active men had sought medical attention for help conceiving a child, with 2.2% of men undergoing an infertility consultation within the past year (16). In 2010, Breyer *et al.* evaluated 1,739 NSFG male respondents, 9.7% of whom had been diagnosed with infertility. Breyer concluded that a diagnosis of male infertility statistically reduced the odds of having a larger family, more so than a diagnosis of female infertility (17). In 2011, Zhang published a book highlighting various issues that have

**Table 1** Strengths and weaknesses of male fertility data sources

Data source	Year established	Affiliation	URL	Strengths	Weaknesses
National Survey of Family Growth (NSFG)	1973	Centers for Disease Control (CDC)	<a href="https://www.cdc.gov/nchs/nsfg/index.htm">https://www.cdc.gov/nchs/nsfg/index.htm</a>	Well-validated survey since 1973; nationally representative for ages 15–49; inclusion of male participants since 2002; approximately 5,000 interviews per year; assesses issues of pregnancy, fertility, contraception, general & reproductive health; over 600 journal articles published using NSFG	Originally designed for female respondents; limited scope related to infertility; limited information available for male infertility
Reproductive Medicine Network (RMN)	1989	Eunice Kennedy Shriver National Institute of Child Health & Human Development—NIH	<a href="https://www.nichd.nih.gov/research/supported/Pages/rmn.aspx">https://www.nichd.nih.gov/research/supported/Pages/rmn.aspx</a>	Designed to perform multi-center clinical trials; landmark papers have been published related to male infertility from RMN trials; over 50 publications since 1990; data can be linked to biologic specimens	Poor recruitment for male trials; majority of focus is on female infertility; failed varicoceleomy trial; not specifically designed for male infertility
Andrology Research Consortium (ARC)	2013	Society for the Study of Male Reproduction (SSMR)	<a href="http://ssmr.org/about/welcome.aspx">http://ssmr.org/about/welcome.aspx</a>	Prospective questionnaire data; 14 different centers specializing in male infertility; specifically designed for male infertility; supported by SSMR; female fertility data available	Limited availability of biologic specimens; requires clinician involvement to ensure ongoing enrollment & completion of questionnaires; no significant publications to date; relatively small patient size compared to other databases (approximately 2,000)
Truven Health MarketScan®	1988	Truven Health Analytics—IBM	<a href="https://truvenhealth.com/markets/life-sciences/products/data-tools/marketscan-databases">https://truvenhealth.com/markets/life-sciences/products/data-tools/marketscan-databases</a>	Over 240 million patients included; population-level data; all claims included in data set; ability to link infertility parameters to other health issues	Retrospective data; limited ability to link male & female partners; not specifically designed for male infertility
Utah Population Database (UPDB)	1987	University of Utah	<a href="https://healthcare.utah.edu/huntsmancancerinstitute/research/updb/">https://healthcare.utah.edu/huntsmancancerinstitute/research/updb/</a>	Information on >8 million individuals in Utah; data linked to >100 million family members; 85% of Utah medical records included in UPDB; 6 generations of pedigree data; biologic specimens available	Retrospective data; not representative of US population; not specifically designed for male infertility
The Hutterites & Ober Lab	1994	University of Chicago	<a href="http://ober.bsd.uchicago.edu/">http://ober.bsd.uchicago.edu/</a>	14-generation pedigree data available; fertility profiles obtained for nearly 300 couples; whole genome genetic data available; data available for males and females	Not representative of US population; founder population; not specifically designed for male infertility

caused men to be overlooked in fertility literature, citing NSFG data throughout the work (18). Data regarding male fertility rates from the 2002 NSFG survey were included in a 2012 publication by Joyner *et al.* evaluating the quality of male fertility data in US surveys (19). Also in 2012, Martinez *et al.* gathered NSFG responses to evaluate fertility outcomes for 10,403 men and 12,279 women, comparing average age at first birth in the 2002 data set to data from 2006–2010. Ultimately, many of the fertility measures including age at first birth were similar for men and women between 2002 and 2006–2010 (20). Eisenberg *et al.* utilized data from the NSFG in a 2013 publication which evaluated survey responses from 25,846 women and 11,067 men and concluded that during the workup for an infertile couple, 18% to 27% of male partners did not undergo a formal evaluation (21). A 2013 study by Chandra *et al.* demonstrated that 9.4% of men aged 15 to 44 and 12% of men aged 25 to 44 in 2006 to 2010 self-reported some form of male infertility, similar to levels seen in 2002 (22). Couple infertility from the male perspective was evaluated in a 2013 publication by Louis *et al.*, concluding that estimated prevalence of couple infertility was approximately 12% based on male responses from the 2002 NSFG survey (23). In 2015, Hotaling *et al.* evaluated data from 11,067 male NSFG respondents and highlighted demographic and socioeconomic differences between the 466 men who sought infertility evaluation and the 326 men who underwent vasectomy. The authors concluded that men seeking infertility evaluation were more likely to be younger, have a college or graduate degree, have fewer children, and were less likely to be married, while men undergoing a vasectomy were more likely to be married and white (24).

The publications outlined above provide valuable insight to the understanding of male fertility, and with ongoing data collection from male subjects, clear opportunities exist for further research within the NSFG. Limitations to the NSFG include a lack of longitudinal data to evaluate the impact of male infertility on subsequent generations, and detailed information regarding the specific cause of male infertility for each respondent is not included in the survey structure. The addition of supplementary male fertility variables to the survey design of the NSFG could strengthen this repository's utility in the future. Changes to the NSFG format over the last several years have improved its applicability to male fertility, but in its current state, the NSFG remains suboptimal for male fertility research. Compared to the number of publications obtained from

the NSFG database regarding contraception, sexually transmitted infections, and female health issues, male infertility publications are lacking, largely due to limitations in the type of data obtained through the survey.

### *Reproductive medicine network*

The Reproductive Medicine Network (RMN) was established in 1989 as a mechanism to carry out large, multicenter clinical trials. RMN clinical trials focus on therapeutic and diagnostic interventions for male and female infertility. Funding for the RMN comes through the Fertility and Infertility (FI) branch of the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). Six primary research sites across the United States and a data coordination center provide the infrastructure to allow investigators to enroll large numbers of patients at multiple centers (25). As opposed to survey-based data sources such as the NSFG, one of the key advantages of the RMN is the ability to collect biologic specimens for further laboratory evaluation. Since 1990, greater than 50 publications have come from RMN clinical trials, helping to advance research related to both male and female infertility (26).

One of the landmark studies in the field of male infertility by Guzick *et al.* in 2001 was the result of collaboration with the RMN. This publication evaluated differences in sperm morphology, motility, and concentration in fertile and infertile men, providing data to better define measurements which can distinguish fertile men from infertile men based on semen analysis parameters (27,28). RMN data was also used to perform a retrospective case-control study to evaluate the association between occupational exposures and male infertility in a publication by Gracia *et al.* in 2005. Although some limitations were present, this study concluded that no significant associations were noted between male factor infertility and exposure to shift work, metal fumes, electromagnetic fields, solvents, lead, paint, pesticides, work-related stress, or vibration (29). A 2013 publication by Jodar *et al.* utilized RMN data to evaluate the role of spermatozoal RNA, improving the understanding of the influences of RNA on fertilization, embryo development, the phenotype of offspring, and RNA's possible impact on future generations (30).

While a primary focus of the RMN is to perform clinical trials investigating male infertility issues, unfortunately the vast majority of RMN research has focused on female infertility. A publication by Trussell *et al.* in 2014 highlighted

the challenges of performing a prospective varicocelectomy trial, reporting that poor recruitment for a planned RMN varicocele trial ultimately led to closure by the Data Safety Monitoring Board. Despite a nearly universal sense from urologists that a prospective varicocelectomy trial would be important and worthwhile, the trial was unable to be completed (31). Poor recruitment and lack of ongoing trials focusing on males are certainly barriers to research within the RMN. At the date of publication, the RMN was actively recruiting for six clinical trials. Only one of the six trials—the Males, Antioxidants and Infertility (MOXI) trial—focuses primarily on male infertility, evaluating whether an antioxidant formulation administered to the male partner can help couples conceive spontaneously (32). While it is encouraging that meaningful research has come out of the RMN, a focus on female-related research continues to predominate. In spite of limitations, the RMN remains one of the only options for large-scale clinical trials related to male infertility.

### ARC

The ARC started in 2013 as an effort by the Society for the Study of Male Reproduction (SSMR). The purpose of the ARC is to provide male-specific fertility data which can be used for clinical, research, educational, and advocacy initiatives. Through ongoing enrollment, the ultimate goal of the ARC is to create a male-specific database analogous to the SART-CDC database (33). Based on information from a 2018 American Urological Association (AUA) accepted abstract, the ARC has collected greater than 4,300 questionnaires from patients at 24 centers specializing in male infertility in North America. The questionnaires capture data regarding age, ethnic and racial background, infertility history, lifestyle factors, and previous fertility therapies. Following enrollment, the information is then transferred to a central database for further analysis (33,34). This data set is unique because of the fact that the questions were specifically designed for male infertility patients. By structuring the questionnaire in this manner, the ARC obtains relevant information for use in future research within the field of male factor infertility.

The first report from the ARC in 2016 was based on responses from 2,108 patients. Findings from the initial report indicated that 16% of respondents had undergone intrauterine insemination (IUI) with their partners while 7.8% of couples had undergone in vitro fertilization (IVF). The study indicated that only 9.8% of couples who had

undergone IUI and only 28% of couples undergoing IVF had a male factor evaluation performed prior to the IUI or IVF procedure. Lifestyle factors were also assessed, documenting relatively high rates of marijuana use (12.8%), tobacco smoking rates of 15.6%, and exogenous testosterone use in 3% of respondents (33).

In a data environment heavily weighted toward female infertility, the ARC is an encouraging indicator of future trends within the field of male infertility research. As the volume of data within the ARC continues to grow, this data source will become increasingly impactful (33). While the questionnaire format of this data source is not conducive to studies requiring biologic specimens, physical exam findings, and longitudinal tracking, the ARC will likely be a useful source of data in the future for male-specific fertility research as ongoing enrollment and utilization answer meaningful clinical questions.

### Truven Health MarketScan®

The Truven Health MarketScan® Research Databases began in 1988, and since that time, data has been collected on over 240 million unique patients. Truven Health Analytics is a commercial entity which collects and then provides healthcare information to researchers, hospitals, clinicians, government agencies, employers, pharmaceutical companies, and medical device companies, among others. Information available through the MarketScan® databases includes medical claims documentation, data points from electronic medical records (EMR), laboratory information, and patient-level data. All data obtained from individual patients are de-identified and HIPAA compliant. Healthcare researchers are able to link claims data to patient-level data, providing a unique opportunity to access an extremely large pool of patients (35). By incorporating information based on insurance claims, the MarketScan® databases are able to capture all claims and provide massive amounts of comprehensive data at the population level.

While fertility is not the primary focus of this data set, medical claims related to infertility are included within the database. The MarketScan® databases have been instrumental in the publication of over 1,400 peer-reviewed journal articles on various topics since inception in 1988. This database provides an excellent foundation for population-level research (35). Within the field of male infertility, researchers have utilized this data set to perform retrospective cohort studies based on infertility data linked to other parameters. In 2015, Eisenberg *et al.* analyzed

subjects from the MarketScan<sup>®</sup> database from 2001 to 2009 to determine whether infertility was associated with increased rates of malignancy. United States claims data for 76,083 infertile men was collected, demonstrating an increased risk of testicular cancer in infertile men as well as an increased risk of all cancers in the years after an infertility evaluation (2). In 2016, a retrospective cohort study of similar design by Eisenberg *et al.* was performed to evaluate the risk of incident chronic medical conditions in a cohort of infertile men. A total of 13,027 men with a diagnosis of male factor infertility were found to have a higher risk of developing diabetes, ischemic heart disease, alcohol abuse, and drug abuse compared to men who only received infertility testing but were not ultimately classified as infertile (36). In 2017, Rao *et al.* utilized the MarketScan<sup>®</sup> database to perform a retrospective, cross-sectional analysis evaluating trends in testosterone replacement therapy among men of reproductive age between the years of 2003 and 2013. Researchers were able to extract information regarding usage of testosterone from an extremely large patient sample, including 5,094,868 men from the year 2013 alone. This study demonstrated that use of testosterone replacement therapy had increased fourfold in men between the ages of 18 and 45 during the time frame from 2003 to 2013 (37).

Overall, strengths of the MarketScan<sup>®</sup> data include a large number of comprehensive records based on claims data which enable researchers to perform population-level analyses. However, the retrospective nature of the data is somewhat limiting, and the data is unable to be linked so that male and female partners can be analyzed as couples. Additionally, while an extremely large number of publications have come from this data source, the database was not specifically designed as a fertility database, so fertility-related publications represent only a small percentage of database output.

### **Utah Population Database**

The Utah Population Database (UPDB), affiliated with the Huntsman Cancer Institute at the University of Utah, is a database providing in-depth information to support research on genetics, epidemiology, demography, and public health. The UPDB started over 30 years ago and has been used since that time to evaluate reproductive outcomes, cancer, genetic mutations, inheritance patterns, and other health conditions from a population level. The UPDB provides access to information on greater than 8 million

individuals within the state of Utah and includes claims data from hospital records, surgery records, and outpatient visits in Utah. Approximately 85% of Utah medical records are incorporated into the UPDB. The UPDB is unique in many ways, but one of the most interesting aspects of this database is that patient information is linked to other family members, providing one of the only sources of transgenerational, longitudinal data within the field of fertility research. Information from greater than 24 million patient records from around the world is being linked to individuals within the UPDB through gathering of family data (38). While not primarily designed for the purpose of fertility research, information related to fertility is one of the parameters collected through the UPDB. The UPDB links information for approximately 50,000 infertile patients to approximately 500,000 of their relatives. In addition to providing claims and outcomes data, the UPDB has a vast collection of biologic specimens (38).

Many valuable publications within the field of male infertility have come from the UPDB. Because of the familial linkage aspect of the UPDB, research from the UPDB has been able to highlight associations between infertility and health outcomes not only in probands but in family members as well. Associations between male infertility and malignancy were reported in a 2016 publication by Hanson *et al.* evaluating 20,433 men undergoing semen analysis and 20,433 fertile controls. Data was obtained from the UPDB. Ultimately, the study found that men with poor sperm quality had elevated rates of testicular cancer. Interestingly, sperm concentration and count in the 90<sup>th</sup> percentiles were also associated with increased rates of melanoma (39). A 2016 publication by Anderson *et al.* used UPDB data to evaluate cancer risk in first- and second-degree relatives of men with poor sperm quality, demonstrating that the first-degree relatives of men who had undergone semen analysis had a 52% increased risk of testicular cancer compared with the first-degree relatives of fertile controls. There did not appear to be a significant difference in testicular cancer risk for the second-degree relatives based on any semen parameters. Additionally, the first-degree relatives and second-degree relatives of azoospermic men had a significantly increased risk of thyroid cancer compared with fertile controls (7). Another study evaluating associations between male infertility and family members was published by Hanson *et al.* in 2017, demonstrating that the first-degree relatives of men with azoospermia had an elevated risk of childhood death related to congenital malformations (8). A 2017 study by Anderson *et al.* also evaluated associations

between male infertility and family members. This study utilized UPDB information to demonstrate that the siblings of men with oligozoospermia were at an increased risk for any-site cancer as well as acute lymphoblastic leukemia (ALL), suggesting a shared genetic or epigenetic connection between infertility and malignancy (6).

While the UPDB has many clear strengths, it is not without flaws. Similar to the MarketScan® databases, the retrospective nature of the data can be limiting for certain studies. Additionally, the population included in the UPDB is highly representative of the state of Utah, but this population is likely unrepresentative of other regions, limiting the external validity of research obtained from the UPDB. The UPDB has proven useful in male infertility research, but like most data sources, the utility of the data for male infertility research is limited by the fact that the database was not originally designed for fertility research. The types of research questions which can be addressed through the database are limited by the retrospective input.

### *The Hutterites and male fertility*

In addition to standard databases, founder populations represent an interesting source of information to evaluate male fertility. One founder population of specific interest is the Hutterites, a conservative Anabaptist religious group that does not believe in any form of contraception. The Hutterites reside in many rural areas in the Midwestern United States and southern Canada (40). The Ober Lab at the University of Chicago was created with the research goal of identifying genetic variants that influence gene expression in tissues related to asthma, chronic rhinosinusitis, and fertility and parturition. The Ober Lab has been studying the Hutterite population since 1994 and has developed a data set that includes greater than 1,400 individuals who are related through a fourteen-generation pedigree. Whole exome data is available for this population, making it feasible to perform studies assessing the impact of genetics on infertility. Prospective fertility profiles as well as genetic information has been obtained for approximately 300 Hutterite couples. Data collection has focused on both males and females, and while this population is not representative of the United States population as a whole, useful publications have resulted from the study of the Hutterites (41).

Using Hutterite data from the Ober Lab, a publication in 2010 by Kosova *et al.* evaluated the contribution of genetic variation on traits of reproduction, concluding that

there appeared to be significant inheritance patterns for reproductive traits in both men and women which could not be entirely accounted for by environmental exposures (42). Kosova *et al.* in 2012 published a genome-wide association study to broadly survey genes contributing to variations in male fertility. Since prior studies had focused primarily on animal models, this study was one of the first efforts to evaluate the impact of genetics on fertility in humans. Kosova's publication concluded that mutations in specific genes may account for a percentage of unexplained male infertility or subfertility within the general population (43). A subsequent publication by Kosova *et al.* in 2014 expounded on the Hutterite data and evaluated genetic variants which had previously been shown to be predictors of family size and birth rate in healthy men. The study evaluated whether changes in specific genes were associated with alterations in sperm morphology, ultimately concluding that single nucleotide polymorphisms (SNPs) in two genes, both of which have roles in nervous system development, were associated with poor sperm morphology (44).

Hutterite data provides one of the best sources of genetic information as it relates to infertility. The relatively insular nature of the Hutterite population makes extensive longitudinal tracking possible through many generations. Additionally, the reproductive histories and genetic information available for couples allows researchers to evaluate infertility at both the individual and partner level. However, as noted previously, the Hutterites are not representative of the general population. Additionally, while the Ober Lab does place an emphasis on fertility research, there is also a high degree of emphasis placed on asthma and chronic rhinosinusitis research. Like many other data sources, the Ober Lab's Hutterite dataset was not specifically designed to solely perform fertility research.

### **Summary**

Each data source outlined above has specific strengths and the potential to continue to provide meaningful contributions to the field of male fertility research. However, at present, male fertility research continues to be limited by the lack of breadth and depth within a centralized database specifically designed to collect male infertility data. The Andrology Research Consortium provides the most promise as a centralized database and will grow in research impact as available data increases. Retrospective data and the inclusion of populations which may not be broadly representative are also limiting factors within the field of



male fertility research. Current data sources were primarily designed for purposes other than male infertility research, which makes application of data challenging.

Female infertility research has been able to benefit from the Society for Assisted Reproductive Technology (SART) clinical summary report and the National ART Surveillance System (NASS) published by the Centers for Disease Control (CDC). This database incorporates information from 95% of IVF cycles in the United States and provides an excellent summary of IVF success from a female perspective (45). While efforts have been made by the Andrology Research Consortium to achieve a superior data set for male factor infertility, the clinical application of these efforts is still being realized. Going forward, male fertility research would benefit from the creation of a centralized database which could compile patient information from clinical intake forms, laboratory results, and EMR. Implementing a system in which male fertility patients could opt out of data collection rather than opt in would dramatically increase the volume of data available. While retrospective data is valuable, having the ability to collect biologic specimens and link patient-level data to these specimens would allow for extensive research opportunities in the future. Using the SART/NASS database as a starting point, male fertility specialists should expand upon this model to build a better, all-inclusive data set for the sole purpose of male fertility research. Recognizing the limitations of the SART/NASS database and tailoring a database specifically to address the needs of male fertility would advance the current research environment dramatically.

Current EMR systems such as Epic have the potential to synthesize vast amounts of integrated data. Through Epic or similar records systems, possibilities exist to link male and female fertility information to other comorbidities. This source of novel patient information is one of the most appealing opportunities for future medical research since healthcare systems have uniformly adopted electronic records in clinical practice. EMR systems provide the opportunity to design automated data fields which have the potential to capture massive amounts of patient information tailored to male fertility issues which can then be pooled for research purposes. To date, the ability to combine patient data from multiple institutions using a shared EMR system has been underutilized. Additionally, responses to electronic versions of the Andrology Research Consortium questionnaire, connections to the Utah Population Database, and patient information linked to biologic specimens are all realistic possibilities through

synchronized EMR systems. The creation of a HIPAA-compliant web portal specifically for male infertility is entirely within the scope of EMR systems. The use of Epic or other similar systems to perform male fertility research has not reached its full potential. While the use of EMR databases to perform research comes with its own set of challenges, this remains one of the most promising avenues for future research within the field of male fertility. The ideal prospective male fertility database would include a combination of demographic information as well as biologic specimens linked to partner information through the EMR. The ability to track couples longitudinally would enhance the value of data and enable researchers to follow outcomes from an infertility diagnosis through multiple rounds of treatment. An integrated, electronic database from multiple institutions is necessary to generate adequate patient volume and increase external validity of research findings. The use of standardized laboratory and clinical metrics at the time of data entry is also necessary to ensure that findings are entered identically in all patient settings. Male infertility researchers have the opportunity to create relevant, clinically applicable results by evaluating the impact of infertility on population-level health. Through a commitment to build systems that increase clinical efficiency while simultaneously generating data for research, researchers can improve the likelihood that quality data will be available in the future for male fertility studies.

### Acknowledgements

None.

### Footnote

*Conflicts of Interest:* The authors have no conflicts of interest to declare.

### References

1. Petok WD. Infertility counseling (or the lack thereof) of the forgotten male partner. *Fertil Steril* 2015;104:260-6.
2. Eisenberg ML, Li S, Brooks JD, et al. Increased risk of cancer in infertile men: analysis of U.S. claims data. *J Urol* 2015;193:1596-601.
3. Brinton LA, Westhoff CL, Scoccia B, et al. Causes of infertility as predictors of subsequent cancer risk. *Epidemiology* 2005;16:500-7.
4. Kort JD, Eisenberg ML, Millheiser LS, et al. Fertility

- issues in cancer survivorship. *CA Cancer J Clin* 2014;64:118-34.
5. Craig JR. Obesity, male infertility, and the sperm epigenome. *Fertil Steril* 2017;107:848-59.
  6. Anderson RE, Hanson HA, Lowrance WT, et al. Childhood cancer risk in the siblings and cousins of men with poor semen quality. *J Urol* 2017;197:898-905.
  7. Anderson RE, Hanson HA, Patel DP, et al. Cancer risk in first- and second-degree relatives of men with poor semen quality. *Fertil Steril* 2016;106:731-8.
  8. Hanson HA, Mayer EN, Anderson RE, et al. Risk of childhood mortality in family members of men with poor sperm quality. *Hum Reprod* 2017;32:239-47.
  9. Levine H, Jørgensen N, Martino-Andrade A, et al. Temporal trends in sperm count: a systematic review and meta-regression analysis. *Hum Reprod Update* 2017;23:646-59.
  10. National Summary Report [database on the Internet]. Society for Assisted Reproductive Technology. 2014. Accessed: September 2016
  11. Kissin DM, Kulkarni AD, Mneimneh A, et al. Embryo transfer practices and multiple births resulting from assisted reproductive technology: an opportunity for prevention. *Fertil Steril* 2015;103:954-61.
  12. Odisho AY, Nangia AK, Katz PP, et al. Temporal and geospatial trends in male factor infertility with assisted reproductive technology in the United States from 1999–2010. *Fertil Steril* 2014;102:469-75.
  13. CDC. National Survey of Family Growth. In: Centers for Disease Control and Prevention. 2017. Available online: <https://www.cdc.gov/nchs/nsfg/index.htm>, accessed Nov 21 2017.
  14. CDC. 2013–2015 National Survey of Family Growth (NSFG): Summary of Design and Data Collection Methods. CDC Website. 2015. Available online: [https://www.cdc.gov/nchs/data/nsfg/NSFG\\_2013-2015\\_Summary\\_Design\\_Data\\_Collection.pdf](https://www.cdc.gov/nchs/data/nsfg/NSFG_2013-2015_Summary_Design_Data_Collection.pdf), accessed 29 March 2018.
  15. CDC. About the National Survey of Family Growth. 2017. Available online: [https://www.cdc.gov/nchs/nsfg/about\\_nsfg.htm](https://www.cdc.gov/nchs/nsfg/about_nsfg.htm), accessed Nov 21 2017.
  16. Anderson JE, Farr SL, Jamieson DJ, et al. Infertility services reported by men in the United States: national survey data. *Fertil Steril* 2009;91:2466-70.
  17. Breyer BN, Smith JF, Shindel AW, et al. The impact of infertility on family size in the USA: data from the National Survey of Family Growth. *Hum Reprod* 2010;25:2360-5.
  18. Zhang L. *Male Fertility Patterns and Determinants*. Springer, 2011.
  19. Joyner K, Peters HE, Hynes K, et al. The quality of male fertility data in major U.S. surveys. *Demography* 2012;49:101-24.
  20. Martinez G, Daniels K, Chandra A. Fertility of men and women aged 15–44 years in the United States: National Survey of Family Growth, 2006–2010. *Natl Health Stat Report* 2012;12:1-28.
  21. Eisenberg ML, Lathi RB, Baker VL, et al. Frequency of the male infertility evaluation: data from the national survey of family growth. *J Urol* 2013;189:1030-4.
  22. Chandra A, Copen CE, Stephen EH. Infertility and impaired fecundity in the United States, 1982–2010: data from the National Survey of Family Growth. *Natl Health Stat Report* 2013;14:1-18.
  23. Louis JF, Thoma ME, Sørensen DN, et al. The prevalence of couple infertility in the United States from a male perspective: evidence from a nationally representative sample. *Andrology* 2013;1:741-8.
  24. Hotaling JM, Patel DP, Brant WO, et al. Demographic and socio-economic differences between men seeking infertility evaluation and those seeking surgical sterilization: from the National Survey of Family Growth. *BJU Int* 2015;116:288-92.
  25. Jarvi K, Lau S, Lo K, et al. First report from the andrology research consortium. *J Urol* 2016;195:e1159.
  26. NICHD. Reproductive Medicine Network (RMN). 2017. Available online: <https://www.nichd.nih.gov/research/supported/Pages/rmn.aspx#Topic>, accessed Nov 22 2017.
  27. RMN. Reproductive Medicine Network (RMN) Publications. 2017. Available online: <http://c2s2.yale.edu/rmn/pubs/index.aspx>, accessed Nov 23 2017.
  28. Guzick DS, Overstreet JW, Factor-Litvak P, et al. Sperm Morphology, Motility, and Concentration in Fertile and Infertile Men. *NEJM* 2001;345:1388-93.
  29. Gracia CR, Sammel MD, Coutifaris C, et al. Occupational Exposures and Male Infertility. *Am J Epidemiol* 2005;162:729-33.
  30. Jodar M, Selvaraju S, Sandler E, et al. The presence, role and clinical use of spermatozoal RNAs. *Hum Reprod Update* 2013;19:604-24.
  31. Trussell JC, Ohl DA, Krawetz SA, et al. Nearly all Surveyed Reproductive Urologists Feel a Prospective Varicocelectomy Trial is Important and Worthwhile. *Fertil Steril* 2014;101:1563-4.e1.
  32. RMN. Reproductive Medicine Network (RMN) Infertility Trials. 2017. Available online: <http://c2s2.yale.edu/rmn/>

- studies/index.aspx, accessed Nov 23 2017.
33. Kim E. SSMR President's Welcome Message. In: Society for the Study of Male Reproduction. 2017. Available online: <http://ssmr.org/about/welcome.aspx>, accessed Nov 23 2017.
  34. AUA. American Urological Association 2018: Abstracts. 2018. Available online: <http://www.aa2018.org/>, accessed 28 March 2018.
  35. Truven. Putting Research Data Into Your Hands with the MarketScan Databases. 2017. Available online: <https://truvenhealth.com/markets/life-sciences/products/data-tools/marketscan-databases>, accessed Nov 23 2017.
  36. Eisenberg ML, Li S, Cullen MR, et al. Increased risk of incident medical conditions in infertile men: analysis of United States claims data. *Fertil Steril* 2016;105:629-36.
  37. Rao PK, Boulet SL, Mehta A, et al. Trends in Testosterone Replacement Therapy Use from 2003 to 2013 among Reproductive-Age Men in the United States. *J Urol* 2017;197:1121-6.
  38. HCI. Utah Population Database Overview. 2017. Available online: <https://healthcare.utah.edu/huntsmancancerinstitute/research/updb/>, accessed Nov 24 2017.
  39. Hanson HA, Anderson RE, Aston KI, et al. Subfertility increases risk of testicular cancer: evidence from population-based semen samples. *Fertil Steril* 2016;105:322-8.e1.
  40. Gemmell AP, Veach PM, MacFarlane I, et al. "If It Helps, It's Worth a Try": an Investigation of Perceptions and Attitudes about Genetic Counseling among Southern Manitoba Hutterites. *J Genet Couns* 2017;26:1357-71.
  41. Ober C. The Ober Lab Website. Chicago, IL. 2017. Available online: <http://ober.bsd.uchicago.edu/index.php>, accessed Nov 25 2017.
  42. Kosova G, Abney M, Ober C. Colloquium papers: Heritability of reproductive fitness traits in a human population. *Proc Natl Acad Sci U S A* 2010;107:1772-8.
  43. Kosova G, Scott NM, Niederberger C, et al. Genome-wide association study identifies candidate genes for male fertility traits in humans. *Am J Hum Genet* 2012;90:950-61.
  44. Kosova G, Hotaling JM, Ohlander S, et al. Variants in DPF3 and DSCAML1 are associated with sperm morphology. *J Assist Reprod Genet* 2014;31:131-7.
  45. SART. Society for Assisted Reproductive Technology (SART) Website. 2017. Available online: <http://www.sart.org/>, accessed Nov 25 2017.

**Cite this article as:** Hanson BM, Hotaling JM. Limitations and opportunities in male fertility databases. *Transl Androl Urol* 2018;7(Suppl 3):S292-S302. doi: 10.21037/tau.2018.04.18