The Canadian Assisted Reproductive Technologies Register (CARTR) Plus database: A validation study

Vanessa Bacal

A thesis submitted to the University of Ottawa in partial fulfillment of the requirements for the degree of Master of Science in Epidemiology

School of Epidemiology and Public Health

Faculty of Medicine

University of Ottawa

Acknowledgements

This thesis and the journey through graduate studies could not have been possible without the help and support from numerous individuals.

Dr. Laura Gaudet has been a role model in her research career, clinical acumen and family life. She has provided me not only with delicious home-cooked dinners, but also invaluable insight into specific issues in my research. Dr. Mark Walker has been enthusiastic and supportive and has provided sound advice since I joined the clinical investigator program. I am grateful to have Dr. Heather Shapiro as a mentor. Her clinical expertise has been indispensable and she has been available every step of the way. Dr. Deshayne Fell has been instrumental in the development and execution of this project. Despite being incredibly busy, she always takes time to share her expertise, making sense of complicated epidemiological problems.

This thesis could not have been possible without the collaboration of the team at BORN Ontario, who helped in the protocol development, facilitated in both the data collection and release. I especially want to thank Dr. Sandy Dunn for her advice on building the protocol, Catherine Riddell for expediting the release of data, Drs. Hao Wang and Mary Guo for working tirelessly to pull data in a timely fashion. To the group at OMNI (Ruth White, Dana Burke, Natalie Rybak), thank you for supporting me by allowing me to work in your office space, but also by bringing me into your research family with open arms and genuine smiles. Finally, to my friend Dr. Andrea Lanes, you have helped at every step of the grad studies process. You have been supportive by providing me not only with SAS code, but as a sounding board for all sorts of ideas and when needed, an extra pair of eyes.

Statement of Originality

I state that all of the work in this thesis, from commencement to completion, represents original work undertaken by me (Vanessa Bacal) as part of the requirements for the MSc degree in Epidemiology. My thesis supervisors and committee members provided guidance on the methodological aspects and clinical relevance of each study as outlined in the Contribution of Authors section.

Table of Contents

Acknowledgements ii

Statement of Originality iii

Preface ix

Title: A systematic review of database validation studies among fertility populations ix

Abstract xi

Chapter 1. Prologue 1

1.1 Background 1

1.1.1 Assisted reproductive technologies 1

1.1.2 Routinely collected data 2

1.1.3 Validation studies 2

1.1.4 CARTR Plus 3

1.2 Research objectives 3

1.3 Rationale 4

1.4 References 5

Chapter 2. Literature Review: [Manuscript 1] A systematic review of database validation studies among fertility populations 7

2.1 Preface to Manuscript 1 7

2.2 Title page 8

2.3 Abstract 9

2.4 Significance 10

2.5 Introduction 11

2.6 Methods 12

2.7 Results 15

2.8 Discussion 17

Acknowledgments 23

Funding 23

Conflicts of interest 23

2.9 References 24

2.10 Figures and tables 29

Appendix A: Supplemental information for Manuscript 1 39

Appendix 1. Medline search strategy 39

Appendix 2. Excluded references 41

Appendix 3. Estimates of the measures of validity from included studies 47

Appendix 4. PRISMA checklist 52

Chapter 3. [Manuscript 2] The Canadian Assisted Reproductive Technologies Register (CARTR) Plus database: A validation study 54

3.1 Preface to Manuscript 2 54

3.2 Title page 55

3.3 Abstract 56

3.4 Introduction 59

3.5 Methods 60

3.5.1 Study design 60

3.5.2 Clinic and chart selection 61

3.5.3 Data extraction 62

3.5.4 Statistical analyses 62

3.5.5 Hypothesis 63

3.5.6 Sample size calculation 63

3.6 Results 64

3.6.1 Patient intake 64

3.6.2 Dates 65

3.6.3 Stimulation 65

3.6.4 Retrieval 66

3.6.5 Embryology 66

3.6.6 Pregnancy 67

3.6.7 Sensitivity analysis - missing charts 68

3.6.8 Sensitivity analysis - assessment of clinic-specific results 68

3.7 Discussion 69

3.7.1 FSH, AMH, AFC 70

3.7.2 Advanced female age 70

3.7.3 Oocyte origin 71

3.7.4 Elective single embryo transfer 72

3.7.5 Chorionicity 73

3.7.6 Reason for cancelled cycle 74

3.7.7 Study limitations 74

3.7.8 What our study adds to current literature 75

3.8 Conclusion 76

Acknowledgments 76

3.9 References 77

3.10 Figures and tables 80

Appendix B: Supplemental information for Manuscript 2 92

Appendix 1- Prevalence estimates, means or medians of selected variables from cycle years 2013-2015 from CARTR Plus 92

Appendix 2-Measures of agreement 94

Appendix 3-Level of agreement for kappa and intraclass correlation coefficients 96

Appendix 4- Sample size calculation sensitivity analysis 97

Appendix 5. Description of study variables by data source for clinic 1 99

Appendix 6. Description of study variables by data source for clinic 2 103

Appendix 7. Description of study variables by data source for clinic 3 107

Appendix 8. Description of study variables by data source for clinic 4 111

Appendix 9. Description of study variables by data source for clinic 5 115

Appendix 10. Description of study variables by data source for clinic 6 119

Appendix 11-Description of missing charts from CARTR Plus 123

Appendix 12. Recommendations for changes to CARTR Plus 125

Appendix 13. Checklist of reporting criteria for studies validating health administrative data algorithms.. 127

Chapter 4. Discussion 129

4.1 Infertility burden 129

4.2 ART monitoring 130

4.3 Reference standards 131

4.4 Population selection 132

4.5 Measures of validity 133

4.6 Implications for future use 134

4.7 Final conclusions 135

4.8 References 137

Appendix C: Certificates of ethical approval 140

TOH approval 140

CHEO exemption 144

Preface

Manuscript 1: Vanessa Bacal, Miguel Russo, Deshayne B Fell, Heather Shapiro, Mark Walker, Laura M Gaudet

Title: A systematic review of database validation studies among fertility populations

Author contributions:

I created the study protocol, developed the search strategy, conducted the search of the literature, did the primary and secondary screen of the articles, extracted data from the included articles, analyzed and interpreted these data, drafted and edited the manuscript. M. Russo performed the primary and secondary screen of the articles, extracted data from included articles and edited the manuscript. D.B. Fell, M. Walker and L.M. Gaudet assisted with the development of the study protocol, interpretation of results and editing of the manuscript. H. Shapiro assisted with interpreting the results and editing the manuscript.

Special approvals:

- None

Manuscript 2: Vanessa Bacal, Deshayne B Fell, Heather Shapiro, Andrea Lanes, Ann E Sprague, Moya Johnson, Mark Walker, Laura M Gaudet

Title: The Canadian Assisted Reproductive Technologies Register (CARTR) Plus database: A validation study

Author contributions:

I created the study protocol and developed the analysis plan with consultation from D.B. Fell, A.E. Sprague. I conducted all analyses, interpreted the results and wrote the manuscript.

D.B. Fell, H. Shapiro, A. Lanes, M. Walker, L.M. Gaudet assisted with interpreting the results and editing the manuscript. A.E. Sprague assisted with editing the manuscript. M. Johnson assisted with data collection and editing the manuscript.

Special approvals:

A. Privacy training

- I completed privacy training at BORN Ontario to gain access to data in CARTR Plus.

B. Research ethics board (REB) approvals

- REB applications were submitted to both the Children's Hospital of Eastern Ontario (CHEO) and The Ottawa Hospital (TOH).
- As a quality improvement project, CHEO provided us with an exemption (Protocol #: 16/141X).
- REB approval was obtained from TOH (Protocol #: 20160862-01H).

Abstract

Background

Maternal and fetal complications of pregnancy, such as ectopic pregnancy, preterm birth, placenta previa, and preeclampsia, are known to occur with greater frequency among women who conceive using assisted reproductive technology (ART) compared with those who conceive naturally. Despite the increased relative risks, since the absolute incidence of these conditions remains low researchers often rely on large cohorts to evaluate complications following the use of ART. Unfortunately, large prospective cohort studies are both time-consuming and expensive to conduct, therefore, administrative databases and registry databases are often used instead. These sources are excellent sources of data for research purposes as they are relatively inexpensive, easily accessible and are collected on a large population scale. Routinely-collected data are generally not collected with the intention of performing research and, therefore, may introduce information bias due to variable misclassification. To provide insight into data quality and inform the extent to which this may be a concern, database validation studies are highly recommended. Notwithstanding increasing utilization of administrative databases and registries in research investigating pregnancy outcomes of fertility treatments, there is a paucity of validation studies in the literature for these routinely-collected data. The objective of this thesis was to perform two studies, a systematic review of database validation studies among fertility populations and a validation study of the Canadian Assisted Reproductive Technologies Register.

Methods

Study 1: We conducted a systematic review to identify validation studies of databases that contain routinely-collected data in populations using ART. In addition to searching Medline,

Embase and CINAHL for relevant literature, we also examined webpages of international ART surveillance programs and databases. Database managers for national ART registries were contacted to obtain unpublished reports of data quality assurance practices. Screening of articles was performed in two steps by two independent reviewers. Data from relevant articles were extracted and results were synthesized qualitatively.

Study 2: Using patient chart reabstraction for *in vitro* fertilization cycles performed in 2015, we then performed a validation study of the Canadian national ART registry, CARTR Plus. Clinics spanning Canada were recruited to participate. We selected twenty-five data elements from CARTR Plus that were deemed clinically relevant. These data were reabstracted from a random sample of patient charts at each of the participating clinics, (considered the reference standard), and compared to those in the database. We calculated agreement between the two data sources using sensitivity, specificity, positive and negative predictive values, kappa and intraclass correlation coefficients.

Results

Nineteen studies met the inclusion criteria for the systematic review, of which one was a validation of a national registry. Seven studies used an ART database to validate another administrative database or maternal questionnaires, while four studies used either maternal questionnaire or an administrative database to validate an element within a registry. Prevalence estimates were generally not reported well for the data variable of interest.

In the CARTR Plus validation study, agreement between reabstracted data and the CARTR Plus was excellent for most of the investigated variables. There were a few variables — namely follicle stimulating hormone (FSH) level, clinical reason for treatment cycle, oocyte origin and elective embryo transfer — that had moderate agreement.

Conclusion

National fertility databases and registries are used for research and feedback to clinics and government and, therefore, the accuracy of these data is essential. Our validation study provides valuable information for Canadian research studies using the database, and can serve as a guide for other databases to perform quality assurance projects.

Chapter 1. Prologue

1.1 Background

1.1.1 Assisted reproductive technologies

Assisted reproductive technologies (ART) involve the manipulation of gametes to achieve pregnancy (1). ART procedures may involve intrauterine insemination, ovarian stimulation with medications, in vitro fertilization (IVF), and fertility preservation (2). IVF includes ovarian stimulation to allow for ovulation of multiple follicles, followed by fertilization of the oocyte with sperm, which, in the appropriate culture media, develops into an embryo. Upon adequate development, the embryo will either be transferred back into the uterus or cryopreserved for storage (3).

Over one million ART cycles per year are initiated worldwide (4). Reasons for undergoing treatment cycles can be attributed to infertility as a result of male factor (poor sperm quality) and female factor (ovulatory dysfunction, tubal and pelvic pathology) accounting for 35% and 50% of causes of infertility, respectively (5). For many infertile couples, the reason for treatment is poorly understood and the exact incidence of causes of infertility varies by population of interest (6). Other indications for treatment include fertility preservation prior to gonadotoxic treatments or for delayed childbearing, allowance of same-sex couples or those without partners to conceive, and for utilization of a gestational surrogate (2).

An estimated 1.1 million babies were born from IVF initiated between 2008 and 2010 globally (4). Pregnancies conceived using ART have been demonstrated to be associated with increased risks of adverse outcomes in ART pregnancies, compared to those conceived naturally, including preterm birth and low birth weight (7,8). It is therefore essential that we accurately monitor these outcomes and optimize treatment strategies to reduce these risks.

1.1.2 Routinely collected data

Routinely collected data are data that are collected in large volumes for billing, surveillance, development of registries, insurance purposes, or other agencies (9,10). They are not collected with the objective of performing research. These data are usually collected indiscriminately on all people who meet the inclusion criteria of that database, and are therefore robust sources of data for research as a secondary use (11). Consequently, the risk of attrition and selection bias is minimized. Compared to questionnaire and survey data, the risk of non-response, reporting and recall bias is also circumvented among administrative health data and registries.

While routinely collected data are collected broadly, they are often collected at tertiary centres, and therefore generalizability may be limited (if they are collected from specialized tertiary centres, for example). Hospital or health administrative databases collect information based on their inclusion criteria, which may be an event (birth, death, hospital admission), which do not necessarily provide relevant risk factor data. When these data are used secondarily for research, important confounding variables may be missing (11). They are also subject to misclassification errors due to incorrect data entry or lack of internationally-recognized definitions (12). In 2007, the Canadian Health Quality Council advocated for the development of strategies to improve the quality of routinely-collected data to better inform policy makers and researchers (13).

1.1.3 Validation studies

Validation of databases or key elements are necessary both to inform the study design of research projects and to determine the extent to which key variables are accurately documented in the database (14). The various methods that exist include comparing variable codes within the database to information in the medical record (15). When multiple databases are linked on a

unique patient identifier or on multiple data elements, the quality of linkage should also be ascertained. The assessment can be done by either comparing patient records from the linkage to the patient chart or by searching for logical inconsistencies between the two data sources (16,17). Finally, to improve the validity of specific treatments or diagnoses within large databases, multiple codes may be combined to create an algorithm which reduces the risk of misclassification bias (18). These algorithms need to be validated to ensure adequate capture of the element of interest, as their accuracy may vary with specific patient variables including age and hospital (19).

1.1.4 CARTR Plus

The Canadian Assisted Reproductive Technologies Register (CARTR) Plus database is the only national database in Canada collecting information on in vitro fertilization cycles with respect to diagnoses, treatments and outcomes. The CARTR Plus database is administered by Better Outcomes Registry & Network (BORN) Ontario and has been collecting national data from cycles since 2013 from 97% of Canadian ART clinics, but it has yet to be validated.

1.2 Research objectives

This thesis is comprised of two studies. The first is a systematic review of the ART literature investigating how national ART registries and large administrative databases are validating their data. The objectives of this study were to determine how national ART registries and databases using fertility populations are validating their data with respect to the variables chosen, methodology and adherence to guidelines on validation studies. We were specifically interested in the rationale for choosing particular data elements, the methodology utilized in determining the validity of those data, and any troubleshooting procedures that were undertaken to optimize the validity.

The systematic review is meant to serve as an introduction to the second study, which is a validation of the Canadian Assisted Reproductive Technologies Register (CARTR) Plus database. We examined IVF cycles from January 1, 2015 to December 31, 2015 using patient chart reabstraction. The objective of this study was to provide the first step in determining the accuracy and reliability of selected data variables in our national ART registry.

1.3 Rationale

Clinical practice changes should be implemented based on accurate and reliable data. High quality data is essential for providing best practice medicine and superior care to patients. Moreover, the Ontario government has been providing funding for in vitro fertilization for couples since 2015 in the amount of 70 million dollars annually (20). Policy decisions related to allocation of provincial funding for fertility treatments and health care for higher risk pregnancies are dependent on reports provided using these data and, therefore, accuracy of information is critical.

1.4 References

1. Fritz M, Speroff L. Assisted reproductive technologies. In: Clinical gynecologic endocrinology and infertility. Philadelphia: Lippincott Williams & Wilkins; 2010. p. 1331–1382.
2. Okun N, Sierra S. Pregnancy outcomes after assisted human reproduction. J Obstet Gynaecol Can. 2014;36(1):64–83.
3. AbdelHafez FF, Desai N, Abou-Setta AM, Falcone T, Goldfarb J. Slow freezing, vitrification and ultra-rapid freezing of human embryos: a systematic review and meta-analysis. Reprod Biomed Online. 2010;20(2):209–222.
4. 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):1588–1609.
5. Fritz M, Speroff L. Female infertility. In: Clinical gynecologic endocrinology and infertility. Philadelphia: Lippincott Williams & Wilkins; 2010. p. 1137–1190.
6. Templeton AA, Penney GC. The incidence, characteristics, and prognosis of patients whose infertility is unexplained. Fertil Steril. 1982;37(2):175–182.
7. Qin J, Liu X, Sheng X, Wang H. Assisted reproductive technology and the risk of pregnancy-related complications and adverse pregnancy outcomes in singleton pregnancies: A meta-analysis of cohort studies. Fertil. Steril. 2016;105(1):73–85e6.
8. Qin J, Wang H, Sheng X, Liang D, Tan H, Xia J. Pregnancy-related complications and adverse pregnancy outcomes in multiple pregnancies resulting from assisted reproductive technology: a meta-analysis of cohort studies. Fertil Steril. 2015;103(6):1492–1497.
9. Iezzoni LI. Assessing Quality Using Administrative Data. Ann Intern Med [Internet] 1997 [cited 2016 Jul 25];127(8_Part_2):666. Available from: http://annals.org/article.aspx?doi=10.7326/0003-4819-127-8_Part_2-199710151-00048
10. Buttner P, Muller R. Epidemiology. 2nd editio. South Melbourne, Australia: Oxford University Press; 2015.
11. Jorm L. Routinely collected data as a strategic resource for research: priorities for methods and workforce. Public Heal Res Pract [Internet] 2015 [cited 2018 Aug 28];25(4). Available from: http://www.phrp.com.au/issues/september-2015-volume-25-issue-4/routinely-collected-data-as-a-strategic-resource-for-research-priorities-for-methods-and-workforce/
12. 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):2683–2687.
13. Health Council of Canada. Health care renewal in Canada: Measuring up? [Internet]. 2007 [cited 2016 Jul 25];Available from: http://tools.hhr-rhs.ca/index.php?option-com_mtree&task=att_download&link_id=10453&cf_id=68&lang=en
14. Grimes DA. Epidemiologic research using administrative databases. Obstet Gynecol. 2010;116(5):1018–1019.
15. Iron K, Manuel D. Quality assessment of administrative data (QuAAD): an opportunity for enhancing Ontario's health data. Inst Clin Eval Stud 2007.
16. Silveira DP da, Artmann E. Accuracy of probabilistic record linkage applied to health databases: systematic review. Rev Saude Publica. 2009;43(5):875–882.

17. Tromp T, Ravelli ACJ, Meray N, Reitsma JB. An efficient validation method of probabilistic record linkage including readmissions and twins. Methods Inf. Med. 2008;47(4):356–363.
18. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. J Clin Epidemiol. 2010;64(2011):821–829.
19. Benchimol EI, Guttmann A, Mack DR, Nguyen GC, Marshall JK, Gregor JC, et al. Validation of international algorithms to identify adults with inflammatory bowel disease in health administrative data from Ontario, Canada. J Clin Epidemiol. 2014;67(8):887–896.
20. Government of Ontario. Newsroom : Ontario Announces 50 Clinics Offering Government-Funded Fertility Treatments [Internet]. [cited 2018 Jan 5];Available from: https://news.ontario.ca/mohltc/en/2015/12/ontario-announces-50-clinics-offering-government-funded-fertility-treatments.html

Chapter 2. Literature Review: [Manuscript 1] A systematic review of database validation studies among fertility populations

2.1 Preface to Manuscript 1

We performed a systematic review to answer the following questions:

1. How are assisted reproductive technology (ART) centres validating their databases?
2. Are they reporting in accordance with the RECORD reporting statement?

2.2 Title page

Title: A systematic review of database validation studies among fertility populations

Running title: Systematic review of ART database validity studies

Authors: Bacal, V[1,2,3]; Russo, M[4,5]; Fell, DB[1,6,7]; Shapiro, H[4,5]; Walker, M[1,2,3]; Gaudet, LM[1,2,3]

[1] School of Epidemiology and Public Health (SEPH) / École d'épidémiologie et de santé publique (EESP), University of Ottawa, 600 Peter Morand Crescent, Ottawa, ON K1G 5Z3

[2] Department of Obstetrics and Gynecology, University of Ottawa, 501 Smyth Road, Ottawa, ON K1H 8L6

[3] Ottawa Hospital Research Institute, 501 Smyth Road, Ottawa, ON K1H 8L6

[4] Department of Obstetrics and Gynecology, University of Toronto, 123 Edward Street, Suite 1200, Toronto, ON M5G 1E2

[5] Mount Sinai Fertility, 250 Dundas Street West, 7th Floor, Toronto, ON M5T 2Z5

[6] Children's Hospital of Eastern Ontario Research Institute, 401 Smyth Road, Ottawa, ON K1H 5B2

[7] Institute for Clinical Evaluative Sciences, 1053 Carling Avenue, Ottawa, ON K1Y 4E9

Corresponding author: Vanessa Bacal

Financial support: None

Funding: None

2.3 Abstract

Introduction

Routinely-collected data, including administrative databases and registries are excellent sources of data. However, these data are subject to misclassification bias due to misdiagnosis or errors in data entry, and therefore need to be validated prior to utilization for clinical or research purposes.

The objective of this study was to assess whether routinely-collected data from fertility populations are adequately validated.

Methods

We conducted a systematic review by searching Medline, Embase and CINAHL from inception to October 6, 2016. Webpages of international assisted reproductive technology (ART) centres were also searched. Keywords and MeSH terms were adapted from previous systematic reviews. Only full-text studies in English were included. Studies were excluded if they did not validate a fertility database or registry. We used the items described by RECORD as important to report as a guide to evaluate whether included studies used rigorous methodology to conduct their validation.

Results

Nineteen studies were included in this review. Two studies validated a fertility database using medical records; seven studies used an ART registry to validate vital records or maternal questionnaires and two studies failed to adequately describe their reference. Seven studies reported the pre-test prevalence of the variable validated; however, only four studies had post-test prevalence estimates within a 2% range of the pre-test estimate.

Discussion

There is a paucity of literature on validation of routinely-collected data from ART populations. Furthermore, the prevalence of the markers validated are not being presented, which can lead to biased estimates. Stakeholders rely on these data for outcomes of treatments, therefore it is essential to ascertain the accuracy of these databases.

2.4 Significance

What is already known on this subject?

Due to the increased risk of adverse outcomes in ART pregnancies compared with naturally conceived pregnancies, it is essential to accurately monitor these outcomes. Clinicians and researchers often rely on data from large databases and registries to conduct these reports and studies. However, if not adequately validated, utilization of these data leads to misclassification bias and unmeasured confounding.

What this study adds?

This study demonstrates a gap in fertility research where accuracy of routinely-collected data is not well-described. Future validation studies need to be conducted for greater transparency in accuracy of research.

2.5 Introduction

Infertility burdens 1.9% to 10.5% of child-seeking women world-wide and was estimated to affect 48.5 million couples in 2010 (1). According to the International Committee for Monitoring Assisted Reproductive Technologies, 1.4 to 1.6 million assisted reproductive technology (ART) cycles were initiated per year from 2008 to 2010 resulting in approximately 800,000 babies born over this time period (2). ART is a rapidly evolving field in medicine with new advances in research and technology. From freezing techniques of gametes and embryos (3–5), to the number of embryos replaced (5,6), and utilization of pre-implantation genetic diagnosis (7), reproductive technologies and guidelines are changing regularly. It is, therefore, prudent to ensure that we can adequately monitor treatment outcomes and adverse events. Studies from the United States and Europe estimate that the prevalence of live births born after in vitro fertilization (IVF) ranges from 1-6% (8,9). The risk of adverse obstetrical events is significantly higher in ART- compared to naturally-conceived pregnancies (10–12). However, the attributable proportion of these complications due to ART, such as ectopic pregnancies, placenta previa and congenital anomalies, is low, with estimates around 1-2% (13,14), 1.6% (15), and 8% (16), respectively. Similarly, neonatal outcomes including small for gestational age, preterm delivery and admission to a critical unit also occur infrequently (10–12). Therefore, in order to adequately understand the implications of such treatment, studies using large sample sizes are required.

Routinely collected data, such as administrative databases and registries are robust sources of data. These databases often contain sociodemographic information, health care utilization, treatment and diagnostic information affiliated with health care visits. However, these data are not collected for a specific research question and are prone to error due to clerical errors, illegible charts and documentation problems (17). If not validated adequately, utilization of these

data for surveillance, quality improvement and research can lead to misclassification bias and unmeasured confounding due to missing data (18).

Many studies that use large administrative and registry databases to identify patients who undergo ART treatments indicate that they are using validated data (19–21). However, the literature is scarce on validation studies and measures performed to ensure accuracy among these databases. There is extensive literature indicating the importance of presenting measures of validity, including sensitivity, specificity, and positive predictive values to reflect whether these data can be reliably used for research and reporting (22–25). With this in mind, we conducted a systematic review to identify validation studies of databases that contain these routinely-collected data (including administrative data and registry data) in an ART setting. Our primary objective was to assess how ART centres validate and report their fertility data, their rationale for choosing specific data elements for validation activities, the extent to which a database is considered valid for use, and actions taken when validity was deemed poor. Our secondary objective was to investigate whether ART centres were reporting their validation studies in accordance with the published reporting guidelines for validation studies (22) with details pertaining to the method of validation and quality control, the variables chosen to validate the database and the outcome measures.

2.6 Methods

This review was conducted in accordance with a protocol developed and registered a priori (International Prospective Register of Systematic Reviews ID: CRD42016048466). This manuscript is not based upon clinical study or patient data. Studies were selected according to the following PICOS criteria:

Population: Fertility patients who have undergone in vitro fertilization (IVF) cycles

Intervention: Chart re-abstractions or self-reported survey data

Comparator: Data from administrative or registry databases

Outcome: Sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratio, kappa coefficient, area under the ROC curve or c-statistic, accuracy or agreement of the selected data elements

Study design: Validation studies of specific codes, or case-finding algorithms within fertility databases or registries; linkage studies between two or more databases that include a fertility registry. Large administrative or registry databases are defined as those that collect data routinely without an a priori research question.

Studies were excluded if they did not use a fertility database (which includes cycle information, specific patient diagnoses, etc.). Only full text articles published in English were considered.

The search strategy was developed with the aid of an information specialist with expertise in clinical research, adapted from previous systematic reviews (22,26). Electronic bibliographic databases, specifically, Medline, Embase and CINAHL were searched using specific vocabulary and MeSH keywords (see Appendix 1 for Medline search strategy). Reference lists of all included articles and relevant systematic reviews were screened to identify additional studies. Webpages for major international fertility surveillance systems were searched to account for validation activities presented within surveillance reports, which are typically not indexed in bibliographic databases. We also contacted these surveillance programs to request reports that were not publicly available. These programs included https://www.belrap.be/Public/Default.aspx?Lg=En (Belgium), https://www.sart.org/ (United States of America), https://www.asrm.org/about-us/contact-us/ (American Society of

Reproductive Medicine), https://npesu.unsw.edu.au/data-collection/australian-new-zealand-assisted-reproduction-database-anzard (Australia and New Zealand), https://www.hfea.gov.uk/ (United Kingdom), https://www.eshre.eu/Home/Contact-us.aspx (European Society of Human Reproduction and Embryology) https://www.icmr.gov.in/icmrnews/art/contact_us.htm (India). Citations were imported into EndNote and managed within Covidence (www.covidence.org). This process was recorded using the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram (27). We performed the final search on October 6, 2016. As validating the dataset is often a secondary objective of studies using routinely collected data, and would therefore not be indexed in MeSH terms, titles or abstracts, this search strategy could not capture all relevant validation studies.

Screening was performed in two steps by two independent reviewers (V.B and M.R.) using the eligibility criteria. Title and abstract screening was performed initially, followed by full-text screening. Disagreements were resolved by consensus or through consultation with a senior expert where consensus could not be reached. Reasons for excluding studies in the full-text screening step were documented.

We extracted data from each included study on country of origin, year of publication, number of clinics involved, number of treatment records, sample size calculation, variables or algorithms used, method of validation (chart review versus survey of patients versus another validated database), whether datasets were linked, how datasets were linked (probabilistic versus deterministic) and validation outcome measures (listed above). Two independent reviewers extracted these data in duplicate.

We used the items described by previously published guidelines for validation studies as important to report as a guide to evaluate whether included studies used rigorous methodology to conduct their validation (22) which was implemented by two independent reviewers. We made a

post-hoc decision after protocol registration to adapt quality assessment tools used by two previously published systematic reviews to assess both reporting and quality of studies (22,28). All results were synthesized qualitatively.

2.7 Results

The electronic search yielded 1074 citations after removing duplicates. Upon applying the inclusion and exclusion criteria, we identified 77 studies for full-text screening after title and abstract screening. Seven additional studies were identified for full-text screening after reviewing the references of pertinent articles and searching webpages. Of these 84 studies, 53 did not meet inclusion criteria for various reasons including wrong study design, comparator or patient population (details can be found in Appendix 2). Nineteen studies were included for final analysis (Figure 1), representing the USA (29–41), Finland (42,43), Denmark (44), the Netherlands (45), Israel (46), and the UK (47). Four studies did not use any reference standard (29,42,43,47), and the reference was poorly described in two studies (30,46) (Table 1). Two studies used medical records to validate a fertility database (30,38); seven studies used an IVF registry as the reference standard to validate either vital records or maternal questionnaires (32,34,36,37,39,44,47); one study utilized maternal report as the reference standard for validation (33) and three studies used vital records (birth and death certificates) as the reference standard (35,40,45). Finally, one study used both IVF registries and vital records as the reference standard depending on the data element validated (41).

Four studies validated method of conception from birth registries (31,34,39,43), two validated diagnoses or treatment variables within the fertility database (30,38), one study created an algorithm to identify a patient population (42), and four studies validated linkage algorithms between a fertility and a second administrative database (29,32,35,47).

Sensitivity was the most commonly reported validation measure. Twelve studies reported sensitivity (31–37,39–41,44,45), nine reported specificity (31,33–37,39,44,45), six reported positive predictive value (31,34,35,37,44,45), one reported negative predictive value (37), five reported the Kappa coefficient (33,35,40,43,45), and seven reported percent agreement (32,33,40,41,43–45) (Table 2). The data quality measures are presented in Appendix 3. Only three studies reported 4 or more measures of validation (33,37,44). Nine studies presented 95% confidence intervals with the estimates (31,32,34,36–38,40,43,45), of which 5 reported confidence intervals for all estimates (32,34,36–38).

The elements of data quality are summarized in Tables 3 and 4. Sixteen studies (84.2%) adequately described their data source and all but one described the type of patient records from which data were extracted (46). The studies predominantly described inclusion and exclusion criteria and their methods for determining the validity of the data. Fifteen studies adequately described their method of patient sampling; while fourteen studies sampled the entire population in the database (29,31–37,39,40,42,44,45,47), one study performed a random sampling strategy (38). Only one group performed their study using an a priori sample size (38) and none provided statistical justification for their sample size.

Where multiple databases were linked using a common patient identifier, the linkage procedures were adequately described in eight (53.3%) of the studies (29,31,32,34,35,40,41,47). The quality of these procedures was described in only seven studies (46.7%) (29,31,32,35,40,42,47).

The pre-test prevalence of the validated variables was provided in seven studies (29,31,33–36,39) (Table 5). The post-test prevalence of these variables was within a 2% range of

the pre-test values for four of the studies (31,34–36); however, in two studies, the post-test prevalence was largely discrepant from pre-test values (33,39).

2.8 Discussion

This study demonstrates there is a paucity of literature on validation of data elements within fertility databases and registries. While there were numerous studies that validated ART information either by maternal report or birth and death certificates by using a fertility database as a reference standard, there was only one study available that demonstrated the validity of a fertility database, which was only obtained after searching websites and contacting the Centre for Disease Control for information on data collection (38). Furthermore, only seven studies published the baseline prevalence of the data element being validated (29,31,33–36,39), of which only 4 studies' sample prevalence approximated that of the population (33,39).

Many studies rely on routinely collected data from fertility registries and birth registries to investigate the implications of assisted reproductive technology. These databases are excellent sources to understand how IVF treatments can lead to specific outcomes, particularly those with low prevalence. Further, as many of these databases collect data on a national or state-wide level at each patient visit, there is little attrition compared to data that would have been collected as part of a prospective cohort study, where loss-to-follow up is a substantial issue. Importantly, these studies are used to inform government policy, guide treatment of patients and stimulate the design of research projects. Therefore, it is essential to ensure the data are accurate and valid.

There is insufficient documentation in the literature with respect to how national fertility registries are validating their databases. The Society of American Reproductive Technologies (SART) publishes a publicly-available report on an annual basis indicating which variables are discrepant between the medical chart and the database (38). However, none of the other national

databases have generated such reports. For example, Human Fertilisation and Embryology Authority (HFEA) in the United Kingdom, Australian & New Zealand Assisted Reproduction Database (ANZARD) and the Belgian Register for Assisted Procreation endorse strict adherence to quality assurance practices; however, no reports were available describing their data validation processes (written communication with Belgium and ANZARD). As all stakeholders, including patients, health care practitioners, researchers, and policy makers, rely on these data to understand the implications of fertility treatments, including the prevalence of disease, practice patterns, complications and outcomes of ART, it is essential that these reports are made publicly available (48–52).

There are a variety of methods that can be used to validate a database or data element of interest. One such way is to validate the codes against the medical record. A study by Dunn et al. (2011) validated a Canadian perinatal database to assess its accuracy (53). Abstractors collected data from patient charts which were subsequently compared to the codes in the database. This methodology allows multiple variables to be assessed at one time and patterns in misclassified charts can be identified and can guide quality assurance projects to improve the accuracy of specific problematic variables. Similarly, the study performed by SART assessed multiple variables at one time comparing SART data to patient charts (38). However, due to the presentation of discrepancy rates without other important measures of validity, such as sensitivity, kappa coefficients or positive predictive values, it is difficult to determine how reliable these data are. A subgroup evaluation by the size of the clinic or geography would be useful to investigate whether specific variables are largely problematic or if there is an issue at a specific clinic. A Canadian study investigating the validity of diagnostic codes in 10 major hospitals found that the sensitivity and specificity were highly dependent on the hospital, where

some had a high accuracy and others demonstrated poor sensitivity (54). Clinics may have specific expertise with respect to their patient populations, and the prevalence of certain conditions or treatments may vary based on health care provider. Predictive tests (PPV, likelihood ratios) are highly dependent on the baseline prevalence of the specific treatment or disease (55). Furthermore, in certain cases, the sensitivity and specificity may vary with the prevalence (56). Only four of the included studies presented post-test prevalence estimates that approximated the reported pre-test prevalence; it therefore puts into question the degree of bias in the estimates presented. As such, it is essential to describe both the source of data and prevalence of the variable of interest to adequately interpret the results.

Particularly when the validity of a specific code is questionable, an algorithm combining different codes can be created to improve reliability and reduce the risk of misclassification bias (22). A recent study by Castro et al. (2015) validated the diagnosis of polycystic ovarian syndrome (PCOS) in a Massachusetts administrative database using International Classification of Disease Version 9 (ICD9) codes (57). An algorithm was created to identify patients with PCOS based on age, gender and other variables. The positive predictive value of the algorithm for definite PCOS was 74% indicating among those with an ICD9 code for PCOS, over one quarter of the patient population would be misclassified. Our review included only one study that designed an algorithm to identify an ART population in Finland, likely due to our restricted search strategy for specific diagnoses in reproductive endocrinology and infertility (42). However, due to privacy laws in Finland, the investigators were unable to validate the accuracy of their algorithm using patient records and must be interpreted cautiously if used in future research.

Studies that link pairs or duplicated charts from two different files or databases are defined as record linkage studies (55). Record linkage can be performed to gather information from more than one source of data for individual patients, providing a more complete clinical history. The linkage can be performed deterministically using a unique patient identifier that is common to both databases. More frequently, however, the linkage is performed probabilistically using multiple data elements such as name, date of birth and date of delivery (58,59). The accuracy of the linkage procedure can be ascertained by comparing the records to an external gold standard (58,60). This methodology is not always feasible, particularly if there is no gold standard available. Other approaches to validating the linkage include either assessment for logical inconsistencies or evaluation of the variables in each database to check for discrepancies between the two (60). Moreover, the validity of this procedure is dependent on the quality of the databases that are linked and how well they are maintained (61). Zhang et al. (2012) validated a probabilistic linkage of ART cycle data from the National ART Surveillance System with birth certificates in a separate database. The linkage performed by Sunderam et al. (2006) using the same population served as their gold standard for validation. However, Sunderam's group validated their study results by assessing for inconsistencies rather than validating against an external reference. Williams et al. (2013) also checked for logical inconsistencies to validate their linkage between their ART and cancer registries. Although external validation requires more time and is more costly, it is the only method of definitively identifying false positive and false negative links, which can also be used for quality improvement of the linkage procedure (60).

Ideally, a validation study uses a gold standard as a measure to guide the accuracy and reliability of the validated variable. Based on the Standards for Reporting of Diagnostic

Accuracy Studies (STARD) guidelines for evaluating diagnostic tests, a gold standard should be the best available test at identifying the condition of interest (62). To this end, the gold standard of determining accuracy of database codes has not been established (18,23,54,63). In the absence of a true gold standard, some argue that the medical record should serve as the reference standard (64,65). Of the included studies in our review, only two used the medical record as the reference standard (30,38) and only one presented measures of validation (38). The others used another database or patient report as the reference. Relying on the medical record is subject to having the physician correctly identify and document the condition(s) validated. Due to the difficulty in diagnosing certain conditions as well as consistently reporting them in the patient's chart, the medical record is susceptible to misclassification bias and may not be an ideal reference standard (66). However, as previously mentioned, using an administrative database or registry as a gold standard relies on that database to be validated. It is clear from this review, that while organizations state their databases are audited, tracking that process and determining which codes are reliable is challenging. Therefore, a gold standard from this source should not be implicitly accepted. Finally, patient report is subject to recall bias, particularly as increasing time has passes from the event to the survey (66).

Our review has several limitations. We restricted our inclusion criteria to published reports in English. As many of the internal processes are likely to occur in the primary language of the registry or organization, it is possible that we were unable to capture validation processes from registries. A comprehensive search on the internet did not yield any results, even in other languages, however. Moreover, only 4 studies were excluded from our database search due to language restriction (67–70). Our study was also limited by the search strategy developed for Medline, Embase and CINAHL. While the strategy was quite general for routinely collected

databases, the list was not exhaustive for specific diagnoses relevant to infertility. Consequently, it is probable that other published studies were not captured in our review.

In spite of these limitations, our study is strengthened by the systematic and comprehensive approach to searching the articles and analyzing the measures of validity. This is the first study to our knowledge to assess the validity of fertility registries. Although many of these reports were not published in indexed bibliographic databases, numerous attempts were made to contact ART surveillance database managers in the UK, Denmark, Belgium, Australia, New Zealand and the United States to obtain unpublished or ad-hoc reports on data maintenance and quality assurance. Additionally, all of the studies included in our review were published in the last 15 years, which is of particular importance given that data completeness tends to improve over time (71).

This review highlights an important gap in the field of fertility research where the validation of largely utilized databases has not been well-described. The national fertility databases and registries are used for research and feedback to clinics and government and, therefore, the accuracy of these data is essential. Furthermore, during the validation process, the prevalence of the variables and the statistical estimates need to be adequately measured. The RECORD reporting guidelines recommend at least 4 measures of validity for each variable assessed, along with an estimate of the error associated with each measure (18). As the prevalence of the condition varies based on health care provider or geographic location, so will these measures. Future studies need to be conducted and published using rigorous methodology that will allow for greater transparency in the accuracy of research within this rapidly evolving field of medicine and research.

Acknowledgments

We would like to thank Risa Shorr who helped in the development of the search strategy.

Funding

This study was not funded

Conflicts of interest

None of the authors have any conflicts of interest to declare

2.9 References

1. Mascarenhas MN, Flaxman SR, Boerma T, Vanderpoel S, Stevens GA. National, regional, and global trends in infertility prevalence since 1990: A systematic analysis of 277 health surveys. Low N (ed). PLoS Med 2012;9:e1001356.
2. Dyer S, Chambers GM, de Mouzon J, Nygren KG, Zegers-Hochschild F, Mansour R, Ishihara O, Banker M, Adamson GD. International Committee for Monitoring Assisted Reproductive Technologies world report: Assisted reproductive technology 2008, 2009 and 2010. Hum Reprod 2016;31:1588–1609.
3. AbdelHafez FF, Desai N, Abou-Setta AM, Falcone T, Goldfarb J. Slow freezing, vitrification and ultra-rapid freezing of human embryos: a systematic review and meta-analysis. Reprod Biomed Online 2010;20:209–222.
4. Loutradi KE, Kolibianakis EM, Venetis CA, Papanikolaou EG, Pados G, Bontis I, Tarlatzis BC. Cryopreservation of human embryos by vitrification or slow freezing: a systematic review and meta-analysis. Fertil Steril 2008;90:186-193.
5. National Institute for Health and Care Excellence. Fertility problems: assessment and treatment. Natl Inst Heal Care Excell 2013. Available at: https://www.nice.org.uk/guidance/cg156/resources/fertility-problems-assessment-and-treatment-35109634660549. Accessed July 27, 2016.
6. Committee of the Society for Assisted Reproductive Technology P, Committee of the American Society for Reproductive Medicine P. Elective single-embryo transfer. Fertil Steril 2012;97:835–842.
7. Harton G, Braude P, Lashwood A, Schmutzler A, Traeger-Synodinos J, Wilton L, Harper JC, European Society for Human Reproduction and Embryology (ESHRE) PGD Consortium. ESHRE PGD consortium best practice guidelines for organization of a PGD centre for PGD/preimplantation genetic screening. Hum Reprod 2011;26:14–24.
8. Sunderam S, Kissin DM, Crawford SB, Folger SG, Jamieson DJ, Warner L, Barfield WD. Assisted reproductive technology surveillance — United States, 2014. MMWR Surveill Summ 2017;66:1–24.
9. Sullivan EA, Zegers-Hochschild F, Mansour R, Ishihara O, De Mouzon J, Nygren KG, Adamson GD. International Committee for Monitoring Assisted Reproductive Technologies (ICMART) world report: assisted reproductive technology 2004. Hum Reprod 2013;28:1375–1390.
10. Qin J, Liu X, Sheng X, Wang H. Assisted reproductive technology and the risk of pregnancy-related complications and adverse pregnancy outcomes in singleton pregnancies: A meta-analysis of cohort studies. Fertil Steril 2016;105:73–85e6.
11. McGovern PG, Llorens AJ, Skurnick JH, Weiss G, Goldsmith LT. Increased risk of preterm birth in singleton pregnancies resulting from in vitro fertilization–embryo transfer or gamete intrafallopian transfer: A meta-analysis. Fertil Steril 2004;82:1514–1520.
12. Sazonova A, Källen K, Thurin-Kjellberg A, Wennerholm U-B, Bergh C. Obstetric outcome after in vitro fertilization with single or double embryo transfer. Hum Reprod 2011;26:442–450.
13. Santos-Ribeiro S, Tournaye H, Polyzos NP. Trends in ectopic pregnancy rates following assisted reproductive technologies in the UK: a 12-year nationwide analysis including 160 000 pregnancies. Hum Reprod 2016;31:393–402.
14. Perkins KM, Boulet SL, Kissin DM, Jamieson DJ, National ART Surveillance (NASS)

Group the NAS (NASS). Risk of ectopic pregnancy associated with assisted reproductive technology in the United States, 2001-2011. Obstet Gynecol 2015;125:70–78.
15. Romundstad LB, Romundstad PR, Sunde A, von Düring V, Skjaerven R, Vatten LJ. Increased risk of placenta previa in pregnancies following IVF/ICSI; a comparison of ART and non-ART pregnancies in the same mother. Hum Reprod 2006;21:2353–2358.
16. Davies MJ, Moore VM, Willson KJ, Van Essen P, Priest K, Scott H, Haan EA, Chan A. Reproductive technologies and the risk of birth defects. N Engl J Med 2012;366:1803–1813.
17. Hierholzer Jr WJ. Health care data, the epidemiologist's sand: Comments on the quantity and quality of data. Am J Med 1991;91:21S–26S.
18. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLoS Med 2015;12.
19. Fedder J, Loft A, Parner ET, Rasmussen S. Neonatal outcome and congenital malformations in children born after ICSI with testicular or epididymal sperm: A controlled national cohort study. Hum Reprod 2013;28:230–240.
20. Traeger-Synodinos J, Coonen E, Goossens V. Data from the ESHRE PGD Consortium. Hum Reprod 2013;28:i18.
21. De Geyter C, Fehr P, Moffat R, Gruber I, Von W. Twenty years' experience with the Swiss data registry for assisted reproductive medicine: outcomes, key trends and recommendations for improved practice. Swiss Med Wkly 2015;145:w14087. Available at: http://doi.emh.ch/smw.2015.14087. Accessed July 29, 2016.
22. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. J Clin Epidemiol 2010;64:821–829.
23. Sørensen HT, Sabroe S, Olsen J. A Framework for Evaluation of Secondary Data Sources for Epidemiological Research. Int J Epidemiol 1996;25:435–442.
24. Van Walraven C, Austin PC, Manuel D, Knoll G, Jennings A, Forster AJ. The usefulness of administrative databases for identifying disease cohorts is increased with a multivariate model. J Clin Epidemiol 2010;63:1332–1341.
25. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol 2010;69:4–14.
26. Shiff NJ, Jama S, Boden C, Lix LM, Virnig B, McBean M, Walraven C van, Austin P, Walraven C Van, Bennett C, et al. Validation of administrative health data for the pediatric population: a scoping review. BMC Health Serv Res 2014;14:236.
27. Moher D, Liberati A, Tetzlaff J, Altman DG, Grp P. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 2009;6:e1000097.
28. Grams ME, Plantinga LC, Hedgeman E, Saran R, Myers GL, Williams DE, Powe NR, CDC CKD Surveillance Team. Validation of CKD and related conditions in existing data sets: A systematic review. Am J Kidney Dis 2011;57:44–54.
29. Sunderam S, Schieve LA, Cohen B, Zhang Z, Jeng G, Reynolds M, Wright V, Johnson C, Macaluso M, Massachusetts Consortium for Assisted Reproductive Technology Epidemiologic Research (MCARTER). Linking birth and infant death records with assisted reproductive technology data: Massachusetts, 1997–1998. Matern Child Health J

2006;10:115–125.
30. Molinaro TA, Shaunik A, Lin K, Sammel MD, Barnhart KT. A strict infertility diagnosis has poor agreement with the clinical diagnosis entered into the Society for Assisted Reproductive Technology registry. Fertil Steril 2009;92:2088–2090.
31. Zhang Z, Macaluso M, Cohen B, Schieve L, Nannini A, Chen M, Wright V, Research MC for ARTE. Accuracy of assisted reproductive technology information on the Massachusetts birth certificate, 1997-2000. Fertil Steril 2010;94:1657–1661.
32. Zhang Y, Cohen B, Macaluso M, Zhang Z, Durant T, Nannini A. Probabilistic linkage of assisted reproductive technology information with vital records, Massachusetts 1997-2000. Matern Child Heal J 2012;16:1703–1708.
33. Buck Louis GM, Hediger ML, Bell EM, Kus CA, Sundaram R, McLain AC, Yeung E, Hills EA, Thoma ME, Druschel CM. Methodology for establishing a population-based birth cohort focusing on couple fertility and children's development, the Upstate KIDS Study. Paediatr Perinat Epidemiol 2014;28:191–202.
34. Cohen B, Bernson D, Sappenfield W, Kirby RS, Kissin D, Zhang Y, Copeland G, Zhang Z, Macaluso M, States Monitoring Assisted Reproductive Technology Collaborative. Accuracy of assisted reproductive technology information on birth certificates: Florida and Massachusetts, 2004-06. Paediatr Perinat Epidemiol 2014;28:181–190.
35. Kotelchuck M, Hoang L, Stern J, Diop H, Belanoff C, Declercq E. The MOSART Database: Linking the SART CORS Clinical Database to the Population-Based Massachusetts PELL Reproductive Public Health Data System. Matern Child Heal J 2014;18:2167–2178.
36. Liberman RF, Stern JE, Luke B, Reefhuis J, Anderka M. Validating assisted reproductive technology self-report. Epidemiology 2014;25:773–775.
37. Buck Louis GM, Druschel C, Bell E, Stern JE, Luke B, McLain A, Sundaram R, Yeung E. Use of assisted reproductive technology treatment as reported by mothers in comparison with registry data: the Upstate KIDS Study. Fertil Steril 2015;103:1461–1468.
38. Centers for Disease Control and Prevention, American Society for Reproductive Medicine, Society for Assisted Reproductive Technology. 2014 Assisted Reproductive Technology Fertility Clinic Success Rates Report. Atlanta, 2016.
39. Luke B, Brown MB, Spector LG. Validation of infertility treatment and assisted reproductive technology use on the birth certificate in eight states. Am J Obstet Gynecol 2016;215:126–127.
40. Stern JE, Gopal D, Liberman RF, Anderka M, Kotelchuck M, Luke B. Validation of birth outcomes from the Society for Assisted Reproductive Technology Clinic Outcome Reporting System (SART CORS): Population-based analysis from the Massachusetts Outcome Study of Assisted Reproductive Technology (MOSART). Fertil Steril 2016;106:717–722.
41. Stern JE, McLain AC, Buck Louis GM, Luke B, Yeung EH. Accuracy of self-reported survey data on assisted reproductive technology treatment parameters and reproductive history. Am J Obstet Gynecol 2016;215:219.e1-6.
42. Hemminki E, Klemetti R, Rinta-Paavola M, Martikainen J. Identifying exposures of in vitro fertilization from drug reimbursement files: a case study from Finland. Med Inform Internet Med 2003;28:279–289.
43. Gissler M, Klemetti R, Sevón T, Hemminki E. Monitoring of IVF birth outcomes in Finland: a data quality study. BMC Med Inform Decis Mak 2004;4.
44. Hvidtjorn D, Grove J, Schendel D, Schieve LA, Ernst E, Olsen J, Thorsen P. Validation of

self-reported data on assisted conception in The Danish National Birth Cohort. Hum Reprod 2009;24:2332–2340.
45. Overbeek A, van den Berg MH, Hukkelhoven CWPM, Kremer LC, van den Heuvel-Eibrink MM, Tissing WJE, Loonen JJ, Versluys AB, Bresters D, Kaspers GJL, et al. Validity of self-reported data on pregnancies for childhood cancer survivors: a comparison with data from a nationwide population-based registry. Hum Reprod 2013;28:819–827.
46. Rosenfeld Y, Strulov A. Clinical reports on IVF cycle rank--reliability and validity. Harefuah 2009b;148:22–24.
47. Williams CL, Bunch KJ, Stiller CA, Murphy MFG, Botting BJ, Wallace WH, Davies M, Sutcliffe AG. Cancer risk among children born after assisted conception. N Engl J Med 2013;369:1819–1827.
48. Canadian Fertility Andrology Society. CARTR Annual Reports. 2014. Available at: https://cfas.ca/cartr-annual-reports/. Accessed August 12, 2017.
49. Harris K, Fitzgerald O, Macaldowie A, Lee E, Chambers G. Assisted reproductive technology in Australia and New Zealand 2014. Sydney Natl Perinat Epidemiol Stat Unit, Univ New South Wales 2016. Available at: https://npesu.unsw.edu.au/sites/default/files/npesu/surveillances/Assisted reproductive technology in Australia and New Zealand 2014.pdf. Accessed August 12, 2017.
50. Butler P. Progress in reproductive health research. World Heal Organ 2003.
51. Chambers GM, Sullivan EA, Ishihara O, Chapman MG, Adamson GD. The economic impact of assisted reproductive technology: a review of selected developed countries. Fertil Steril 2009;91:2281–2294.
52. Human Fertilisation and Embryology Authority. Annual report and accounts 2015/16. 2016. Available at: http://ifqtesting.blob.core.windows.net/umbraco-website/1183/56071_hc_380_web_v02.pdf. Accessed August 12, 2017.
53. Dunn S, Bottomley J, Ali A, Walker M. 2008 Niday Perinatal Database quality audit: report of a quality assurance project. Chronic Dis Inj Can 2011;32:32–42.
54. Juurlink D, Preyra C, Croxford R, Chong A, Austin P, Tu J, Laupacis A. Enhancing the effectiveness of health care for Ontarians through research Canadian Institute for Health Information Discharge Abstract Database: A Validation Study ICES Investigative Report. Toronto Inst Clin Eval Sci 2006.
55. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. BMJ 1994;309:102.
56. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. Stat Med 1997;16:981–991.
57. Castro V, Shen Y, Yu S, Finan S, Pau CT, Gainer V, Keefe CC, Savova G, Murphy SN, Cai T, et al. Identification of subjects with polycystic ovary syndrome using electronic health records. Reprod Biol Endocrinol 2015;13:1–8.
58. Silveira DP da, Artmann E. Accuracy of probabilistic record linkage applied to health databases: systematic review. Rev Saude Publica 2009;43:875–882.
59. Méray N, Reitsma JB, Ravelli ACJ, Bonsel GJ. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. J Clin Epidemiol 2007;60:883.e1-883.e11.
60. Tromp T, Ravelli ACJ, Meray N, Reitsma JB. An efficient validation method of probabilistic record linkage including readmissions and twins. Methods Inf Med 2008;47:356–363.
61. Blom JW, Vanderschoot JPM, Oostindier MJ, Osanto S, Van Der Meer FJM, Rosendaal FR. Incidence of venous thrombosis in a large cohort of 66 329 cancer patients: results of a

record linkage study. J Thromb Haemost 2006;4:529–535.
62. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HCW, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. Clin Chem 2003;49:1–6.
63. Lain SJ, Hadfield RM, Raynes-Greenow CH, Ford JB, Mealing NM, Algert CS, Roberts CL. Quality of data in perinatal population health databases: a systematic review. Med Care 2012;50:e7-e20.
64. Widdifield J, Labrecque J, Lix L, Paterson JM, Bernatsky S, Tu K, Ivers N, Bombardier C. Systematic review and critical appraisal of validation studies to identify rheumatic diseases in health administrative databases. Arthritis Care Res (Hoboken) 2013;65:1490–1503.
65. Frosst G, Hutcheon J, Joseph K, Kinniburgh B, Johnson C, Lee L. Validating the British Columbia perinatal data registry: a chart re-abstraction study. BMC Pregnancy Childbirth 2015;15:123.
66. Leong A, Dasgupta K, Bernatsky S, Lacaille D, Avina-Zubieta A, Rahme E. Systematic review and meta-analysis of validation studies on a diabetes case definition from health administrative records. Barengo NC (ed). PLoS One 2013;8:e75256.
67. Pierron A, Revert M, Goueslard K, Vuagnat A, Cottenet J, Benzenine E, Fresson J. Evaluation of the metrological quality of the medico-administrative data for perinatal indicators: A pilot study in 3 university hospitals. Rev Epidemiol Sante Publique 2015;63:237–246.
68. Ameri H, Alizadeh S. Assessing the effects of infertility treatment drugs using clustering algorithms and data mining techniques. J Maz Univ Med Sci 2014;24:26–35.
69. Lidegaard O, Hammerum MS. The National Patient Registry as a tool for continuous production and quality control. Ugeskr Laeger 2002;164:4420–4423.
70. Rosenfeld Y, Strulov A. Improvement of accuracy of clinical reports--the case of IVF cycle rank. J Assist Reprod Genet 2009a;26:99–103.
71. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol 2015;7:449–490.

2.10 Figures and tables

List of tables

Table 1. Descriptive characteristics of included studies

Table 2. Summary of reported validity measures

Table 3. Reporting quality of methodology of included studies

Table 4. Reporting quality of the results of included studies

Table 5. Description of the pre- and post-test prevalence of measured estimates of validity in included studies

List of figures

Figure 1. PRISMA flow diagram

Table 1. Descriptive characteristics of included studies

Author	Year	Country	Data source being validated	Reference standard	Population	Sample size
Buck Louis (33)	2014	USA	Administrative database (Perinatal Data System)	Questionnaire	Mothers who had live births in Upstate New York between July 2008 and May 2010 in whom "Infertility treatment" was checked on birth certificate and multiple births matched to singleton infants whose treatment box was not checked	4989
Buck Louis (37)	2015	USA	Questionnaire (Upstate New York Infant Development Screening Program Study)	IVF registry (SART CORS)	Mothers who had live births in Upstate New York between July 2008 and May 2010 in whom "Infertility treatment" was checked on birth certificate and multiple births matched to singleton infants whose treatment box was not checked	5034
CDC (38)	2016	USA	Fertility database (SART)	Medical record	ART cycle data from 458 fertility clinics in the US during the 2014 cycle year. A random selection of 34 clinics were selected	1996
Cohen (34)	2014	USA	Administrative database (Birth certificates)	IVF registry (NASS)	Live births to Florida or Massachusetts resident mothers that occurred in state from March 2004 to December 2006	856165
Gissler (43)	2004	Finland	Administrative database (Medical Birth Record)	NA (Compared ad hoc IVF research and IVF statistics, no reference standard)	Newborns from fertility treatments from 1996 to 1998	176698
Hemminki (42)	2003	Finland	Administrative database (Drug Reimbursement Register)	Internal examination of data and linkage to Birth Register	Women exposed to ART between 1996 and 1998	24318

Hvidtjørn (44)	2009	Denmark	Administrative database	IVF Registry	Women who participated in the first Danish National Birth Cohort (study) interview with a pregnancy resulting in a live born child between October 2007 and June 2003	88151
Kotelchuck (35)	2014	USA	IVF registry (SART)	Administrative database (PELL)	Children born to Massachusetts resident women in MA hospitals from July 2004 to December 2008 conceived by ART	10138
Liberman (36)	2014	USA	Questionnaire (National Birth Defects Prevention Study)	IVF registry	Women who completed the National Birth Defects Prevention Study (NBDPS) with in-state deliveries between September 2004 and December 2008	77
Luke (39)	2016	USA	Administrative database (Birth certificates)	IVF registry	Live births in Florida, Massachusetts, New York, Pennsylvania, Texas, California, Ohio and Colorado between 2004 and 2009. IVF cycles from SART CORS were linked to birth certificates	716103
Molinaro (30)	2009	USA	IVF registry	Medical records	IVF patients enrolled for other studies at the University of Pennsylvania between December 2003 and June 2006.	590
Overbeek (45)	2013	Netherlands	Questionnaire (DCOG LATER-VEVO Study-nationwide cohort study)	Administrative database (Netherlands Perinatal Registry)	Childhood cancer survivors who achieved pregnancy and their sibling controls	524
Rosenfeld (46)	2009	Israel	IVF reporting system	Medical record	Women who receive fertility treatment in the District of Haifa and Western Galilee of the General Health Services	108
Stern (40)	2016	USA	IVF registry (SART)	Administrative database (Massachusetts BDMP Registry)	ART deliveries from July 1, 2004 to December 31, 2008 in Massachusetts	9092

Stern (41)	2016	USA	Questionnaire (Upstate New York Infant Development Screening Program Study)	SART database for current cycle; Questionnaire for prior treatment information	Mothers who participated in Upstate KIDS Study linked with SART CORS	617
Sunderam (29)	2006	USA	Administrative database	IVF registry	Infants born in 1997 and 1998 in MA, RI, NH, CT to MA-resident mothers who used ART clinics in MA or RI	2703
Williams (47)	2013	UK	Administrative database (National Registry of Childhood Tumours)	IVF registry (HFEA)	Children born between January 1, 1992 and December 31, 2008	106013
Zhang (32)	2012	USA	Administrative database	IVF registry (NASS)	Live births to MA-resident mothers that occurred in MA during 1997-2000	6139
Zhang (31)	2010	USA	Administrative database (Massachusetts Registry of Vital Records and Statistics-MBC)	IVF registry (NASS)	Live births to MA-resident mothers that occurred in MA during 1997-2000	5190

ART-Assisted reproductive technology; BDMP- Birth Defects Monitoring Program; CT-Connecticut; IVF-*in vitro* fertilization; MA-Massachusetts; NASS-National ART Surveillance System; NBDP-National Birth Defects; NH-New Hampshire; PELL-Pregnancy to Early Life Longitudinal data system; RI-Rhode Island; SART CORS-Society for Assisted Reproductive Technology Clinical Outcomes Reporting System; UK-United Kingdom; USA-United States of America

Table 2. Summary of reported validity measures

Study	Sensitivity	Specificity	PPV	NPV	Kappa	% agreement	ICC	AUC/ c-statistic	Likelihood ratios	4 or more measures of validity	Number of measures	95% CI
Buck Louis (33)	1/4	1/4	No	No	1/4	4/4	No	No	No	1/4	4	0/4
Buck Louis (37)	10/10	10/10	10/10	10/10	No	No	No	No	No	10/10	4	10/10
CDC (38)	No	No	No	No	No	No	No	No	No	No	1	18/18
Cohen (34)	2/2	2/2	2/2	No	No	No	No	No	No	No	3	2/2
Gissler (43)	No	No	No	No	1/2	2/2	No	No	No	No	2	1/2
Hemmink (42)	No	No	No	No	No	No	No	No	No	No	0	NA
Hvidtjørn (44)	3/3	3/3	3/3	No	No	3/3	No	No	No	3/3	4	0/3
Kotelchuk (35)	3/3	1/3	3/3	No	3/3	No	No	No	No	No	4	0/3
Liberman (36)	5/5	5/5	No	No	No	No	No	No	No	No	2	5/5
Luke (39)	1/1	1/1	No	No	No	No	No	No	No	No	2	No
Molinaro (30)	No	No	No	No	No	No	No	No	No	No	0	NA
Overbeek (45)	10/26	10/26	10/26	No	16/26	16/26	No	No	No	No	4	16/26
Rosenfeld (46)	No	No	No	No	No	No	No	No	No	No	2	No
Stern (40)	6/11	No	No	No	2/11	2/11	No	No	No	No	3	6/11
Stern (41)	13/13	No	No	No	No	3/13	No	No	No	No	5	No
Sunderam (29)	No	No	No	No	No	No	No	No	No	No	0	NA
Williams (47)	No	No	No	No	No	No	No	No	No	No	1	No
Zhang (32)	1/1	No	No	No	No	No	No	No	No	No	2	1/2
Zhang (31)	1/1	1/1	1/1	No	No	No	No	No	No	No	3	3/3

AUC-Area under the curve; CI-Confidence intervals; ICC-Intraclass correlation coefficient; NPV-Negative predictive value; PPV-positive predictive value

Table 3. Reporting quality of methodology of included studies

Methods	**Frequency**	**%**
Describes the data source		
Yes	16/19	84.2%
Incomplete	2/19	10.5%
Unclear	1/19	5.3%
Describes type of records (inpatient, outpatient, linked records)		
Yes	18/19	94.7%
Unclear	1/19	5.3%
Describes setting and locations where data were collected		
Yes	18/19	94.7%
Incomplete	1/19	5.3%
Reports a priori sample size		
Yes	1/19	5.3%
Provides statistical justification for the sample size		
Yes	0/19	0.0%
Describe recruitment procedure of validation cohort (from a database, based on diagnostic codes)		
Yes	17/19	89.5%
Unclear	2/19	10.5%
Describe patient sampling (Random, consecutive, all)		
Random sampling	1/19	5.3%
All	14/19	73.7%
Unclear	2/19	10.5%
Incomplete	2/19	10.5%
Describe how participants were chosen for data collection and analysis		
Yes	15/19	78.9%
Unclear	2/19	10.5%
Describes inclusion/exclusion criteria		
Yes	14/19	73.7%
Incomplete	1/19	5.3%
Describes who identified patients (for patients identified from medical records)		
Yes	1/19	5.3%
Incomplete	1/19	5.3%
Describes who collected data		
Yes	3/19	15.8%
Describes use of a priori data collection form		
Yes	13/19	68.4%
Unclear	1/19	5.3%
Use of a split sample or an independent sample (revalidation using a separate cohort)		
Yes	1/19	5.3%
Describes the reference standard		
Yes	13/17	76.5%
Reports the number of persons reading the reference standard		
Yes	2/17	11.8%

Describes the training or expertise of persons reading reference standard		
Yes	1/17	5.9%
Readers of the reference standard were blinded to the results of the classification by routinely collected data for that patient (reference standard: medical records)		
Yes	1/17	5.9%
Reports a measure of concordance if >1 persons reading the reference standard		
Yes	0/17	0.0%
Describes the linkage procedure, if done (probabilistic/deterministic)		
Yes	8/15	50.0%
Incomplete	6/15	37.5%
Describes the methods of linkage quality evaluation		
Yes	7/15	46.7%
Incomplete	2/15	13.3%
Describes explicit methods for calculating or comparing measures of accuracy and statistical methods used to quantify uncertainty		
Yes	13/19	68.4%

Table 4. Reporting quality of the results of included studies

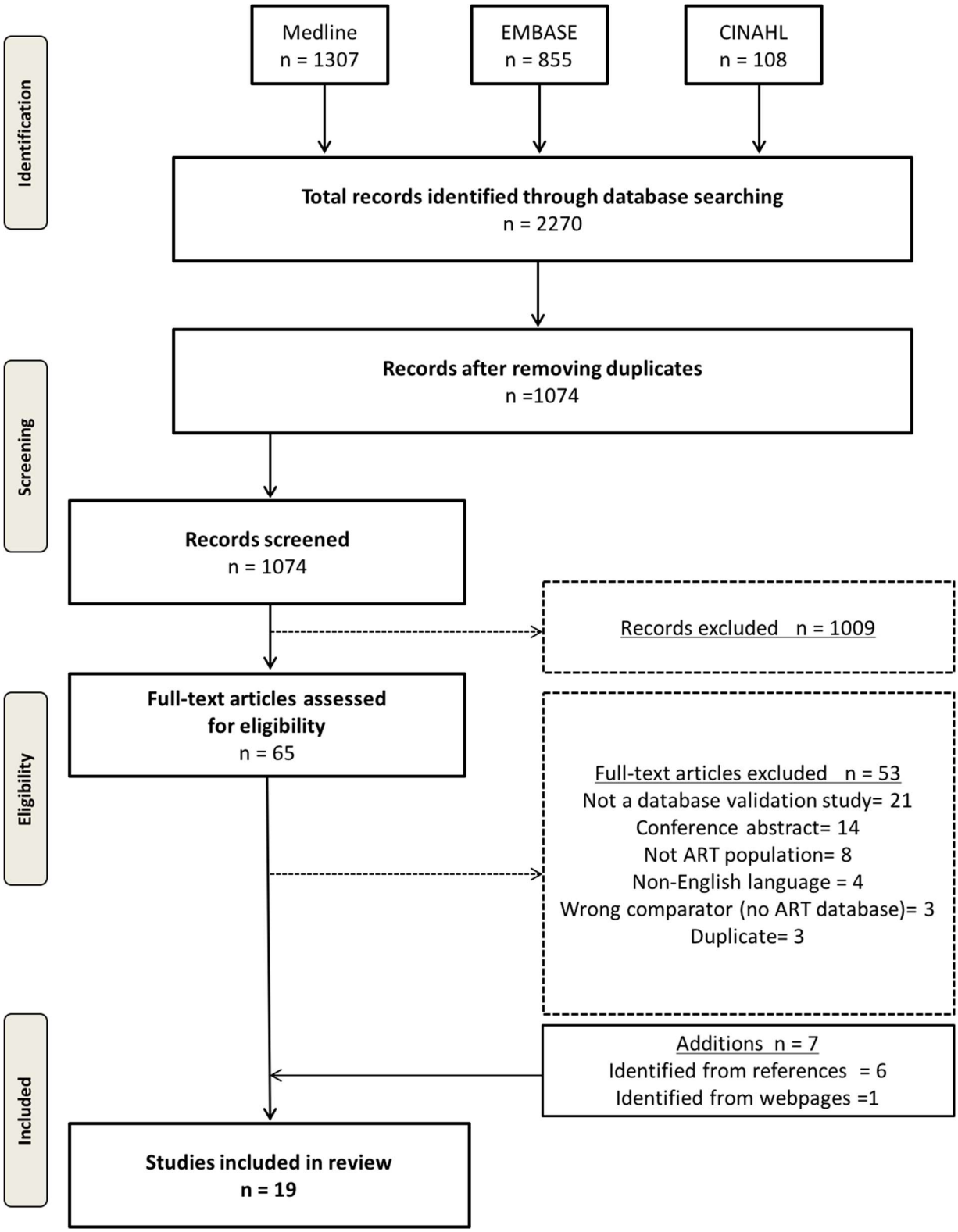
Reports the number of participants satisfying the inclusion/ exclusion criteria		
Yes	13/18	68.4%
Incomplete	1/18	5.6%
Describes the characteristics of misclassified patients (false-positives and/or false negatives)		
Yes	13/18	68.4%
Unclear	2/18	11.1%
Provides a study flow diagram		
Yes	4/19	21.1%
Reports the number of records unable to link		
Yes	11/12	91.7%
Incomplete	1/12	8.3%
Reports missing medical records or reports the number of patients unwilling to participate		
Yes	10/19	52.6%
Reports incomplete records		
Yes	13/19	68.4%
Presents a cross tabulation of results of the validated source to the reference standard		
Yes	11/19	57.9%
Incomplete	1/19	5.3%
Reports the pretest prevalence in the study sample		
Yes	5/19	26.3%
Incomplete	2/19	10.5%
Tests and reports results of multiple algorithms		
Yes	6/15	40.0%
Reports estimates of test reproducibility of the split or independent sample if done		
Yes	0/19	0.0%

Table V. Description of the pre- and post-test prevalence of measured estimates of validity in included studies

Study	Prevalence estimate reported	Pre-test prevalence	Post-test prevalence*
Buck Louis (33)	ART conceived infant	1.4%	14%
Buck Louis (37)	No	-	-
CDC (38)	No	-	-
Cohen (34)	ART conceived infant	1.40%	0.45%
Gissler (43)	No	-	-
Hemminki (42)	No	-	-
Hvidtjørn (44)	No	-	-
Kotelchuk (35)	ART conceived infant	1.60%	2.72%
Liberman (36)	ART conceived infant in MA	4.30%	5.30%
Luke (39)	ART conceived infant	1.70%	9.8%
Molinaro (30)	No	-	-
Overbeek (45)	No	-	-
Rosenfeld (46)	No	-	-
Stern (40)	Incomplete	-	-
Stern (41)	No	-	-
Sunderam (29)	Yes	3%	-
Williams (47)	No	-	-
Zhang (32)	No	-	-
Zhang (31)	ART Live birth deliveries	3%	1.7%

*Based on reference standard
ART-Assisted reproductive technology; MA-Massachusetts

Figure 1. PRISMA flow diagram



Appendix A: Supplemental information for Manuscript 1

Appendix 1. Medline search strategy

1. exp infertility/
2. infertil*.tw.
3. exp infertility therapy/
4. (assist* adj2 reproduc*).tw.
5. (reproduc* adj2 technolog*).tw.
6. ivf.tw.
7. in vitro fertil*.tw.
8. icsi.tw.
9. intracytoplasmic sperm injection*.tw.
10. ovulation induc*.tw.
11. (ovari* adj2 stimulat*).tw.
12. reproductive procedure/ or fertility preservation/ or follicular aspiration/ or oocyteretrieval/
13. maternal age/
14. advanc* maternal age.tw.
15. advanc* reproduc* age.tw.
16. premature ovarian failure/
17. (premature ovar* adj (insufficien* or failure)).tw.
18. diminish* ovar* reserve.tw.
19. *twin pregnancy/
20. chorionicity.tw.
21. (dichorionic or monochorionic).tw.
22. or/1-21
23. data base/ or factual database/
24. database*.tw.
25. (admin* adj2 data*).tw.
26. (data adj2 warehouse*).tw.
27. data base*.tw.
28. exp medical record/
29. (medic* adj2 record*).tw.
30. (utili?ation data* or claims data* or managed care data* or physician billing data*or hospitali?ation data* or linked data*).tw.
31. health record*.tw.
32. (patient* adj2 record*).tw.
33. (emr or epr or ehr).tw.
34. (physician* adj2 claim*).tw.

35. coding/ or patient coding/
36. ((clinic* or patient) adj2 (code or coding)).tw.
37. (admin* adj2 bill*).tw.
38. exp "international classification of diseases"/
39. (icd9* or icd10* or icd).tw.
40. register/ or disease registry/
41. (registry or registries).tw.
42. or/23-41
43. 22 and 42
44. validation study/
45. (validit* or validation).tw.
46. "sensitivity and specificity"/
47. (sensitivit* or specificit*).tw.
48. predictive value/
49. (predictiv* adj2 value*).tw.
50. accuracy/
51. (accurate or accuracy).tw.
52. or/44-51
53. 43 and 52

Appendix 2. Excluded references

Ameri H, Alizadeh S. Assessing the effects of infertility treatment drugs using clustering algorithms and data mining techniques. J Maz Univ Med Sci 2014;24:26–35.
Reason for exclusion: Non-English

Anazodo AC, Stern CJ, McLachlan RI, Gerstl B, Agresta F, Cohn RJ, Jayasinghe Y, Wakefield CE, Daly G, Chan D, et al. A Study Protocol for the Australasian Oncofertility Registry: Monitoring Referral Patterns and the Uptake, Quality, and Complications of Fertility Preservation Strategies in Australia and New Zealand. J Adolesc Young Adult Oncol 2016;5:215–225.
Reason for exclusion: Not a database validation study

Baldwin E, Johnson K, Berthoud H. Linking mothers and infants within electronic health records: A comparison of deterministic and probabilistic algorithms. Pharmacoepidemiol Drug Saf 2015;24:45–51.
Reason for exclusion: Not ART population

Blenstrup LT, Knudsen LB. Danish registers on aspects of reproduction. Scand J Public Health 2011;39:79–82.
Reason for exclusion: Not a database validation study

de Boer EJ, den Tonkelaar I, Burger CW, van Leeuwen FE, Group OP. Validity of self-reported causes of subfertility. Am J Epidemiol 2005;161:978–986.
Reason for exclusion: Wrong comparator (no ART database)

Boyer P, Gervoise-Boyer M, Meddeb L, Rossin B. Modelization of growth between birth and 6 years of age in children born after ART in a French monocentric cohort compared to references growth curves. Hum Reprod 2010;25:i254.
Reason for exclusion: Conference abstract

Castilla JA, Luceno F, Gomez-Palomares JL, Marqueta J, Hernandez J, Cabello Y, Herrero J, Vidal E, Fernandez-Shaw S. Differences between a voluntary ART register and an official one in Spain. J fur Reproduktionsmedizin und Endokrinol 2010;7:329.
Reason for exclusion: Conference abstract

Dejoy S, Pekow P, Bertone-Johnson E. Validation of a Certified Nurse-Midwifery Database for Use in Quality Monitoring and Outcomes Research. J Midwifery Women's Heal 2014;59:438–446.
Reason for exclusion: Not ART population

Dick M-LB, Bain CJ, Purdie DM, Siskind V, Molloy D, Green AC. Self-reported difficulty in conceiving as a measure of infertility. Hum Reprod 2003;18:2711–2717.
Reason for exclusion: Wrong comparator (no ART database)

Eisenberg M, Pastuszak AW, Langlois P, Moffitt K, Lamb DJ. The risk of congenital birth

defects is not associated with semen parameters. J Urol 2014;191:e798-e799.
Reason for exclusion: Conference abstract

Eisenberg M, Pastuszak AW, Langlois P, Moffitt K, Lamb DJ. The risk of congenital birth defects is not associated with semen parameters or mode of conception in offspring of men visiting a reproductive clinic. Andrology 2015;3:48.
Reason for exclusion: Conference abstract

Ellison GT, de Wet T, Matshidze KP, Cooper P. The reliability and validity of self-reported reproductive history and obstetric morbidity amongst Birth to Ten mothers in Soweto. Curationis 2000;23:76–80.
Reason for exclusion: Wrong comparator (no ART database)

ESHRE, European IVF-Monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology, Calhaz-Jorge C, de Geyter C, Kupka MS, de Mouzon J, Erb K, Mocanu E, Motrenko T, Scaravelli G, et al. Assisted reproductive technology in Europe, 2012: results generated from European registers by ESHRE. Hum Reprod 2016;31:1638–1652.
Reason for exclusion: Not a validation study

Guzick DS, Boles J, Schadle R. Database management system for assisted reproduction. J In Vitro Fert Embryo Transf 1990;7:236–240.
Reason for exclusion: Not a validation study

Hilder L, Moser K, Dattani N, Macfarlane A. Pilot linkage of NHS Numbers for Babies data with birth registrations. Health Stat Q 2007:25–33.
Reason for exclusion: Not ART population

Joffe M, Villard L, Li Z, Plowman R, Vessey M. A time to pregnancy questionnaire designed for long term recall: validity in Oxford, England. J Epidemiol Community Heal 1995;49:314–319.
Reason for exclusion: Not ART population

Kelley-Quon LI, Tseng C-H, Janzen C, Shew SB. Congenital malformations associated with assisted reproductive technology: a California statewide analysis. J Pediatr Surg 2013;48:1218–1224.
Reason for exclusion: Not a validation study

Kesmodel US, Ingerslev HJ, Lemmen JG, Rasmussen, I.A. Adverse effects in pregnancy after treatment with preimplantation genetic diagnosis-a Danish national multicenter followup study. Hum Reprod 2015;30:i359-i360.
Reason for exclusion: Conference abstract

Kohli KL, Al Omaim M. Some indirect estimates of fertility from 1980 census data in Kuwait. Popul Bull U N Econ Comm West Asia 1985:39–62.
Reason for exclusion: Not a validation study

Kupka MS, Dorn C. Development of electronic data collection as a tool for quality assessment in reproductive techniques in Germany. Reprod Technol 2001;10:332–334.
Reason for exclusion: Not a validation study

Lidegaard O, Hammerum MS. The National Patient Registry as a tool for continuous production and quality control. Ugeskr Laeger 2002;164:4420–4423.
Reason for exclusion: Non-English

Liu J, Tuvblad C, Li L, Raine A, Baker LA. Medical record validation of maternal recall of pregnancy and birth events from a twin cohort. Twin Res Hum Genet 2013;16:845–860.
Reason for exclusion: Not ART population

Luceno F, Castilla JA, Gomez-Palomares JL, Cabello Y, Hernandez J, Marqueta J, Herrero J, Vidal E, Fernandez-Shaw S, Coroleu B. Comparison of IVF cycles reported in a voluntary ART registry with a mandatory registry in Spain. Hum Reprod 2010;25:3066–3071.
Reason for exclusion: Not a validation study

Luceno Maestre F, Castilla Alcala JA, Gomez-Palomares JL, Cabello Y, Hernandez J, Marqueta J, Herrero J, Vidal E, Fernandez-Shaw S. Validation of a voluntary assisted reproductive technology register by comparison with an official one. Hum Reprod 2010;25:i179.
Reason for exclusion: Conference abstract

Luke B, Brown MB, Wantman E, Lederman A, Gibbons W, Schattman GL, Lobo RA, Leach RE, Stern JE. Cumulative Birth Rates with Linked Assisted Reproductive Technology Cycles. N Engl J Med 2012;366:2483–2491.
Reason for exclusion: Not a validation study

Luke B, Cabral H, Cohen BB, Hoang L, Plummer K. Comparison of measures in sart database and massachusetts vital statistics. Fertil Steril 2012;98:S260.
Reason for exclusion: Conference abstract

Margulis AV, Setoguchi S, Mittleman MA, Glynn RJ, Dormuth CR. Algorithms to estimate the beginning of pregnancy in administrative databases. Pharmacoepidemiol Drug Saf 2013;22:16–24.
Reason for exclusion: Not ART population

Mneimneh AS, Boulet SL, Sunderam S, Zhang Y, Jamieson DJ, Crawford S, McKane P, Copeland G, Mersol-Barg M, Grigorescu V, et al. States Monitoring Assisted Reproductive Technology (SMART) Collaborative: Data Collection, Linkage, Dissemination, and Use. J Women's Heal 2013;22:571–577.
Reason for exclusion: Not a validation study

Morse CB, Sammel MD, Dokras A, Coutifaris C, Barnhart K. Weighing the evidence: Accuracy of perinatal outcomes reported to the society for assisted reproductive technologies (SART) database. Fertil Steril 2012;98:S100-S101.

Reason for exclusion: Conference abstract

De Neubourg D, Bogaerts K, Blockeel C, Coetsier T, Delvigne A, Devreker F, Dubois M, Gillain N, Gordts S, Wyns C. How do cumulative live birth rates and cumulative multiple live birth rates over complete courses of assisted reproductive technology treatment per woman compare among registries?. Hum Reprod 2016;31:93–99.
Reason for exclusion: Not a validation study

Ohm Kyvik K, Derom C. Data collection on multiple births -- establishing twin registers and determining zygosity. Early Hum Dev 2006;82:357–363.
Reason for exclusion: Not a validation study

Pierron A, Revert M, Goueslard K, Vuagnat A, Cottenet J, Benzenine E, Fresson J. Evaluation of the metrological quality of the medico-administrative data for perinatal indicators: A pilot study in 3 university hospitals. Rev Epidemiol Sante Publique 2015;63:237–246.
Reason for exclusion: Non-English

Priskorn L, Jensen TK, Lindahl-Jacobsen R, Skakkebaek NE, Bostofte E, Eisenberg ML. Parental age at delivery and a man's semen quality. Hum Reprod 2014;29:1097–1102.
Reason for exclusion: Not a validation study

Purohit D, Thompson S, Choi J, Harrison L, Jones A. Evaluation of an ICD-9 search algorithm for identifying women with polycystic ovary syndrome. Endocr Rev 2011;32:no pagination.
Reason for exclusion: Conference abstract

Rosenfeld Y, Strulov A. Clinical reports on IVF cycle rank--reliability and validity. Harefuah 2009;148:22–24.
Reason for exclusion: Non-English

Rubenstein J. How Will the Transition to ICD-10 Affect Urology Coding? An Analysis of ICD-9 Code Use from a Large Group Practice. Urol Pract 2015;2:312–315.
Reason for exclusion: Not a validation study

Schieve L, Cohen B, Nannini A, Ferre C, Reynolds L, Zhang Z, Jeng G, Macaluso M, Wright V. A population-based study of maternal and perinatal outcomes associated with assisted reproductive technology in Massachusetts. Matern Child Health J 2007;11:517–525.
Reason for exclusion: Not a validation study

Simpson JL. Registration of congenital anomalies in ART populations: pitfalls. Hum Reprod 1996;11:81–88.
Reason for exclusion: Not a validation study

Stern JE, Gopal D, Kotelchuck M. Validation of birth defects data in the SART CORS to birth defects registry data in Massachusetts. Fertil Steril 2015;104:e205-e206.
Reason for exclusion: Conference abstract

Stern JE, Gopal D, Liberman RF, Anderka M, Kotelchuck M, Luke B. Validation of birth outcomes from the Society for Assisted Reproductive Technology Clinic Outcome Reporting System (SART CORS): population-based analysis from the Massachusetts Outcome Study of Assisted Reproductive Technology (MOSART). Fertil Steril 2016;106:717–722.e2.
Reason for exclusion: Duplicate

Stern JE, Hickman TN, Kinzer D, Penzias AS, Ball GD. Can the Society for Assisted Reproductive Technology Clinic Outcome Reporting System (SART CORS) be used to accurately report clinic total reproductive potential (TRP)? Fertil Steril 2012;97:886–889.
Reason for exclusion: Not a validation study

Stern JE, Hickman TN, Kinzer D, Penzias AS, Ball GD, Gibbons WE. Can the Society for Assisted Reproductive Technology Clinic Outcome Reporting System (SART CORS) be used to accurately report clinic total reproductive potential (TRP)?. Fertil Steril 2012;97:886–889.
Reason for exclusion: Duplicate

Stern JE, Kotelchuck M, Luke B, Declercq E, Belanoff C. Alternate methods of estimating gestational age from national databases may overestimate rates of preterm birth following art. Fertil Steril 2012;98:S50.
Reason for exclusion: Conference abstract

Stuebe A, Dorman K, Wapner R, DiVito M, Tita A, Biggio J, R. L, Chapman V, Saade G, Salazar A, et al. Creation of a multiinstitutional CTSA-sponsored obstetrics registry for adverse rare events (RARE). Clin Transl Sci 2014;7:270.
Reason for exclusion: Conference abstract

Sullivan EA, Wang YA, Abeywardana S. Congenital anomalies following assisted reproductive technology in Australia. Hum Reprod 2011;26:i119-i120.
Reason for exclusion: Conference abstract

Sullivan EA, Zegers-Hochschild F, Mansour R, Ishihara O, De Mouzon J, Nygren KG, Adamson GD. International Committee for Monitoring Assisted Reproductive Technologies (ICMART) world report: assisted reproductive technology 2004. Hum Reprod 2013;28:1375–1390.
Reason for exclusion: Not a validation study

Thomas FS, Stanford JB, Sanders JN, Gurtcheff SE, Gibson M, Porucznik CA. Development and initial validation of a fertility experiences questionnaire. Reprod Health 2015;12:62.
Reason for exclusion: Not ART population

Thomas FS, Stanford JB, Sanders JN, Gurtcheff SE, Gibson M, Porucznik CA, Simonsen SE. Development and initial validation of a fertility experiences questionnaire. Reprod Health 2015;12:62.

Reason for exclusion: Duplicate

Toner JP, Coddington CC, Doody K, Van Voorhis B, Seifer DB, Ball GD, Luke B, Wantman E. Society for Assisted Reproductive Technology and assisted reproductive technology in the United States: a 2016 update. Fertil Steril 2016;106:541–546.
Reason for exclusion: Not a validation study

Vega M, Breborowicz A, Morris S, Sirota I, Gonzales E. Peak estradiol (E2) at the time of HCG trigger as a predictor of small for gestational age (SGA). Fertil Steril 2013;100:S326.
Reason for exclusion: Conference abstract

Wang ET, Ozimek JA, Greene N, Ramos L, Vyas N, Kilpatrick SJ, Pisarska MD. Impact of fertility treatment on severe maternal morbidity. Fertil Steril 2016;106:423–426.
Reason for exclusion: Not a validation study

Wang YA, Nikravan R, Smith HC, Sullivan EA. Higher prevalence of gestational diabetes mellitus following assisted reproduction technology treatment. Hum Reprod 2013;28:2554–2561.
Reason for exclusion: Not ART population

Yeung EH, Sundaram R, Bell EM, Druschel C, Kus C, Ghassabian A, Bello S, Xie Y, Buck Louis GM. Examining Infertility Treatment and Early Childhood Development in the Upstate KIDS Study. JAMA Pediatr 2016;170:251–258.
Reason for exclusion: Not a validation study

Appendix 3. Estimates of the measures of validity from included studies

Author	Validation	Variable	Sensitivity	Specificity	Kappa	NPV	PPV	% Agreement	ICC	AUC/c-statistic	LR
Buck Louis (33)	Diagnosis/ Variable		0.55	0.99	0.807	No	No		No	No	No
		No infertility treatment						97.0%			
		Yes, any infertility treatment						81.0%			
		Yes, infertility treatment (specifically drugs or ART)						73.0%			
Buck Louis (37)	Diagnosis/ Variable	Overall	0.93 (0.87-0.97)	0.99 (0.99-1.00)	No	1.00 (1.00-1.00)	0.80 (0.73-0.86)	No	No	No	No
		Private insurance-No	0.98 (0.81-1.00)	0.99 (0.99-1.00)	No	1.00 (1.00-1.00)	0.34 (0.10-0.65)	No	No	No	No
		Private insurance-Yes	0.93 (0.87-0.97)	0.99 (0.99-1.00)	No	1.00 (0.99-1.00)	0.84 (0.77-0.89)	No	No	No	No
		Race-Other	1.00 (1.00-1.00)	0.99 (0.97-1.00)	No	1.00 (1.00-1.00)	0.74 (0.39-0.95)	No	No	No	No
		Race-White	0.93 (0.87-0.97)	0.99 (0.99-1.00)	No	1.00 (1.00-1.00)	0.83 (0.75-0.89)	No	No	No	No
		Race-Black	0.81 (0.35-0.99)	0.98 (0.96-0.99)	No	1.00 (0.98-1.00)	0.53 (0.20-0.84)	No	No	No	No
		Education (<HS)	0.73 (0.36-0.95)	0.99 (0.98-1.00)	No	1.00 (0.99-1.00)	0.44 (0.19-0.71)	No	No	No	No
		Education (>College)	0.95 (0.89-0.98)	0.99 (0.99-1.00)	No	1.00 (1.00-1.00)	0.84 (0.77-0.90)	No	No	No	No
		Maternal age (< 30)	0.82 (0.51-0.97)	0.99 (0.99-1.00)	No	1.00 (1.00-1.00)	0.56 (0.33-0.77)	No	No	No	No
		Maternal age (> 30)	0.95 (0.89-0.98)	0.99 (0.98-0.99)	No	1.00 (0.99-1.00)	0.84 (0.76-0.90)	No	No	No	No
		Paternal Age (<30)	0.79 (0.31-0.99)	0.99 (0.99-1.00)	No	1.00 (1.00-1.00)	0.36 (0.14-0.62)	No	No	No	No
		Paternal Age (>30)	0.94 (0.88-0.97)	0.99 (0.99-1.00)	No	1.00 (0.99-1.00)	0.86 (0.79-0.92)	No	No	No	No
		Parental smoking-No	0.93 (0.87-0.97)	0.99 (0.99-1.00)	No	1.00 (1.00-1.00)	0.83 (0.75-0.88)	No	No	No	No
		Parental smoking-Yes	1.00 (1.00-1.00)	0.99 (0.98-1.00)	No	1.00 (1.00-1.00)	0.42 (0.12-0.76)	No	No	No	No
		Risk factors during pregnancy-No	0.96 (0.88-0.99)	0.99 (0.99-1.00)	No	1.00 (1.00-1.00)	0.79 (0.68-0.87)	No	No	No	No
		Risk factors during pregnancy-Yes	0.90 (0.79-0.97)	0.99 (0.99-1.00)	No	1.00 (0.99-1.00)	0.82 (0.71-0.90)	No	No	No	No
		WIC Participation-No	0.93 (0.86-0.97)	0.99 (0.99-1.00)	No	1.00 (0.99-1.00)	0.85 (0.78-0.90)	No	No	No	No
		WIC Participation-Yes	0.98 (0.84-1.00)	0.99 (0.99-1.00)	No	1.00 (1.00-1.00)	0.40 (0.17-0.67)	No	No	No	No

		Abnormal fetal conditions-No	0.95 (0.88-0.98)	0.99 (0.99-1.00)	No	1.00 (1.00-1.00)	0.78 (0.69-0.86)	No	No	No	No
		Abnormal fetal conditions-Yes	0.91 (0.78-0.98)	0.99 (0.98-0.99)	No	0.99 (0.98-1.00)	0.85 (0.71-0.94)	No	No	No	No
		Fetal presentation-cephalic	0.92 (0.84-0.96)	0.99 (0.99-1.00)	No	1.00 (1.00-1.00)	0.81 (0.73-0.88)	No	No	No	No
		Fetal presentation-Breech	0.99 (0.95-1.00)	0.99 (0.97-1.00)	No	1.00 (0.98-1.00)	0.87 (0.69-0.96)	No	No	No	No
		Fetal presentation-Other	1.00 (1.00-1.00)	0.96 (0.88-0.99)	No	1.00 (1.00-1.00)	0.12 (0.01-0.43)	No	No	No	No
CDC (38)	Diagnosis/ Variable										
		Patient date of birth	No	No	No	No	No	No	No	No	No
		Cycle intention	No	No	No	No	No	No	No	No	No
		Cycle cancellation	No	No	No	No	No	No	No	No	No
		Number of eggs/embryos transferred	No	No	No	No	No	No	No	No	No
		Outcome of ART treatment	No	No	No	No	No	No	No	No	No
		Number of fetal hearts on ultrasound	No	No	No	No	No	No	No	No	No
		Pregnancy outcome	No	No	No	No	No	No	No	No	No
		Date of pregnancy outcome	No	No	No	No	No	No	No	No	No
		Number of infants born	No	No	No	No	No	No	No	No	No
		Cycle count	No	No	No	No	No	No	No	No	No
		Tubal factor	No	No	No	No	No	No	No	No	No
		Ovulatory dysfunction	No	No	No	No	No	No	No	No	No
		Diminished ovarian reserve	No	No	No	No	No	No	No	No	No
		Endometriosis	No	No	No	No	No	No	No	No	No
		Uterine factor	No	No	No	No	No	No	No	No	No
		Male factor	No	No	No	No	No	No	No	No	No
		Other factor	No	No	No	No	No	No	No	No	No

		Unknown factor	No	No	No	No	No	No	No	No	No
Cohen (34)	Diagnosis/ Variable										
		ART (yes/no)-MA	41.4 (31.7-51.8)	99.0 (98.8-99.2)	No	No	54.6 (51.9-57.2)	No	No	No	No
		ART (yes/no)-FL	28.9 (23.2-35.3)	99.8 (99.7-99.8)	No	No	45.5 (41.7-49.4)	No	No	No	No
Gissler (43)	Diagnosis/ Variable										
		IVF status (yes/no)	No	No	0.75 (0.74-0.76)	No	No	98.7%	No	No	No
		IVF (in IVF children)			No			60.2%			No
Hemminki (42)	Algorithm		No	No	No	No	No	No	No	No	No
Hvidtjørn (44)	Diagnosis/ Variable										
		OI/Insemination	74	99	No	No	79	No	No	No	No
		IVF/ICSI	84	99	No	No	76	No	No	No	No
		Assisted conception	83	99	No	No	88	No	No	No	No
Kotelchuck (35)	Linkage										
		ART Status	36.9	No	0.453	No	0.626	No	No	No	No
		Plurality at delivery	98.9	No	0.973	No	0.995	No	No	No	No
		Fetal loss/live birth associated delivery	47.4	No	0.457	No	0.692	No	No	No	No
Liberman (36)	Diagnosis/ Variable										
		IVF	91 (82-96)	No	No	No	No	No	No	No	No
		ICSI	71 (48-89)	87 (66-97)	No	No	No	No	No	No	No
		Donor egg	67 (22-96)	100 (95-100)	No	No	No	No	No	No	No
		Frozen cycle	100 (40-100)	99 (92-100)	No	No	No	No	No	No	No
		Multiple fetuses	100 (91-100)	94 (81-99)	No	No	No	No	No	No	No
Luke (39)	Diagnosis/ Variable										
		ART, IVF or GIFT	28.2	99.7	No	92.8	89	No	No	No	No
Molinaro (30)	Diagnosis/ Variable										
		Endometriosis	No	No	No	No	No	No	No	No	No
		Tubal factor	No	No	No	No	No	No	No	No	No

		PCOS	No	No	No	No	No	No	No	No	No
		Diminished ovarian reserve	No	No	No	No	No	No	No	No	No
		Male factor	No	No	No	No	No	No	No	No	No
		Unexplained	No	No	No	No	No	No	No	No	No
		Uterine	No	No	No	No	No	No	No	No	No
Overbeek (45)	Diagnosis/ Variable										
		Natural conception-CCS	No	No	0.82 (0.67-0.97)	No	No	No	No	No	No
		Hormonal conception-CCS	No	No	0.56 (0.12-1.01)	No	No	No	No	No	No
		IUI-CCS	No	No	0.85 (0.57-1.14)	No	No	No	No	No	No
		IVF/ICSI-CCS	No	No	0.81 (0.61-1.02)	No	No	No	No	No	No
		Natural conception-Control	No	No	0.71 (0.39-1.02)	No	No	No	No	No	No
		Hormonal conception-Control	No	No	0.62 (0.04-1.28)	No	No	No	No	No	No
		IUI-Control	No	No	N/A	No	No	No	No	No	No
		IVF/ICSI-Control	No	No	0.85 (0.56-1.14)	No	No	No	No	No	No
Rosenfeld (46)	Diagnosis/ Variable										
		IVF cycle rank	No	No	No	No	No	No	No	No	No
Stern (40)	Diagnosis/ Variable										
		Maternal race/ethnicity	No	No	0.86 (0.83-0.88)	No	No	No	No	No	No
		Delivery outcome	No	No	NR	No	No	No	No	No	No
		Plurality at birth	No	No	0.99 (0.98-0.99)	No	No	No	No	No	No
		Outcome/delivery date difference	No	No	No	No	No	92.6%	No	No	No
		Birthweight of singletons difference	No	No	No	No	No	86.7%	No	No	No
		Infants with a birth defect	38.6 (30.3-47.5)	No	No	No	No	No	No	No	No
		Cleft palate	40.0 (5.3-85.3)	No	No	No	No	No	No	No	No
		Genetic	43.5 (23.2-65.5)	No	No	No	No	No	No	No	No
		Cardiac	27.5 (14.6-43.9)	No	No	No	No	No	No	No	No
		Limb	50.0 (1.3-98.7)	No	No	No	No	No	No	No	No

		Other	18.4 (11.8-26.8)	No	No	No	No	No	No	No	No
Stern (41)	Diagnosis/ Variable										
		Vaginal ultrasound	80	No	No	No	No	No	No	No	No
		Vaginal ultrasound (removing FET)	80	No	No	No	No	No	No	No	No
		Medication any administration	92	No	No	No	No	No	No	No	No
		IVF no ICSI	81	No	No	No	No	No	No	No	No
		IVF with ICSI	78	No	No	No	No	No	No	No	No
		Donor egg	82	No	No	No	No	No	No	No	No
		Donor sperm	82	No	No	No	No	No	No	No	No
		Donor embryos	0	No	No	No	No	No	No	No	No
		Assisted hatching	38	No	No	No	No	No	No	No	No
		FET	93	No	No	No	No	No	No	No	No
		Medication any administration prior to index birth with gonadotropins	76	No	No	No	No	No	No	No	No
		Medication any administration prior to index birth with gonadotropins fresh or frozen	75	No	No	No	No	No	No	No	No
		IVF any/GIFT/ZIFT	79	No	No	No	No	No	No	No	No
Sunderam (29)	Linkage		No	No	No	No	No	No	No	No	No
Williams (47)	Linkage		No	No	No	No	No	No	No	No	No
Zhang (32)	Linkage		96.4 (95.7-96.9)	No	No	No	95.6	No	No	No	No
Zhang (31)	Diagnosis/ Variable		27.1 (25.9-28.4)	99.7 (99.7-99.7)	No	No	59.3 (57.3-61.2)	No	No	No	No

ART-Assisted reproductive technology; AUC-Area under the curve; CCS-Childhood cancer survivors; FET-Frozen embryo transfer; GIFT-Gamete intrafallopian tube transfer; FL-Florida; HS-high school; ICC-Intraclass correlation; ICSI-Intracytoplasmic sperm injection; IVF-in vitro fertilization; LR-Likelihood ratio; MA-Massachusetts; NPV-Negative predictive value; OI-Ovulation induction; PCOS-Polycystic ovarian syndrome; PPV-Positive predictive value; WIC-Women, Infants and Children; ZIFT-Zygote intrafallopian tube transfer

Appendix 4. PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	8
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	9
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	11-12
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	12-13
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	12
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	12-13
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	13-14
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	14
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	14-15
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	14
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	15
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	NA

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	15
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	NA
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	NA
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	15, Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	15 -16 Table 1 Appendix 3
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Table 3 Table 4
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Table 2
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	NA
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	NA
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	NA
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	17
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	21
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	22
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	23

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

Chapter 3. [Manuscript 2] The Canadian Assisted Reproductive Technologies Register (CARTR) Plus database: A validation study

3.1 Preface to Manuscript 2

We performed a validation study of the CARTR Plus database to answer the following questions: Are data in CARTR Plus accurate and reliable?

This manuscript has not yet been submitted for publication.

3.2 Title page

Title: The Canadian Assisted Reproductive Technologies Register (CARTR) Plus database: A validation study

Running title: CARTR Plus validation study

Authors: V. Bacal[1,2,3], D.B. Fell[2,4], H. Shapiro[5,6], A. Lanes[3,7], A. Sprague[4,7], M. Johnson[7], M. Walker[1,2,3,7], L.M. Gaudet[1,2,3].

[1]University of Ottawa, Obstetrics and Gynecology, Ottawa, Canada.

[2]University of Ottawa, School of Epidemiology and Public Health SEPH/ École d'épidémiologie et de santé publique EESP, Ottawa, Canada.

[3]The Ottawa Hospital, Ottawa Hospital Research Institute, Ottawa, Canada.

[4]Children's Hospital of Eastern Ontario, Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada.

[5]University of Toronto, Department of Obstetrics and Gynecology, Toronto, Canada.

[6]Mount Sinai Hospital, Mount Sinai Fertility, Toronto, Canada.

[7]Children's Hospital of Eastern Ontario, Better Outcomes Registry & Network Ontario, Ottawa, Canada.

Corresponding author: Vanessa Bacal

Financial support: This project was partially funded in the amount of $5000 by the Canadian Fertility and Andrology Society Research Seed Grant

Capsule:

This validation study of the CARTR Plus database demonstrated that the majority of data elements verified are of very high quality

3.3 Abstract

Study question:

Are data in the Canadian Assisted Reproductive Technologies Register (CARTR) Plus database valid and reliable?

Summary answer:

Markers of validity were strong for the majority of variables evaluated. Those with moderate agreement were FSH levels, oocyte origin and elective single embryo transfer.

What is known already:

Health databases and registries are excellent sources of data. However, as these databases are typically not established for the primary purpose of performing research, they should be validated prior to utilization for research both to inform the study design and to determine the extent to which key study variables are accurately documented in the database. CARTR Plus is Canada's national register for collecting extensive information on in vitro fertilization (IVF) and corresponding pregnancy outcomes, and it has yet to be validated.

Study design, size, duration:

This validation study of the CARTR Plus database examined IVF cycles performed in 2015 using patient chart reabstraction. Six clinics across Canada were recruited to participate using a purposive sampling strategy. Fixed random sampling was employed to select 146 patient cycles at each clinic, representing unique patients.

Participants/materials, setting, methods:

Twenty-five data elements were reabstracted from patient charts, which was declared the reference standard. Data were reabstracted by two independent auditors with relevant clinical knowledge after confirming inter-rater reliability. Data from the chart were then compared to

those in CARTR Plus. We calculated kappa coefficients, sensitivity, specificity, positive predictive value, and negative predictive value with 95% confidence intervals (CI) for categorical variables, and mean differences, paired t-tests and intraclass coefficients for continuous variables.

Main results and the role of chance:

Six clinics agreed to participate in this study representing 5 Canadian provinces. The mean age of patients was 35.5 years, which was similar between the two data sources, resulting in a near perfect level of agreement (ICC=0.99; 95% CI: 0.99, 0.99). The agreement for FSH was moderate, ICC=0.68 (95% CI: 0.64, 0.72). There was nearly perfect agreement for cycle type, kappa=0.99 (95% CI: 0.98, 1.00). Over 90% of the cycles in the reabstracted charts used autologous oocytes; however, data on oocyte source was missing for 13% of cycles in CARTR Plus, resulting in a moderate degree of agreement, kappa=0.45 (95% CI: 0.37, 0.52). Embryo transfer and number of embryos transferred had nearly perfect agreement with kappa coefficients greater than 0.90, whereas that for elective single or double embryo transfer was much lower (kappa=0.55; 95% CI: 0.49, 0.61). Validity markers for pregnancy type, number of fetal sacs, and number of fetal hearts on ultrasound were nearly perfect, with kappa coefficients greater than 0.90.

Limitations, reasons for caution:

CARTR Plus contains over 200 variables, of which only 25 were selected for this study. This study, therefore, represents the first step in the validation process for the national database. This foundational validation work should be extended to other database variables in future studies.

Wider implications of the findings:

This study provides the first assessment of the quality of the CARTR Plus database and we found very high data quality for the majority of the variables that were analyzed. The rigorous methodology used can serve as a guide for future validation projects.

Trial registration number:

Not applicable

3.4 Introduction

In 2001, the Royal Commission for New Reproductive Technology estimated that one quarter of a million Canadian couples have difficulty conceiving (1). More recent data from the Canadian Community Health Survey in 2010 estimated that infertility affects 11.5-15.7% of the Canadian population, representing half a million Canadian couples (2). The decision to delay childbearing has become more common due to competing goals of advancing education or pursuing employment opportunities, a trend that is increasing the number of couples that need to rely on assisted reproductive technologies (ART) to conceive. The public health burden and indirect costs of infertility treatments can largely be attributed to the maternal and fetal complications of pregnancy, specifically preterm delivery and multiple gestation, which can increase the cost of care by three-fold (3). In Canada, the incidence of preterm birth in treated infertile couples in 2014 was 24-28% (4), three times higher than the general obstetrical population (5). Other complications that contribute to the indirect cost include ectopic pregnancy, placenta previa, and preeclampsia, which also have a higher incidence among ART pregnancies (6–8). Despite the elevated risks for these conditions, their absolute incidence remains low (9), forcing researchers to often rely on large database-derived cohorts for ART studies for practical reasons.

Both health administrative and registry databases are excellent sources of data for research purposes as they are relatively inexpensive, easily accessible and are collected on a population scale (10,11). These data can be used for evaluating access and quality of health care, health service planning, reporting to governing bodies, and clinical research (10). However, routinely-collected data are generally not collected with the intent of performing research (11). As a result, studies reliant on these data are subject to misclassification, unmeasured

confounding due to missing variables and missing data (12). The accuracy of routinely-collected data is subject to errors from inter-observer discrepancies, documentation problems, illegible charts, missingness of data elements and timeliness of input into the database (13).

In order to establish adequate quality data and avoid misclassification bias in research studies, validation studies that measure the kappa coefficient, sensitivity, specificity, and predictive value of variables contained within health databases are highly recommended as per the RECORD reporting guideline (12). The validity of such databases can be assessed by reabstraction of medical charts (10).

The Canadian Assisted Reproductive Technologies Register (CARTR) Plus is a national database, administered by Better Outcomes Registry & Network (BORN) Ontario, which has collected individual patient data related to *in vitro* fertilization (IVF) and corresponding pregnancy outcomes since 2013 from all 33 ART clinics across Canada. CARTR Plus is the only database in Canada to contain national IVF data, and it has not yet been validated. Because these data may be used to inform policy makers regarding ART funding decisions and as a source of information for clinicians and researchers about current fertility practices and effectiveness and safety of ART treatments in Canada, it was both prudent and timely to conduct a validation study of CARTR Plus. The primary objective of our study was to validate a subset of clinically relevant variables from CARTR Plus to determine the extent to which key study variable are accurately documented in the database.

3.5 Methods

3.5.1 Study design

This validation of the CARTR Plus database examined IVF cycles from January 1, 2015 to December 31, 2015 using patient chart reabstraction as the gold reference standard.

3.5.2 Clinic and chart selection

Upon obtaining ethics approval from The Ottawa Hospital (approval # 20160862-01H), a targeted sample of clinics across Canada was selected and invited to participate in this validation study. Six clinics (out of 33 operating at the time) were selected using purposive sampling to maximize clinic variation in annual cycle volume, geography and mode of data entry into CARTR Plus (i.e., manual entry through a secure web portal versus data upload through an electronic medical record (EMR) system directly to BORN Ontario). The identifier for each clinic was encoded in the database by a 3rd party not involved in the clinic selection and its name was only revealed after the clinic was chosen. We selected our 6 clinics from 5 Canadian provinces. Three of the clinics uploaded their data manually and the other three uploaded data through various EMR systems. We chose two clinics from each "small", "medium", and "large" based on annual cycle volume of ≤500, 501-999, ≥1000, respectively. We only selected from clinics that were considered "good" or "excellent" in their completeness and timeliness in data input. Five of the initial six clinics agreed to participate. A sixth clinic with similar characteristics to the clinic that declined (in the same province, using the same data entry method and with a similar cycle volume) was invited in its place and agreed to participate.

At each study site, a fixed random sample of 146 patient cycles was drawn centrally by a data analyst at BORN Ontario who was not involved in the data extraction or analysis of this project (see below for sample size calculation). Only a single treatment cycle record from a unique patient at each clinic was considered during chart selection. The identified charts were then pulled by the clinic.

3.5.3 Data extraction

We identified twenty-five key data elements from CARTR Plus for validation, chosen based on clinical importance using guidance from the literature, and the consensus of a clinical expert group from the CARTR Plus Steering Committee, Data Elements Committee, and Data Quality Committees (see Appendix 1 for complete list of variables and means or prevalence estimates from 2013-2015). Database variables with missingness greater than 30% were not considered for validation as these elements are likely to have high agreement, but provide little insight into the mechanism behind the missingness (14). Moreover, BORN Ontario has a policy of not reporting data with missingness above this threshold.

Data from each selected chart were abstracted by one of two independent auditors who were blinded to data from CARTR Plus (V.B. and M.J.). To establish inter-rater reliability, the auditors first pilot tested the reabstraction process using 15 patient records at each study site after standard definitions and processes for chart reabstraction were developed. Differences between abstractors were discussed and resolved. Upon reaching 95% agreement for all variables, each auditor then separately abstracted data from remaining sampled charts. The abstracted data were entered and managed in REDCap (hosted at The Children's Hospital of Eastern Ontario Research Institute) with de-identified patient information and each REDCap entry of chart data was double-checked for errors (15).

3.5.4 Statistical analyses

We analyzed characteristics of the sample groups using frequencies for categorical variables, and means and standard deviations or medians and interquartile ranges (IQR) for continuous variables, stratified by source of data (reabstracted versus database-derived). True positives, true negatives, false positives and false negatives were ascertained based on the

comparison between the database code and the element from the clinical chart, where reabstracted data from charts were considered the reference standard. For categorical variables, kappa coefficients, sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV) with 95% confidence intervals (CI) were calculated (see Appendix 2 for sample tables of calculations). For continuous variables, mean differences, using paired t-tests, and ICC were calculated. Kappa coefficients and ICCs were graded according to the levels described by Landis and Koch (see Appendix 3 for full description) (16). For dates, a paired t-test was performed to calculate the mean discrepancy in days between the two sources. The exact date was required to demonstrate agreement. Percent agreement was calculated for the each of the indicators. The primary analysis included combined data from each clinic to determine the agreement across all sites. We then performed sensitivity analyses to assess the measures of agreement at the level of the individual clinics for variables with low measures of validity (i.e., where the kappa coefficient was less than 0.80 or the ICC was less than 0.90). Statistical analyses were performed using SAS statistical software version 9.4 (SAS Institute Inc., Cary, NC).

3.5.5 Hypothesis

We hypothesized that the calculated kappa coefficients and ICCs would be at least 0.80 and 0.90 respectively.

3.5.6 Sample size calculation

We based our sample size calculations on estimating an ICC of 0.90 with a 95% confidence interval yielding a margin of error ± 0.10, using two sources of measurement (database and chart review) for continuous variables (17). For categorical variables, the calculation was based on an anticipated kappa coefficient of 0.80 with a 95% CI yielding a

margin of error ± 0.10 with two sources of measurement, and estimated proportions of each of the categorical variables (18–20); these calculations generated a minimum total sample size of 726. These parameters were based on an earlier perinatal database validation study, and additionally taking feasibility and budget constraints into account (see Appendix 4 for sample size calculation sensitivity analysis) (14). Data elements within categorical variables with a prevalence less than 1% were excluded in the sample size calculations.

Finally, in order to account for missingness of up to 20% for some data elements (based on data from CARTR Plus from 2013-2015), we increased the total sample size to 876 patient charts to guarantee the ± 0.10 margin of error. To ensure adequate accuracy at each site, a fixed sampling approach was undertaken, thus 146 charts were randomly sampled at each of the six participating clinics.

3.6 Results

Six clinics agreed to participate in this study representing 5 Canadian provinces. The cycle volume per clinic in 2015 ranged from 329 to 2212. There were 12 charts that were not retrievable at one clinic site, which were assumed to be missing completely at random. To ensure adequate sample size, we randomly selected an additional 12 charts at this clinic to replace those that could not be retrieved.

3.6.1 Patient intake

The mean age of the patients and oocyte providers (either autologous or donors) was 35.5 and 34.6 years, respectively, and these values were similar between the two data sources (Table 1). The estimated ICCs for patient age and oocyte provider age were 0.99 and 0.86, respectively, indicating almost perfect agreement (Table 2). Among the subset of records with complete information documented on antral follicle count (AFC) and antimüllerian hormone (AMH) in

both data sources, there was almost perfect agreement between CARTR Plus and the reabstracted data with ICCs greater than 0.90 (Table 3). The ICC for follicle stimulating hormone (FSH) level was lower at 0.68 (95% CI: 0.64-0.72), though still in a range indicating strong agreement, and the mean difference between the two sources was not statistically significant. The kappa coefficient for diminished ovarian reserve as a reason for treatment indicated strong agreement (κ=0.72, 95% CI: 0.66, 0.78) and that for advanced female age was moderate (κ=0.60, 95% CI: 0.53, 0.67). See Appendix 5 for complete 2x2 contingency tables for categorical variables. Of note, the PPV for advanced female age was only 0.56 (95% CI: 0.48, 0.63), indicating that if the patient was labelled as such in CARTR plus, there is a 56% probability that she is actually ≥35 years of age.

3.6.2 Dates

The mean difference in days for patient date of birth was 30.6 (Table 2); however, there were only 10 disagreements (1.14%). There were 101 disagreements for cycle start date and 37 disagreements for oocyte collection date. Notably, the mean difference between the two sources for each of these dates was less than one day.

3.6.3 Stimulation

Approximately two thirds of the cycles were fresh IVF cycles and just under one third of the cycles were frozen embryo transfers (FET) in both data sources (Table 1). Seven percent of the cycles were cancelled, of which low ovarian response was the most commonly-cited reason for cancellation.

There was nearly perfect agreement for cycle type, with all measures of validity approaching 1 and percent agreement greater than 99% (Table 3). Two cycles that were documented in the chart as cancelled were reported in CARTR Plus as completed cycles; one of

which was initiated one month after the reabstracted start date, and the other was documented in CARTR Plus as an IVF cycle that did not have an embryo transfer. The weighted kappa coefficient for reason for cancellation was considered moderate.

3.6.4 Retrieval

According to the reabstracted chart data, over 90% of cycles used autologous oocytes, while CARTR Plus reported autologous oocyte use in 82% of cycles (Table 1). Oocyte source was missing for 13% of the cycles in CARTR Plus. The kappa coefficient for oocyte origin was in the moderate range, with overall percent agreement of 87% (Table 3). There was good agreement for fresh autologous oocytes and fresh donor oocytes between the two data sources. Most of the missing data on oocyte source in CARTR Plus was determined to be fresh own oocytes according to patient charts (Appendix 5).

3.6.5 Embryology

An embryo transfer was performed in 80% of cycles that were not cancelled in both data sources (Table 1). Fifty-five percent of these transfers were single embryo transfers and 42% were double embryo transfers. There was nearly perfect to perfect agreement for all measures of validity for both embryo transfer (yes/no) and the number of embryos transferred when performed (Table 3). The kappa coefficient for elective single or double embryo transfer was moderate at 0.55 (95% CI: 0.49, 0.61), with sensitivity, specificity and positive predictive values greater than 0.75. The negative predictive value was much lower, though, at 0.58 (95% CI: 0.52, 0.64).

Embryos were transferred predominantly on day 5 of development, but ranged from day 2 to 6 in both fresh and frozen cycles. In CARTR Plus, embryo transfer day was either not reported for frozen cycles or was reported as day 0. For fresh cycles, there was nearly perfect

agreement for day of transfer between the two data sources. However, as transfer day for frozen cycles is either not recorded or is recorded as day 0, the ICC was unmeasurable (Table 2).

Over 90% of embryos were frozen using the vitrification technique after a fresh cycle. Cryopreservation technique for frozen cycles was derived from the primary fresh cycle. Approximately 80% of embryos were cryopreserved by vitrification in the FETs. Eighty-three cycles were missing method of cryopreservation in the CARTR Plus database (Appendix 5). The weighted kappa coefficients for cryopreservation technique overall were quite strong for both frozen and fresh cycles. Among FET cycles, the percent agreement was much lower than for IVF cycles. When broken down by technique, the kappas for vitrification and slow-freeze in frozen cycles were moderate. There was almost perfect agreement between CARTR Plus and the reabstracted data for number of embryos thawed and number of utilizable embryos after thawing.

3.6.6 Pregnancy

Thirty-three percent of all initiated cycles and 44% of cycles with an embryo transfer resulted in a clinical intrauterine pregnancy (Table 1) according to the reabstracted data. Among these clinical pregnancies, ultrasound assessment detected one fetal sac in 78% and a single fetal heart in 71% according to the reabstracted data. Chorionicity was only reported for multi-fetal gestations, representing 65 pregnancies, of which dichorionicity was most prevalent. There was very strong agreement for pregnancy type, number of fetal sacs on ultrasound and number of fetal hearts on ultrasound (Table 3). The overall kappa coefficients for all three variables were 0.90 or higher. Among the multi-fetal gestation pregnancies, however, the agreement for chorionicity was only 74% with a kappa coefficient of 0.29 (95% CI: 0.08, 0.51).

3.6.7 Sensitivity analysis - missing charts

An analysis of the 12 missing charts using data from CARTR Plus revealed similar patient and oocyte provider ages to the 876 charts included in this study. The mean FSH, AFC and AMH of the patients with missing charts were similar to that in the CARTR Plus. The oocyte source was fresh autologous for all ongoing cycles. There was a higher prevalence of advanced female age in the patients with missing charts compared to the study population. Lastly, all the frozen cycles were cryopreserved using slow-freeze technology rather than vitrification (see Appendix 6 for full results).

3.6.8 Sensitivity analysis - assessment of clinic-specific results

Patient intake

For advanced female age, one of the six clinics had a percent agreement lower than 85% (Table 4). There were 52 disagreements where the patient was reported as being advanced age in CARTR Plus, but not in the reabstracted dataset. However, if advanced female was reported in the reabstracted data, they were consistently documented as such in CARTR Plus.

Follicle stimulating hormone levels, antimüllerian hormone levels and antral follicle count demonstrated particularly low agreement in one clinic. However, percent agreement was generally poor for FSH and AFC across all clinics with estimates ranging from 22.6% to 81.5%.

Stimulation

The most common reason for a cancelled cycle was low ovarian response, followed by premature ovulation, and other, which included patient request. There was no particular trend in the reason for a cancelled cycle among the clinics.

Retrieval

The estimates for oocyte origin were attenuated largely due to missing data. The degree of missingness was higher in the clinics that entered their data automatically through an EMR, where the percent agreement ranged from 74% to 84%. For the clinics that entered data manually, the percent agreement ranged from 91% to 99%.

Embryology

The estimates for cryopreservation were much stronger for IVF compared to FET cycles. Among FET cycles, agreement was poor in four of the six clinics. A post-hoc sensitivity analysis was performed to determine if there was a difference between elective single embryo transfers (eSET) and elective double embryo transfers (eDET) (Table 5). The percent agreement for eSET and eDET was much lower at one particular clinic compared to the others (Table 6). Furthermore, there was a stronger agreement among eSET on day 3 of transfer compared to eDET. There was little difference between eSET and eDET among the other clinics or when stratified by cycle type.

3.7 Discussion

The CARTR Plus database, which began collecting data in 2013, is the only national database in Canada that collects detailed clinical information on IVF treatments, diagnoses and outcomes. Our study, which assessed the validity of CARTR Plus data compared with reabstracted patient chart data, demonstrated that for most of the data elements selected, markers of validity were quite strong. The areas with moderate agreement were FSH levels, reason for treatment cycle, reason for a cancelled cycle, oocyte origin, elective single or double embryo transfer and chorionicity.

3.7.1 FSH, AMH, AFC

The value of testing for diminished ovarian reserve remains controversial in the literature. Ovarian reserve tests are used to predict how well women will respond to infertility treatment and identify patients who are at risk of diminished ovarian reserve (21). Elevated levels of FSH, low AFC and AMH levels have been associated with diminished reserve (22–25). In our study, we identified a minimum of 120 more lab test results and ultrasound reports in the chart reabstraction for FSH, AMH and AFC than were recorded in the CARTR Plus database. The FSH and AMH tests are often performed at laboratories off-site and results then scanned into patients' charts once they become available. These results may be challenging to find on the chart if there are many tests performed, or they may be recorded incorrectly or not at all. Furthermore, the estimates of agreement were especially poor for FSH. When values were recorded for AFC and AMH, there was good agreement. During the data reabstraction process, we noted that many patients underwent frequent FSH tests, which led to numerous disagreements between the auditors. As such, it is not surprising there was significant disagreement between the two data sources. We would not recommend using these variables in future research projects until the data entry process into CARTR Plus can be clarified and improved.

3.7.2 Advanced female age

As a woman ages, the number of oocytes remaining in her ovaries decreases and the probability of conception diminishes (26). The Society of Obstetricians and Gynaecologists of Canada (SOGC) recommends referral to an IVF clinic among women aged 35 or older after 6 months of trying to conceive (21). However, there is no specific definition for "advanced female age", likely a result of the continuous and progressive decline in live birth rates with advancing age. Problematically, in the CARTR Plus data dictionary, there was no specified definition for

this variable until 2016, now delineated by age greater than or equal to 35 years. The lack of consistent designation likely contributed to a poorer degree of agreement.

The estimated kappa coefficient for advanced female age was 0.60 while the percent agreement between the two sources was 82.1%. The discrepancy between the kappa coefficient and percent agreement demonstrates the importance of reporting multiple measures of validity to determine whether the data element is utilizable or whether changes are required to the database or data entry procedures. While the percent agreement is a crude estimate, the kappa statistic adjusts for agreement due to chance, making it a more robust measure. The kappa coefficient, however, is affected by the distribution of positive and negative agreements (27). Additionally, if the estimated prevalence of a condition is unequal between two data sources, the kappa coefficient will be biased, leading to a larger estimate(28). Juurlink *et al* (2006) argue that in certain cases, PPV and sensitivity are more valuable than the kappa coefficient (11,29). With no specific guideline indicating which measure is ideal for reporting and heterogeneity in the literature on the chosen measurement, Benchimol *et al* (2010) encourage reporting a minimum of four different measures of validity with corresponding confidence intervals (11).

3.7.3 Oocyte origin

The overall kappa for oocyte origin and specific kappa coefficients for fresh own oocytes were in the moderate range. However, the other markers of agreement, including percent agreement, sensitivity, specificity and PPV were more in keeping with strong agreement between the two data sources. Importantly, 106 cycles, (10% of charts) in CARTR Plus were missing this element when the cycle was not cancelled. The chart reabstraction data indicated that these missing values in the CARTR Plus database were predominantly fresh own oocytes, followed by "other", which was largely frozen donor oocytes. We would, therefore, recommend if this

variable is used in future research or surveillance projects, that an imputation strategy be considered that weights the probability that missing values were fresh own oocytes more heavily, followed by frozen donor oocytes.

3.7.4 Elective single embryo transfer

Elective single or double embryo transfer is defined as the selection of one (eSET) or two (eDET) cleavage- or blastocyst-stage embryos to transfer from a larger pool of viable embryos (30). The risk of multiple pregnancy after single embryo transfer is significantly reduced for both cleavage- and blastocyst-stage embryos compared to double embryo transfer(31). However, the decision to proceed with eSET, eDET or multiple embryo transfer is based on a number of factors, including policy recommendations to reduce the risk of twin or high-order multiple gestations, patient prognosis, and embryo quality (31–34).

The Canadian Fertility and Andrology Society (CFAS), the American Society of Reproductive Medicine (ASRM) and the National Institute for Health and Care Excellence (NICE) have published recommendations to reduce the risk of multiple pregnancies by minimizing the number of embryos that are transferred in a single cycle while maintaining an adequate live birth rate (30,34,35). Our study demonstrated poor agreement in the measures of validity for eSET or eDET. Upon further examination, this difference was largely attributed to one clinic, where the individual kappa was 25%. While the overall percent agreement for eSET was 83.2%, more than 80% of the disagreements were a result of the clinics mislabeling the transfer as non-elective when it was truly elective based on the reference standard. Though, the agreement for the number of embryos transferred was nearly perfect. Newer studies are demonstrating that pregnancy rates may be higher and the prevalence of low birth weight in neonates lower for eSET compared to non-elective SET (36,37). Based on the error trend we

found, the risk of poor pregnancy outcomes with non-elective SET would likely be attenuated if such as study were carried out using the CARTR Plus database, assuming non-differential misclassification (38).

3.7.5 Chorionicity

Sixty-five cycles were classified as multiple gestation, representing 20% of ongoing clinical intrauterine pregnancies where there was more than one fetal heart on ultrasound. Among these pregnancies, dichorionicity was most prevalent. In 2015, the prevalence of multiple gestation in the Canadian ART population was estimated to be 11% of ongoing clinical pregnancies (39). Thus, our sample over-represents multiple gestation pregnancies compared to the overall ART population. Although there were few disagreements between the two data sources, the PPVs for two fetal hearts and two fetal sacs should be interpreted cautiously, since PPV is highly influenced by the prevalence in the population (40). With a sample prevalence greater than the true population prevalence, we expect that our PPV estimate was higher than the true value.

Additionally, for some of these pregnancies, the number of fetal sacs, hearts and placenta were based on an ultrasound performed in the clinic, at which point chorionicity may not have been visible. Other patients went to outside clinics for a later ultrasound, especially if they were undergoing treatments far from their place of residence. Thirteen cycles identified as dichorionic in CARTR Plus could not be corroborated on reabstraction. These entries may have been based on either the assumption that two embryos transferred should be dichorionic or from a postpartum pathology report that described two distinct placentas. These speculations cannot be verified as these data were not available at the time of reabstraction. Monochorionicity confers a significantly increased risk to the pregnancy than dichorionicity with respect to intrauterine fetal

demise, preterm delivery, placental insufficiency (41). These complications, which are more prevalent in ART pregnancies (42,43), may be inappropriately described if not accurately reported in the registry.

3.7.6 Reason for cancelled cycle

Reason for cycle cancellation had a kappa coefficient considered "moderate" due to a combination of small sample size and missing data in the patient chart (16). The reason for cancellation was not documented in 18% of the cancelled cycles. By this standard, caution would be advised for any research or data reporting using this variable at the present time.

3.7.7 Study limitations

Our sample of clinics was assembled to represent the Canadian population undergoing ART from clinics of varying sizes and varying regions of the country, and using different modalities of data entry. For practical reasons, we also selected clinics that were most adherent to timely and complete data submission to CARTR Plus. As such, our results may represent more reliable clinics from a data collection perspective, which may limit generalizability. However, upon initiating improvements with respect to the way elements are entered, including training to those who input the data, these estimates will serve as targets for the rest of the clinics. Although only 25 data elements were evaluated in this project, CARTR Plus contains over 200 data elements. Nevertheless, our study represents the first formal assessment of data quality in CARTR Plus and we specifically selected variables for inclusion in our study based on high clinical importance. Ideally, a formal validation should be performed for other database variables prior to use.

Finally, each data entry per clinic was input by one or few people per site; these cycles are, therefore, clustered by site (i.e. they are not independent from one another). As a result, the

generated confidence intervals around the estimates are likely narrower than they would be had we analyzed the data as correlated observations (44,45). Due to limited resources and funds, we were unable to collect data on the scale required for minimum sample size. Future validation studies should account for the intracluster correlation.

3.7.8 What our study adds to current literature

Notwithstanding these limitations, this study is strengthened by the rigorous methodology we adopted to ensure that abstractors were meticulous in the data reabstraction process. Definitions for data collection processes were created prior to initiation, inter-rater reliability was confirmed prior to abstracting data independently, and each chart was double-checked to reduce clerical errors. Moreover, the participating clinics were open and compliant with record sharing. Finally, this is the first study to our knowledge evaluating the validity of a national ART database performed in accordance with recommendations for reporting measures of both validation of administrative databases and diagnostic accuracy studies (12,46).

Despite increasing utilization of health administrative databases and registries in research investigating pregnancy outcomes of fertility treatments, there is a paucity of validation studies in the literature for these routinely-collected data. The Society for Assisted Reproductive Technology (SART) in the United States publishes an annual surveillance report with an appendix indicating only the percentage disagreement of selected variables in the American fertility database when compared with a sample of medical charts (47). As previously described, percent disagreement does not account for agreement/disagreement due to chance, limiting its measurability of accuracy. Additionally, the recommended measures of validity including kappa coefficients, sensitivity, specificity or negative and positive predictive values were not utilized,

thereby making it difficult to interpret the accuracy of the presented information or to compare with our own results (11,12).

3.8 Conclusion

In conclusion, our study provides the first assessment of the quality of the CARTR Plus database. This is also the first evaluation of validity of an ART database adherent to reporting guidelines for validation studies. The methodologic rigor utilized in the design and analysis should serve as a guideline for future studies of this nature. The majority of elements we assessed demonstrated a high level of validity which can be used for future projects. We have identified key data points that are either too often lacking or inconsistent with chart data, indicating changes in the data entry process may be required.

Utilization of CARTR Plus data is important in the analysis of Canadians' access to this aspect of the health care system, and determination of the implications of fertility treatments on pregnancy outcomes. Quality improvement initiatives including benchmarking and dashboards for clinics also rely on these data. Our study provides direction for further refinement and improvement for data collection and entry into a national database. This will allow for accurate, meaningful clinical research and health policy initiatives in the future.

Acknowledgments

We would like to thank Dr. Monica Taljaard who helped with the sample size calculation. We would also like to thank Drs. Hao Wang and Mary Guo for all their hard work in pulling data in an expedited fashion. Finally, we are very appreciative of all the participating clinics who graciously welcomed us into their office spaces.

3.9 References

1. Norris S. Reproductive infertility: Prevalence, causes, trends and treatments. Parliam Res Branch Libr Parliam. 2001. http://publications.gc.ca/collections/Collection-R/LoPBdP/EB-e/prb0032-e.pdf. Accessed December 20, 2016.
2. Bushnik T, Cook JL, Yuzpe AA, Tough S, Collins J. Estimating the prevalence of infertility in Canada. Hum Reprod. 2012;27(3):738-746. doi:10.1093/humrep/der465.
3. Connolly MP, Hoorens S, Chambers GM. The costs and consequences of assisted reproductive technology: An economic perspective. Hum Reprod Update. 2010;16(6):603-613. doi:10.1093/humupd/dmq013.
4. Canadian Assisted Reproductive Technologies Register (CARTR) Plus. Preliminary Treatment Cycle Data for 2014. Ottawa, ON; 2015.
5. Public Health Agency of Canada. Perinatal Health Indicators for Canada 2017.; 2017.
6. Pandey S, Shetty A, Hamilton M, Bhattacharya S, Maheshwari A. Obstetric and perinatal outcomes in singleton pregnancies resulting from IVF/ICSI: a systematic review and meta-analysis. Hum Reprod Update. 2012;18(5):485-503. doi:10.1093/humupd/dms018.
7. Romundstad LB, Romundstad PR, Sunde A, von Düring V, Skjaerven R, Vatten LJ. Increased risk of placenta previa in pregnancies following IVF/ICSI; a comparison of ART and non-ART pregnancies in the same mother. Hum Reprod. 2006;21(9):2353-2358. doi:10.1093/humrep/del153.
8. Maymon R. Controversies and problems in the current management of tubal pregnancy. Hum Reprod Update. 1996;2(6):541-551. doi:10.1093/humupd/2.6.541.
9. Sazonova A, Källen K, Thurin-Kjellberg A, Wennerholm U-B, Bergh C. Obstetric outcome after in vitro fertilization with single or double embryo transfer. Hum Reprod. 2011;26(2):442-450. doi:10.1093/humrep/deq325.
10. Iron K, Manuel D. Quality assessment of administrative data (QuAAD): an opportunity for enhancing Ontario's health data. Inst Clin Eval Stud. 2007.
11. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. J Clin Epidemiol. 2010;64(2011):821-829. doi:10.1016/j.jclinepi.2010.10.006.
12. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLoS Med. 2015;12(10). doi:10.1371/journal.pmed.1001885.
13. Hierholzer Jr WJ. Health care data, the epidemiologist's sand: Comments on the quantity and quality of data. Am J Med. 1991;91(3):21S-26S.
14. Dunn S, Bottomley J, Ali A, Walker M. 2008 Niday Perinatal Database quality audit: report of a quality assurance project. Chronic Dis Inj Can. 2011;32(1):32-42.
15. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377-381. doi:10.1016/j.jbi.2008.08.010.
16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-174. http://www.jstor.org/stable/2529310. Accessed December 15, 2017.
17. Zou G. Sample size formulas for estimating intraclass correlation coefficients with precision

and assurance. Stat Med. 2012;31(29):3972-3981. doi:10.1002/sim.5466.
18. Donner A, Eliasziw M. Sample size requirements for reliability studies. Stat Med. 1987;6(4):441-448. doi:10.1002/sim.4780060404.
19. Donner A, Zou G. Interval estimation for a difference between intraclass kappa statistics. Biometrics. 2002;58(1):209-215. doi:10.1111/j.0006-341X.2002.00209.x.
20. Bartfay E, Donner A. Statistical inferences for interobserver agreement studies with nominal outcome data. J R Stat Soc Ser D Stat. 2001;50(2):135-146. doi:10.1111/1467-9884.00266.
21. Liu K, Case A, Cheung AP, et al. Advanced Reproductive Age and Fertility. J Obstet Gynaecol Canada. 2011;33(11):1165-1175. doi:10.1016/S1701-2163(16)35087-3.
22. Sherman BM, West JH, Korenman SG. The Menopausal Transition: Analysis of LH, FSH, Estradiol, and Progesterone Concentrations During Menstrual Cycles of Older Women. J Clin Endocrinol Metab. 2009;42(4):629-636.
23. Ng EHY, Tang OS, Ho PC. The significance of the number of antral follicles prior to stimulation in predicting ovarian responses in an IVF programme. Hum Reprod. 2000;15(9):1937-1942. doi:10.1093/humrep/15.9.1937.
24. Chang M, Chiang C, Hsieh T, Soong Y, Hsu K. Use of the Antral Follicle Count to Predict the Outcome of Assisted Reproductive Technologies. Fertil Steril. 1998;69(3):505-510.
25. Durlinger A, Visser J, Themmen A. Regulation of ovarian function: the role of anti-Mullerian hormone. Reproduction. 2002;124(5):601-609. doi:10.1530/rep.0.1240601.
26. Broekmans FJ, Soules MR, Fauser BC. Ovarian Aging: Mechanisms and Clinical Consequences. Endocr Rev. 2009;30(5):565-593.
27. Feinstein AR, Cicchetti D V. High agreement but low Kappa: I. the problems of two paradoxes. J Clin Epidemiol. 1990;43(6):543-549. doi:10.1016/0895-4356(90)90158-L.
28. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol. 1993;46(5):423-429. doi:10.1016/0895-4356(93)90018-V.
29. Juurlink D, Preyra C, Croxford R, et al. Canadian Institute for Health Information discharge abstract database: a validation study. Toronto Inst Clin Eval Sci. 2006.
30. Committee of the Society for Assisted Reproductive Technology P, Committee of the American Society for Reproductive Medicine P. Elective single-embryo transfer. Fertil Steril. 2012;97:835-842. doi:10.1016/j.fertnstert.2011.11.050.
31. Pandian Z, Marjoribanks J, Ozturk O, Serour G, Bhattacharya S. Number of embryos for transfer following in vitro fertilisation or intra-cytoplasmic sperm injection. Pandian Z, ed. Cochrane Database Syst Rev. 2013;(7). doi:10.1002/14651858.CD003416.pub4.
32. Greenblatt E. Advisory Process for Infertility Services Key Recommendations Report.; 2015. http://www.health.gov.on.ca/en/public/programs/ivf/docs/ivf_report.pdf. Accessed July 26, 2016.
33. Peeraer K, Debrock S, Laenen A, et al. The impact of legally restricted embryo transfer and reimbursement policy on cumulative delivery rate after treatment with assisted reproduction technology. Hum Reprod. 2014;29(2):267-275. doi:10.1093/humrep/det405.
34. Min J, Sylvestre C. Guidelines on the number of embryos transferred. Can Fertil Androl Soc. 2013. http://www.cfas.ca/images/stories/pdf/cfas_cpg_embryo_transfer_2013.pdf. Accessed July 25, 2016.
35. National Institute for Health and Care Excellence. Fertility problems: assessment and treatment. Natl Inst Heal Care Excell. 2013. https://www.nice.org.uk/guidance/cg156/resources/fertility-problems-assessment-and-treatment-35109634660549. Accessed July 27, 2016.

36. Mersereau J, Stanhiser J, Coddington C, Jones T, Luke B, Brown MB. Patient and cycle characteristics predicting high pregnancy rates with single-embryo transfer: an analysis of the Society for Assisted Reproductive Technology outcomes between 2004 and 2013. Fertil Steril. 2017;108(5):750-756. doi:10.1016/J.FERTNSTERT.2017.07.1167.
37. Styer AK, Luke B, Vitek W, et al. Factors associated with the use of elective single-embryo transfer and pregnancy outcomes in the United States, 2004–2012. Fertil Steril. 2016;106(1):80-89. doi:10.1016/J.FERTNSTERT.2016.02.034.
38. Armstrong BG. Effect of measurement error on epidemiological studies of environmental and occupational exposures. Occup Environ Med. 1998;55(10):651-656. http://www.ncbi.nlm.nih.gov/pubmed/9930084. Accessed December 11, 2017.
39. Canadian Assisted Reproductive Technologies Registry (CARTR) Plus. Final Treatment Cycle and Pregnancy Outcome Data for 2015. Ottawa, ON; 2017.
40. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. BMJ. 1994;309(6947):102. http://www.ncbi.nlm.nih.gov/pubmed/8038641. Accessed July 17, 2017.
41. Hack K, Derks J, Elias S, et al. Increased perinatal mortality and morbidity in monochorionic versus dichorionic twin pregnancies: clinical implications of a large Dutch cohort study. BJOG An Int J Obstet Gynaecol. 2007;115(1):58-67. doi:10.1111/j.1471-0528.2007.01556.x.
42. Mateizel I, Santos-Ribeiro S, Done E, et al. Do ARTs affect the incidence of monozygotic twinning? Hum Reprod. 2016;31(11):2435-41. doi:10.1093/humrep/dew216.
43. Kanter JR, Boulet SL, Kawwass JF, Jamieson DJ, Kissin DM. Trends and correlates of monozygotic twinning after single embryo transfer. Obstet Gynecol. 2015;125(1):111-117. doi:10.1097/AOG.0000000000000579.
44. Zyzanski SJ, Flocke SA, Dickinson LM. On the nature and analysis of clustered data. Ann Fam Med. 2004;2(3):199–200.
45.Galbraith S, Daniel JA, Vissel B. A study of clustered data and approaches to its analysis. J Neurosci. 2010;30(32):10601–10608.
46. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. Clin Chem. 2003;49(1):1-6.
47. Centers for Disease Control and Prevention, American Society for Reproductive Medicine, Society for Assisted Reproductive Technology. 2014 Assisted Reproductive Technology Fertility Clinic Success Rates Report. Atlanta; 2016.
48. Almog B, Shehata F, Suissa S, et al. Age-related normograms of serum antimüllerian hormone levels in a population of infertile women: a multicenter study. Fertil Steril. 2011;95(7)2359-2363.

3.10 Figures and tables

Table 1. Description of study variables by data source

Table 2. Measures of agreement for continuous variables

Table 3. Measures of agreement for categorical variables

Table 4. Sensitivity analysis: Percent agreement of problematic variables by clinic

Table 5. Prevalence of elective single or double embryo transfer by data source

Table 6. Sensitivity analysis: Percent agreement of elective single embryo transfer and double embryo transfer by clinic, cycle type, and day of transfer

Table 1. Description of study variables by data source

	CARTR Plus						Reabstracted Data					
Variable	**N**	**%**	**Mean**	**SD**	**Median**	**IQR**	**N**	**%**	**Mean**	**SD**	**Median**	**IQR**
Intake of patient												
Patient age (years)	876		35.5	4.63	35	(32-39)	876		35.5	4.66	35	(32-39)
Oocyte provider age (years)	860		34.2	4.58	31	(34-38)	874		34.7	7.06	34	(31-38)
Reason for treatment cycle												
Diminished ovarian reserve												
Yes	149	17.0					150	17.1				
No	727	83.0					726	82.9				
Advanced female age												
Yes	187	21.4					125	14.3				
No	689	78.7					751	85.7				
FSH (IU/L)	509		7.20	3.59	6.70	(5.20-8.00)	734		6.81	2.95	6.30	(5.00-7.90)
AFC (# follicles)	362		16.2	11.8	13.5	(8.00-21.0)	583		16.3	11.2	14.0	(9.00-21.0)
AMH (ng/dL)*	77		2.20	2.14	1.30	(0.90-2.80)	202		2.46	2.56	1.60	(0.80-3.30)
Stimulation												
Cycle type												
IVF	607	69.3					602	68.7				
FET	245	28.0					245	28.0				
Frozen oocyte IVF	14	1.60					18	2.05				
Oocyte banking	10	1.14					11	1.26				
Cancelled cycle												
Yes	60	6.85					62	7.08				
No	816	93.2					814	92.9				
Reason for cancelled												
Low ovarian response	47	78.3					38	61.3				
Premature ovulation	3	5.00					3	4.84				
Other	10	16.7					10	16.1				
Missing	0	0.00					11	17.7				

Retrieval												
Oocyte origin												
Fresh own oocytes	672	82.4					750	92.1				
Fresh donor oocytes	35	4.29					39	4.79				
Other	3	0.37					23	2.83				
Missing	106	13.0					2	0.25				
Embryo transfer												
Embryo transfer												
Yes	655	80.3					654	80.3				
No	152	18.6					160	19.7				
Missing	9	1.10					0	0.00				
ET day	655		2.57	2.12			647		4.39	1.16		
Fresh cycles	418		4.02	1.09	5.00	(3.00-5.00)	418		4.00	1.10	5.00	(3.00-5.00)
2	26	6.22					27	6.46				
3	166	39.7					168	40.2				
5	225	53.8					222	53.1				
6	1	0.24					1	0.24				
Frozen cycles	237		0.00	0.00			229		5.10	0.93	5.00	(5.00-6.00)
2	0						1	0.42				
3	0						21	8.90				
4	0						13	5.51				
5	0						120	50.9				
6	0						69	29.2				
>6	0						5	2.11				
Missing							7	2.97				
# Embryos transferred												
1	362	55.3					362	55.4				
2	273	41.7					273	41.7				
3	17	2.60					16	2.45				
4	3	0.46					3	0.46				
eSET or eDET												
Yes	387	59.1					475	72.6				
No	268	40.9					177	27.1				
Missing	0	0.00					2	0.31				

Embryology												
Embryo cryopreservation												
IVF												
Vitrification	267	94.0					267	94.0				
Slow-freeze	17	5.99					16	5.63				
Mixed	0	0.00					1	0.35				
FET												
Vitrification	142	58.0					194	79.2				
Slow-freeze	20	8.16					49	20.0				
Mixed	0	0.00					1	0.41				
Missing	83	33.9					1	0.41				
# embryos thawed	245		1.95	1.76	1.00	(1.00-2.00)	244		1.95	1.77	1.00	(1.00-2.00)
# embryos utilizable after thaw	245		1.47	0.99	1.00	(1.00-2.00)	244		1.46	0.92	1.00	(1.00-2.00)
Pregnancy												
Pregnancy type												
Not pregnant	522	59.6					519	59.3				
Biochemical	45	5.14					52	5.94				
Clinical intrauterine	285	32.5					289	33.0				
Other	2	0.23					2	0.23				
Unknown	22	2.51					14	1.60				
# fetal sac												
1	221	77.5					222	76.8				
2	63	22.1					62	21.5				
3	1	0.35					2	0.69				
Missing	0	0.00					3 †	1.04				
# fetal heart												
0	26	9.12					28	9.69				
1	202	70.9					206	71.3				
2	56	19.7					52	18.0				
3	1	0.35					1	0.35				
Missing	0	0.00					2	0.69				

Chorionicity				
1	4	5.97	5	6.15
2	59	88.1	45	69.2
3	1	1.49	1	1.54
Missing	3	4.48	15	23.1

AFC: Antral follicle count, AMH: Antimüllerian hormone, CARTR Plus: Canadian Assisted Reproductive Technologies Register Plus, eDET: Elective double embryo transfer, eSET: Elective single embryo transfer, ET: Embryo transfer, FSH: Follicle stimulating hormone; FET: Frozen embryo transfer, IQR: Interquartile range, IVF: *in vitro* fertilization, N: Number or patients, SD: Standard deviation

* AMH levels were converted from pmol/L to ng/dL using a conversion factor of 7.14 (48)

ŧ There was an error in data entry which was recoded as missing

Table 2. Measures of agreement for continuous variables

	N	ICC	95% CI	Mean difference	95% CI	% Agreement
Dates						
Patient date of birth	876			30.58 days	(-7.92, 69.1)	98.9
Cycle start date	876			0.24 days	(0.03, 0.46)	88.5
Oocyte collection date	798			0.44 days	(-4.26, 5.14)	95.8
Intake						
Patient age (years)	876	0.99	(0.99, 0.99)	0.04	(0.01, 0.06)	96.6
Oocyte provider age (years)	858	0.86	(0.84, 0.88)	0.17	(0.01, 0.33)	92.5
FSH (IU/L)	503	0.68	(0.64, 0.72)	0.14	(-0.02, 0.30)	64.4
AFC (# follicles)	342	0.92	(0.91, 0.94)	0.59	(0.27, 0.93)	62.1
AMH (ng/dL)	69	0.92	(0.89, 0.95)	0.13	(-0.04, 0.29)	83.2
Embryo transfer						
ET Day						
Fresh cycles (days)	402	0.98	(0.98, 0.99)	1.79	(1.60, 1.99)	99.5
Frozen cycles (days)	229	0.00		5.10	(4.97, 5.22)	3.27*
Embryology						
# embryos thawed	244	1.00	(0.99, 1.00)	0.00	(-0.02, 0.02)	99.4
# embryos utilizable after thaw	244	0.93	(0.92, 0.95)	0.01	(-0.04, 0.05)	96.2

AFC: Antral follicle count, AMH: Antimullerian hormone, CI: Confidence interval, ET: Embryo transfer, FSH: Follicle stimulating hormone, ICC: Intraclass correlation coefficient, N: Number of patients

* There was no either no recorded day of transfer or ET day was missing for FET cycles in the CARTR Plus database.

Table 3. Measures of agreement for categorical variables

	κ	95% CI	SN	95% CI	SP	95% CI	PPV	95% CI	NPV	95% CI	% Agreement
Diminished ovarian reserve	0.72	(0.66, 0.78)	0.77	(0.69,0.83)	0.95	(0.94, 0.97)	0.77	(0.70, 0.84)	0.95	(0.93, 0.97)	92.1
Advanced female age	0.60	(0.53, 0.67)	0.83	(0.75, 0.89)	0.89	(0.86, 0.91)	0.56	(0.48, 0.63)	0.97	(0.95, 0.98)	82.1
Cycle type	0.99	(0.98, 1.00)									99.4
IVF	0.99	(0.98, 1.00)	1.00	(0.99, 1.00)	0.98	(0.96, 0.99)	0.99	(0.98, 1.00)	1.00	(0.99, 1.00)	99.4
FET	1.00	(1.00, 1.00)	1.00	(0.99, 1.00)	1.00	(0.99, 1.00)	1.00	(0.99, 1.00)	1.00	(0.99, 1.00)	100
Frozen oocyte IVF	0.87	(0.75, 1.00)	0.78	(0.52, 0.94)	1.00	(1.00, 1.00)	1.00	(0.77, 1.00)	1.00	(0.99, 1.00)	99.5
Oocyte banking	0.95	(0.86, 1.00)	0.91	(0.59, 1.00)	1.00	(1.00, 1.00)	1.00	(0.69, 1.00)	1.00	(0.99, 1.00)	99.9
Cancelled cycle	0.98	(0.96, 1.00)	0.97	(0.89, 1.00)	1.00	(1.00, 1.00)	1.00	(0.94, 1.00)	1.00	(0.99, 1.00)	99.8
Reason cancelled	0.47*	(0.28, 0.67)									72.6
Low ovarian response	0.60	(0.40, 0.80)	0.97	(0.86, 1.00)	0.58	(0.37, 0.78)	0.79	(0.64, 0.89)	0.93	(0.68, 1.00)	82.3
Premature ovulation	0.65	(0.20, 1.00)	0.67	(0.09, 0.99)	0.98	(0.91, 1.00)	0.67	(0.09, 0.99)	0.98	(0.91, 1.00)	96.8
Other	0.52	(0.23, 0.81)	0.60	(0.26, 0.88)	0.92	(0.81, 0.98)	0.60	(0.26, 0.88)	0.92	(0.81, 0.98)	87.1
Oocyte origin	0.45	(0.37, 0.52)									86.7
Fresh own oocyte	0.56	(0.48, 0.64)	0.89	(0.87, 0.91)	0.98	(0.92, 1.00)	1.00	(0.99, 1.00)	0.44	(0.36, 0.52)	89.9
Fresh donor oocyte	0.89	(0.81, 0.96)	0.85	(0.69, 0.94)	1.00	(0.99, 1.00)	0.94	(0.81, 0.94)	0.99	(0.98, 1.00)	99.0
Other	0.15	(-0.04, 0.33)	0.09	(0.01, 0.28)	1.00	(0.99, 1.00)	0.67	(0.09, 0.67)	0.97	(0.96, 0.99)	97.3
Embryo transfer	1.00	(1.00, 1.00)	1.00	(0.99, 1.00)	1.00	(0.98, 1.00)	1.00	(0.99, 1.00)	1.00	(0.98, 1.00)	98.9
# embryos transferred	1.00	(0.99, 1.00)									99.9
1	1.00	(1.00, 1.00)	1.00	(0.99, 1.00)	1.00	(0.99, 1.00)	1.00	(0.99, 1.00)	1.00	(0.99, 1.00)	100
2	1.00	(0.99, 1.00)	1.00	(0.98, 1.00)	1.00	(0.99, 1.00)	1.00	(0.99, 1.00)	1.00	(0.99, 1.00)	99.9
3	0.97	(0.91, 0.97)	1.00	(0.79, 1.00)	1.00	(0.99, 1.00)	0.94	(0.71, 0.94)	1.00	(0.99, 1.00)	99.9
4	1.00	(1.00, 1.00)	1.00	(0.29, 1.00)	1.00	(0.99, 1.00)	1.00	(0.29, 1.00)	1.00	(0.99, 1.00)	100
ET day: fresh cycles	0.99	(0.97, 1.00)									99.3
2	0.98	(0.94, 1.00)	0.96	(0.81, 1.00)	1.00	(0.99, 1.00)	1.00	(0.87, 1.00)	1.00	(0.99, 1.00)	99.8
3	0.99	(0.98, 1.00)	0.99	(0.96, 1.00)	1.00	(0.99, 1.00)	1.00	(0.98, 1.00)	0.99	(0.97, 1.00)	99.5

5	0.99	(0.97, 1.00)	1.00	(0.98, 1.00)	0.98	(0.96, 1.00)	0.99	(0.96, 1.00)	1.00	(0.98, 1.00)	99.3
6	1.00	(1.00, 1.00)	1.00	(0.03, 1.00)	1.00	(0.99, 1.00)	1.00	(0.03, 1.00)	1.00	(0.99, 1.00)	100
eSET or eDET	0.55	(0.49, 0.61)	0.76	(0.72, 0.80)	0.88	(0.82, 0.92)	0.94	(0.91, 0.96)	0.58	(0.52, 0.64)	83.2
Embryo cryopreservation-IVF	0.91*	(0.78, 1.00)									98.2
Vitrification	0.86	(0.73, 0.98)	0.99	(0.96, 1.00)	0.94	(0.71, 1.00)	1.00	(0.98, 1.00)	0.80	(0.56, 0.94)	98.2
Slow-freeze	0.97	(0.90, 1.00)	1.00	(0.79, 1.00)	1.00	(0.98, 1.00)	0.94	(0.71, 1.00)	1.00	(0.99, 1.00)	99.7
Mixed	0.00		0.00		1.00		-		-		62.1
Embryo cryopreservation-FET	0.89*	(0.77, 1.00)									64.8
Vitrification	0.49	(0.38, 0.59)	0.72	(0.65, 0.79)	0.96	(0.86, 1.00)	0.99	(0.95, 1.00)	0.47	(0.37, 0.57)	76.6
Slow-freeze	0.49	(0.35, 0.64)	0.39	(0.25, 0.54)	0.99	(0.97, 1.00)	0.95	(0.75, 1.00)	0.87	(0.81, 0.91)	87.3
Mixed	0.00		0.00		1.00		-		1.00		92.1
Pregnancy type	0.90	(0.88, 0.93)									94.9
Not pregnant	0.91	(0.88, 0.94)	0.97	(0.95, 0.98)	0.94	(0.91, 0.97)	0.96	(0.94, 0.98)	0.95	(0.92, 0.97)	95.8
Biochemical	0.84	(0.76, 0.92)	0.79	(0.65, 0.89)	1.00	(0.99, 1.00)	0.91	(0.79, 0.98)	0.99	(0.98, 0.99)	98.3
Intrauterine	0.97	(0.95, 0.99)	0.97	(0.95, 0.99)	0.99	(0.98, 1.00)	0.99	(0.96, 1.00)	0.99	(0.97, 0.99)	98.6
Other	1.00	(1.00, 1.00)	1.00	(0.16, 1.00)	1.00	(1.00, 1.00)	1.00	(0.16, 1.00)	1.00	(1.00, 1.00)	100
Unknown	0.26	(0.07, 0.46)	0.36	(0.11, 0.61)	0.98	(0.97, 0.99)	0.23	(0.05, 0.40)	0.99	(0.98, 1.00)	97.0
# fetal sac	0.93	(0.88, 0.98)									94.8
1	0.87	(0.80, 0.93)	0.96	(0.92, 0.98)	0.93	(0.83, 0.98)	0.98	(0.95, 0.99)	0.87	(0.77, 0.94)	95.2
2	0.93	(0.88, 0.98)	0.95	(0.87, 0.99)	0.98	(0.96, 1.00)	0.94	(0.88, 0.98)	0.99	(0.96, 1.00)	97.6
3	0.67	(0.05, 1.00)	0.50	(0.01, 0.99)	1.00	(0.99, 1.00)	1.00	(0.03, 1.00)	1.00	(0.99, 1.00)	99.7
# fetal heart	0.90	(0.85, 0.96)									93.4
0	0.85	(0.75, 0.96)	0.82	(0.63, 0.94)	0.99	(0.97, 1.00)	0.92	(0.74, 0.99)	0.98	(0.96, 0.99)	97.6
1	0.87	(0.80, 0.93)	0.95	(0.91, 0.97)	0.94	(0.87, 0.98)	0.98	(0.94, 0.99)	0.88	(0.79, 0.94)	94.5
2	0.91	(0.85, 0.97)	0.96	(0.87, 1.00)	0.97	(0.95, 0.99)	0.89	(0.78, 0.96)	0.99	(0.97, 1.00)	93.4
3	1.00	(1.00, 1.00)	1.00	(0.03, 1.00)	1.00	(0.99, 1.00)	1.00	(0.03, 1.00)	1.00	(0.99, 1.00)	100
Chorionicity	0.29	(0.08, 0.51)									73.9
1	0.55	(0.09, 1.00)	0.50	(0.07, 0.93)	0.98	(0.91, 1.00)	0.67	(0.09, 0.99)	0.97	(0.89, 1.00)	95.4

2	0.34	(0.11, 0.57)	0.98	(0.88, 1.00)	0.30	(0.12, 0.54)	0.76	(0.63, 0.86)	0.86	(0.42, 1.00)	76.9
3	1.00	(1.00, 1.00)	1.00	(0.03, 1.00)	1.00	(0.94, 1.00)	1.00	(0.03, 1.00)	1.00	(0.94, 1.00)	100

CI: Confidence interval, eDET: Elective double embryo transfer, eSET: Elective single embryo transfer, FET: Frozen embryo transfer, IVF: *in vitro* fertilization, κ: Kappa coefficient, NPV: Negative predictive value, PPV: Positive predictive value, SN: Sensitivity, SP: Specificity

* weighted Kappa coefficient

Table 4. Sensitivity analysis: Percent agreement of problematic variables by clinic

Clinic	Data entry method	Advanced female age	Diminished ovarian reserve	FSH	AMH	AFC	Reason cancelled	Oocyte origin	eSET/ eDET	Cryopreservation technique-FET
1	Manual	95.2%	95.2%	67.8%	89.0%	71.2%	100.0%	99.3%	92.7%	96.8%
2	EMR	89.0%	91.8%	22.6%	50.0%	49.3%	95.9%	74.0%	84.3%	50.0%
3	Manual	64.4%	87.0%	72.6%	89.7%	74.7%	93.8%	91.0%	84.2%	53.9%
4	EMR	95.2%	95.9%	81.5%	93.8%	64.4%	99.3%	78.4%	59.0%	53.2%
5	EMR	87.0%	86.3%	60.3%	87.0%	63.0%	99.3%	83.9%	95.8%	66.7%
6	Manual	98.0%	96.6%	81.5%	89.7%	50.0%	100.0%	92.9%	82.1%	76.9%

AFC: Antral follicle count, AMH: Antimüllerian hormone, eSET/ eDET: elective single embryo transfer or elective double embryo transfer, FET-Frozen embryo transfer

Table 5. Prevalence of elective single or double embryo transfer by data source

	CARTR Plus		**Reabstracted Data**	
eSET or eDET*	N	%	N	%
SET	n=362		n=362	
Elective	249	68.8	274	75.7
Non-elective	113	31.2	86	23.8
Missing	0	0.00	2	0.55
eDET	n=273		n=273	
Elective	138	50.6	201	73.6
Non-elective	135	49.5	72	26.4

eDET: Elective double embryo transfer, eSET: Elective single embryo transfer

Table 6. Sensitivity analysis: Percent agreement of elective single embryo transfer and double embryo transfer by clinic, cycle type, and day of transfer

	eSET	eDET
Clinic		
1	94.9%	97.8%
2	89.8%	94.5%
3	97.7%	86.5%
4	88.1%	70.9%
5	99.3%	96.5%
6	95.0%	87.1%
Cycle Type		
Fresh	97.3%	88.6%
Frozen	86.9%	89.3%
Transfer Day		
Day 3	97.9%	79.4%
Day 5	93.6%	88.0%

eDET: Elective double embryo transfer, eSET: Elective single embryo transfer

Appendix B: Supplemental information for Manuscript 2

Appendix 1- Prevalence estimates, means or medians of selected variables from cycle years 2013-2015 from CARTR Plus

Variable	Prevalence (%)	Mean (SD)/ Median(IQR)
Intake of patient		
Patient age (years)		35.6 (SD: 4.82)
Patient DOB		
Oocyte provider age (years)		34.7(SD: 4.88)
Reason for treatment cycle (picklist: 20 items)		
Advanced female age	13.6%	
Diminished ovarian reserve	16.7%	
FSH day 2-4 of oocyte provider (IU/L)		7.49 (SD: 4.52)
AFC of oocyte provider (# of follicles)		14.7 (SD: 10.64)
AMH level of oocyte provider (ng/mL)		5.28 (SD: 9.04)
Stimulation		
Type of cycle (picklist: 5 items)		
FET	34.7%	
Frozen oocyte IVF	0.83%	
IVF	63.4%	
IVM	0.27%	
Oocyte banking	0.80%	
Current cycle start date		
Oocyte collection date		
Cycle cancelled	7.17%	
Reason for cancelled (picklist: 12 items)		
Low ovarian response	73.3%	
Premature ovulation	7.10%	
Other	10.9%	
Retrieval		
Oocyte origin (picklist: 9 items)		
Fresh donor	2.86%	

Fresh own	74.9%	
Embryology		
Embryo cryopreservation method (picklist: 3 items)		
Vitrification	79.0%	
Slow-freeze	7.33%	
Mixed	0.07%	
# embryos thawed		1.00 (IQR 1.00-2.00)
# of utilizable embryos after thaw		1.00 (IQR 1.00-2.00)
Embryo transfer		
Embryo transfer (Y)	75.6%	
ET day		3.00 (IQR 0.00-5.00)
# embryos transferred		1.00 (IQR 1.00-2.00)
eSET or eDET	47.9%	
Pregnancy		
Type of pregnancy		
Biochemical	8.58%	
Clinical intrauterine	37.7%	
Not pregnant	52.9%	
# US fetal sacs		1.00 (IQR 1.00-2.00)
# US fetal hearts		1.00 (IQR 1.00-2.00)
Chorionicity		2.00 (IQR 2.00-2.00)

AFC: antral follicle count, AMH: anti-mullerian hormone, DOB: date of birth, eDET: elective double embryo transfer, eSET: elective single embryo transfer, ET: embryo transfer, FET: frozen embryo transfer, FSH: follicle stimulating hormone, IQR: interquartile range, IVF: in vitro fertilization, IVM: in vitro maturation, US: ultrasound, SD: standard deviation, Y: yes

Appendix 2-Measures of agreement

		Re-abstracted Chart		
		Disease present	Disease absent	Total
Database	Disease present	A	B	A+B
	Disease absent	C	D	C+D
	Total	A+C	B+D	A+B+C+D

Sensitivity

- The percentage of patients who are coded as having disease among all those with disease
- A/(A+C)

Specificity

- The percentage of patients who are coded as being disease-free among all those who do not have disease
- D/(B+D)

Positive predictive value

- The percentage of patients who have disease among all those who are coded with having disease
- A/(A+B)

Negative predictive value

- The percentage of patients who do not have disease among all those who are coded as being disease-free
- D/(C+D)

Kappa coefficient

- The percent agreement between two measures (the code and the re-abstracted variable) while adjusting for the agreement due to chance

- Observed agreement (OA):
    - (A+D)/(A+B+C+D)
- Agreement due to chance (AC):
    - [(A+B)/(A+B+C+D) * (A+C)/(A+B+C+D)] + [(B+C)/ (A+B+C+D) * (C+D)/ (A+B+C+D)]
- Kappa:
    - (OA-AC)/(1-AC)

Intraclass correlation coefficient (ICC)

- The ratio of the variance between all the sets of pairs (coded variable and re-abstracted variable) to the total variation in the measurements

Appendix 3-Level of agreement for kappa and intraclass correlation coefficients

Measured Kappa or ICC	**Degree of agreement beyond chance**
0.00-0.20	Poor
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Strong
0.81-1.00	Almost perfect

ICC: intraclass correlation coefficient

*Adapted from Landis and Koch (16)

Appendix 4- Sample size calculation sensitivity analysis

A. Sample size calculations for categorical variables, example of 3 data elements

Variable	Anticipated K	K lower limit	K upper limit	N
Premature ovulation (p=0.071), binary	0.70	0.60	0.80	890
	0.75	0.70	0.80	2903
	0.80	0.70	0.90	726
	0.85	0.75	0.95	628
Embryo transfer (p=0.756), binary	0.70	0.60	0.80	328
	0.75	0.70	0.80	1054
	0.80	0.70	0.90	264
	0.80	0.75	0.85	726
	0.85	0.75	0.95	227
Pregnancy outcome (p=0.087, 0.380, 0.533), 3 categories	0.70	0.60	0.80	207
	0.75	0.70	0.80	670
	0.80	0.70	0.90	168
	0.80	0.76	0.84	726
	0.85	0.75	0.95	145

K: kappa coefficient, N: sample size, p: prevalence

B. Sample size calculations for continuous variables

Anticipated ICC	ICC lower limit	ICC upper limit	N
0.70	0.55	0.85	128
0.70	0.60	0.80	260
0.70	0.65	0.75	929
0.70	0.67	0.73	726
0.75	0.60	0.90	64
0.75	0.65	0.85	202
0.75	0.70	0.80	705
0.75	0.73	0.77	726
0.90	0.80	1.00	58
0.90	0.85	0.95	170
0.90	0.89	0.91	726

ICC: intraclass correlation coefficient, N: sample size

Appendix 5. Description of study variables by data source for clinic 1

	CARTR Plus						Reabstracted Data					
Variable	**N**	**%**	**Mean**	**SD**	**Median**	**IQR**	**N**	**%**	**Mean**	**SD**	**Median**	**IQR**
Intake of patient												
Patient age (years)	146		36.6	4.70	37	(33-40)	146		36.6	4.70	37	(33-40)
Oocyte provider age (years)	146		36.2	7.30	37	(32-39)	146		34.7	7.06	37	(32-39)
Reason for treatment cycle												
Diminished ovarian reserve												
Yes	57	39.0					54	37.0				
No	89	61.0					92	63.0				
Advanced female age												
Yes	29	19.9					24	16.4				
No	117	80.1					122	83.6				
FSH (IU/L)	108		7.99	4.17	6.85	(5.55-9.15)	77		9.01	6.23	7.40	(6.00-10.6)
AFC (# follicles)	118		12.3	11.5	10	(6-16)	90		11.80	9.13	9	(6-17)
AMH (ng/dL)*	20		1.80	2.60	0.80	(0.45-1.65)	6		2.95	3.61	0.85	(0.80-6.30)
Stimulation												
Cycle type												
IVF	113	77.4					115	78.8				
FET	31	21.2					31	21.2				
Frozen oocyte IVF	2	1.37					0	0.00				
Oocyte banking	0	0.00					0	0.00				
Cancelled cycle												
Yes	9	6.16					9	6.16				
No	137	93.8					137	93.8				
Reason for cancelled												
Low ovarian response	5	55.6					5	55.6				
Premature ovulation	2	22.2					2	22.2				
Other	2	22.2					2	22.2				
Missing	0	0.00					0	0.00				

Retrieval												
Oocyte origin												
Fresh own oocytes	127	92.7					127	92.7				
Fresh donor oocytes	8	5.84					9	6.57				
Other	2	1.46					1	0.73				
Missing	0	0.00					0	0.00				
Embryo transfer												
Embryo transfer												
Yes	111	81.0					111	81.0				
No	26	19.0					26	19.0				
Missing	0	0.00					0	0.00				
ET day	111		3.50	1.09	3	(3-5)	111		2.45	1.68	3	(0-3)
Fresh cycles	83		3.27	1.02	3	(3-3)	83		3.25	1.00	3	(3-3)
2	15	18.1					15	18.1				
3	49	59.0					50	60.2				
5	19	22.9					18	21.7				
6	0	0.00					0	0.00				
Frozen cycles							28		4.25	1.04	5	(3-5)
2							0	0.00				
3							11	39.3				
4							0	0.00				
5							16	57.1				
6							1	3.57				
>6							0	0.00				
Missing							0	0.00				
# Embryos transferred												
1	79	71.2					79	71.2	79	71.2		
2	30	27.0					30	27.0	30	27.0		
3	2	1.80					2	1.80	2	1.80		
4	0	0.00					0	0.00	0	0.00		
eSET or eDET												
Yes	82	73.9					80	72.1	82	73.9		
No	29	26.1					31	27.9	29	26.1		
Missing	79	71.2					79	71.2	79	71.2		

Embryology												
Embryo cryopreservation												
IVF												
Vitrification	32	100					30	100				
Slow-freeze	0	0.00					0	0.00				
Mixed	0	0.00					0	0.00				
FET												
Vitrification	30	96.8					31	100				
Slow-freeze	1	3.23					0	0.00				
Mixed	0	0.00					0	0.00				
Missing	0	0.00					0	0.00				
# embryos thawed	31		3.13	2.06	2	(2-4)	31		3.13	2.06	2	(2-4)
# embryos utilizable after thaw	31		2.29	1.83	2	(1-3)	31		2.42	2.08	2	(1-3)
Pregnancy												
Pregnancy type												
Not pregnant	89	61.0					91	62.3				
Biochemical	15	10.3					14	9.59				
Clinical intrauterine	40	27.4					41	28.1				
Other	0	0.00					0	0.00				
Unknown	2	1.37					0	0.00				
# fetal sac												
1	35	87.5					37	90.2				
2	4	10.0					4	9.76				
3	0	0.00					0	0.00				
Missing	0	0.00										
# fetal heart												
0	5	12.5					5	12.2				
1	32	80.0					33	80.5				
2	3	7.50					3	7.32				
3	0	0.00					0	0.00				
Missing	0	0.00										

Chorionicity				
1	0	0.00	0	0.00
2	3	75.0	3	75.0
3	0	0.00	0	0.00
Missing	1	25.0	1	25.0

AFC: Antral follicle count, AMH: Antimüllerian hormone, CARTR Plus: Canadian Assisted Reproductive Technologies Register Plus, eDET: Elective double embryo transfer, eSET: Elective single embryo transfer, ET: Embryo transfer, FSH: Follicle stimulating hormone; FET: Frozen embryo transfer, IQR: Interquartile range, IVF: *in vitro* fertilization, N: Number or patients, SD: Standard deviation

Appendix 6. Description of study variables by data source for clinic 2

	CARTR Plus						Reabstracted Data					
Variable	**N**	**%**	**Mean**	**SD**	**Median**	**IQR**	**N**	**%**	**Mean**	**SD**	**Median**	**IQR**
Intake of patient												
Patient age (years)	146		36.2	4.41	36.00	(24-39)	146		36.1	4.41	36	(34-39)
Oocyte provider age (years)	145		36.6	11.40	35.00	(33-38)	131		34.8	4.13	35	(33-37)
Reason for treatment cycle												
Diminished ovarian reserve												
Yes	19		13.0				11		7.53			
No	127		87.0				135		92.5			
Advanced female age												
Yes	24		16.4				26		17.80			
No	122		83.6				120		82.20			
FSH (IU/L)	0						113		5.77	2.72	5.40	(4.30-6.80)
AFC (# follicles)	0						74		18.1	9.94	16.0	(11-25)
AMH (ng/dL)*	0						73		2.71	2.95	2.10	(0.80-3.70)
Stimulation												
Cycle type												
IVF	84		57.5				86		58.90			
FET	50		34.3				50		34.30			
Frozen oocyte IVF	5		3.42				3		2.05			
Oocyte banking	7		4.79				7		4.79			
Cancelled cycle												
Yes	19		13.0				18		12.3			
No	127		87.0				128		87.7			
Reason for cancelled												
Low ovarian response	11		57.9				16		88.9			
Premature ovulation	1		5.26				0		0.00			
Other	4		21.1				2		11.1			
Missing	3		15.80									

Retrieval												
Oocyte origin												
Fresh own oocytes	113		89.0				89		69.5			
Fresh donor oocytes	4		3.15				4		3.13			
Other	8		6.30				1		0.78			
Missing	2		1.57				34		25.6			
Embryo transfer												
Embryo transfer												
Yes	79		62.2				79		61.7			
No	48		37.8				43		33.6			
Missing							6		4.69			
ET day	72		5.44	0.71	6	(5-6)	79		1.81	2.40	0	(0-5)
Fresh cycles	29		4.93	0.37	5	(5-5)	29		4.93	0.37	5	(5-5)
2	0	0.00					0	0.00				
3	1	3.45					1	3.45				
5	26	96.6					28	96.6				
6	0	0.00					0	0.00				
Frozen cycles							50		5.79	0.67	6	(6-6)
2							0	0.00				
3							2	4.00				
4							0	0.00				
5							3	6.00				
6							38	76.0				
>6							0	0.00				
Missing							7	14.0				
# Embryos transferred												
1	52		65.8				52		65.8			
2	27		34.2				27		34.2			
3	0		0.00				0		0.00			
4	0		0.00				0		0.00			
eSET or eDET												
Yes	56		70.9				56		70.9			
No	21		26.6				23		29.1			
Missing	2		2.53				0		0.00			

Embryology										
Embryo cryopreservation										
IVF										
Vitrification	55	100.0				58	100			
Slow-freeze	0	0.00				0	0.00			
Mixed	0	0.00				0	0.00			
FET										
Vitrification	50	100.0				25	50.0			
Slow-freeze	0	0.00				0	0.00			
Mixed	0	0.00				0	0.00			
Missing	0	0.0				25	50.0			
# embryos thawed	49	1.33	0.52	1	(1-2)	50	1.34	0.52	1	(1-2)
# embryos utilizable after thaw	49	1.27	0.45	1	(1-2)	50	1.28	0.45	1	(1-2)
Pregnancy										
Pregnancy type										
Not pregnant	110	75.3				102	69.6			
Biochemical	4	2.74				2	1.37			
Clinical intrauterine	27	18.5				23	15.8			
Other	0	0.00				0	0.00			
Unknown	5	3.42				19	13.0			
# fetal sac										
1	19	70.4				16	69.6			
2	7	25.9				7	30.4			
3	0	0.00				0	0.00			
Missing	1	3.70								
# fetal heart										
0	2	7.41				1	4.35			
1	18	66.7				17	73.9			
2	6	22.2				5	21.7			
3	0	0.00				0	0.00			
Missing	1	3.70								

Chorionicity				
1	0	0.00	0	0.00
2	7	100	7	100
3	0	0.00	0	0.00
Missing	0	0.00	0	0.00

AFC: Antral follicle count, AMH: Antimüllerian hormone, CARTR Plus: Canadian Assisted Reproductive Technologies Register Plus, eDET: Elective double embryo transfer, eSET: Elective single embryo transfer, ET: Embryo transfer, FSH: Follicle stimulating hormone; FET: Frozen embryo transfer, IQR: Interquartile range, IVF: *in vitro* fertilization, N: Number or patients, SD: Standard deviation

Appendix 7. Description of study variables by data source for clinic 3

	CARTR Plus						Reabstracted Data					
Variable	**N**	**%**	**Mean**	**SD**	**Median**	**IQR**	**N**	**%**	**Mean**	**SD**	**Median**	**IQR**
Intake of patient												
Patient age (years)	146		34.4	3.97	34.0	(31-38)	146		34.4	3.90	34	(31-37)
Oocyte provider age (years)	146		33.9	4.16	34.0	(31-37)	146		33.8	4.13	34	(31-37)
Reason for treatment cycle												
Diminished ovarian reserve												
Yes	12		8.22				29		19.9			
No	134		81.8				117		80.1			
Advanced female age												
Yes	10		6.85				62		42.5			
No	136		93.2				84		57.5			
FSH (IU/L)	143		7.29	2.62	6.90	(5.70-8.40)	128		7.33	3.07	7.10	(5.55-8.60)
AFC (# follicles)	123		15.8	9.30	14	(10-21)	126		15.40	8.76	14	(9-20)
AMH (ng/dL)*	48		2.07	2.01	1.20	(0.75-2.50)	44		1.89	1.98	1.05	(0.70-2.10)
Stimulation												
Cycle type												
IVF	120		82.2				120		82.20			
FET	26		17.8				26		17.80			
Frozen oocyte IVF	0		0.00				0		0.00			
Oocyte banking	0		0.00				0		0.00			
Cancelled cycle												
Yes	13		8.90				13		8.90			
No	133		91.1				133		91.1			
Reason for cancelled												
Low ovarian response	4		30.8				7		53.9			
Premature ovulation	0		0.00				1		7.69			
Other	1		7.69				5		38.5			
Missing	8		61.5									

Oocyte origin												
Fresh own oocytes	132		99.3				120		90.2			
Fresh donor oocytes	1		0.75				1		0.75			
Other	0		0.00				0		0.00			
Missing	0		0.00				12		9.02			
Embryo transfer												
Embryo transfer												
Yes	116		87.2				116		87.2			
No	17		12.8				17		12.8			
Missing	0		0.00				0		0.00			
ET day	116		3.95	1.24	3	(3-5)	116		2.87	1.65	3	(3-3)
Fresh cycles	93		3.56	0.90	3	(3-5)	93		3.58	0.91	3	(3-5)
2	0	0.00					0	0.00				
3	66	71.0					67	72.0				
5	27	29.0					26	28.0				
6	0	0.00					0	0.00				
Frozen cycles							23		5.52	1.20	6	(4-6)
2							0	0.00				
3							0	0.00				
4							6	26.1				
5							5	21.7				
6							7	30.4				
>6							5	21.7				
Missing							0	0.00				
# Embryos transferred												
1	52		65.8				52		65.8			
2	27		34.2				27		34.2			
3	0		0.00				0		0.00			
4	0		0.00				0		0.00			
eSET or eDET												
Yes	56		70.9				56		70.9			
No	21		26.6				23		29.1			
Missing	2		2.53									

Embryology										
Embryo cryopreservation										
IVF										
Vitrification	21	70.0				19	65.5			
Slow-freeze	9	30.0				10	34.5			
Mixed	0	0.00				0	0.00			
FET										
Vitrification	0	0.00				0	0.00			
Slow-freeze	26	100				14	53.9			
Mixed	0	0.00				0	0.00			
Missing	0	0.00				12	46.2			
# embryos thawed	26	1.42	0.64	1	(1-2)	26	1.35	0.63	1	(1-2)
# embryos utilizable after thaw	26	1.00	0.49	1	(1-1)	26	1.00	0.49	1	(1-1)
Pregnancy										
Pregnancy type										
Not pregnant	86	58.9				94	64.4			
Biochemical	11	7.53				11	7.53			
Clinical intrauterine	40	27.4				39	26.7			
Other	2	1.37				2	1.37			
Unknown	7	4.79				0	0.00			
# fetal sac										
1	32	80.0				31	79.5			
2	7	17.5				8	20.5			
3	1	2.50				0	0.00			
Missing	0	0.00								
# fetal heart										
0	5	12.5				5	12.8			
1	27	67.5				26	66.7			
2	7	17.5				7	18.0			
3	1	2.50				1	2.56			
Missing	21	70.0				19	65.5			

Chorionicity				
1	0	0.00	0	0.00
2	3	37.5	8	100
3	0	0.00	0	0.00
Missing	5	62.5	0	0.00

AFC: Antral follicle count, AMH: Antimüllerian hormone, CARTR Plus: Canadian Assisted Reproductive Technologies Register Plus, eDET: Elective double embryo transfer, eSET: Elective single embryo transfer, ET: Embryo transfer, FSH: Follicle stimulating hormone; FET: Frozen embryo transfer, IQR: Interquartile range, IVF: *in vitro* fertilization, N: Number or patients, SD: Standard deviation

Appendix 8. Description of study variables by data source for clinic 4

	CARTR Plus						Reabstracted Data					
Variable	**N**	**%**	**Mean**	**SD**	**Median**	**IQR**	**N**	**%**	**Mean**	**SD**	**Median**	**IQR**
Intake of patient												
Patient age (years)	146		35.6	4.72	35.0	(32-39)	146		35.6	4.71	35	(32-39)
Oocyte provider age (years)	146		34.2	7.01	33.0	(30-37)	146		33.6	4.60	33	(30-37)
Reason for treatment cycle												
Diminished ovarian reserve												
Yes	18		12.3				16		11.0			
No	128		87.7				130		89.0			
Advanced female age												
Yes	34		23.3				33		22.60			
No	112		76.7				113		77.40			
FSH (IU/L)	133		6.72	2.32	6.90	(5.00-8.00)	116		6.85	2.5	7.00	(5.10-8.00)
AFC (# follicles)	125		23.0	13.80	20	(13-30)	117		22.30	14.5	19	(12-30)
AMH (ng/dL)*	10		2.41	2.11	2.05	(1.20-3.50)	1		1.20			
Stimulation												
Cycle type												
IVF	89		61.0				90		61.6			
FET	47		32.2				47		32.2			
Frozen oocyte IVF	9		6.16				9		6.16			
Oocyte banking	1		0.68				0		0.00			
Cancelled cycle												
Yes	12		8.22				12		8.22			
No	134		91.8				134		91.80			
Reason for cancelled												
Low ovarian response	11		91.7				12		100			
Premature ovulation	0		0.00				0		0.00			
Other	1		8.33				0		0.00			
Missing	0		0.00									

Oocyte origin												
Fresh own oocytes	121		90.3				105		78.4			
Fresh donor oocytes	3		2.24				0		0.00			
Other	10		7.46				0		0.00			
Missing	0		0.00				29		21.6			
Embryo transfer												
Embryo transfer												
Yes	123		91.8				123		91.8			
No	11		8.2				11		8.21			
Missing	0		0.00				0		0.00			
ET day	123		4.30	1.23	5	(3-5)	123		2.45	2.13	3	(0-5)
Fresh cycles	76		3.92	1.15	5	(3-5)	76		3.96	1.14	5	(3-5)
2	7	9.21					8	10.5				
3	29	38.2					29	38.2				
5	40	52.6					39	51.3				
6	0	0.00					0	0.00				
Frozen cycles							47		5.15	0.96	5	(5-6)
2							0	0.00				
3							6	26.1				
4							0	0.00				
5							22	46.8				
6							19	40.4				
>6							0	0.00				
Missing												
# Embryos transferred												
1	57		46.3				57		46.3			
2	53		43.10				53		43.1			
3	11		8.94				11		8.94			
4	2		1.63				2		1.63			
eSET or eDET												
Yes	80		65.0				25		20.3			
No	43		35.0				98		79.7			
Missing	57		46.3				57		46.3			

Embryology										
Embryo cryopreservation										
IVF										
Vitrification	40	83.3				41	85.4			
Slow-freeze	7	14.60				7	14.6			
Mixed	1	2.08				0	0.00			
FET										
Vitrification	37	78.7				23	48.9			
Slow-freeze	9	19.2				4	8.51			
Mixed	1	2.13				0	0.00			
Missing	0	0.00				20	42.6			
# embryos thawed	47	2.68	3.01	2	(1-3)	47	2.68	3.01	2	(1-3)
# embryos utilizable after thaw	47	1.55	0.75	1	(1-2)	70	1.04	0.95	1	(0-2)
Pregnancy										
Pregnancy type										
Not pregnant	72	49.3				71	48.6			
Biochemical	10	6.85				11	7.53			
Clinical intrauterine	64	43.8				64	43.8			
Other	0	0.00				0	0.00			
Unknown	0	0.00				0	0.00			
# fetal sac										
1	45	70.3				47	73.4			
2	17	26.6				16	25.0			
3	1	1.56				1	1.56			
Missing	1	1.56								
# fetal heart										
0	4	6.25				5	7.81			
1	44	68.8				43	67.2			
2	15	23.4				16	25.0			
3	0	0.00				0	0.00			
Missing	1	1.56								

Chorionicity				
1	0	0.00	1	5.56
2	15	83.3	16	88.9
3	1	5.56	1	5.56
Missing	2	11.10	0	0.00

AFC: Antral follicle count, AMH: Antimüllerian hormone, CARTR Plus: Canadian Assisted Reproductive Technologies Register Plus, eDET: Elective double embryo transfer, eSET: Elective single embryo transfer, ET: Embryo transfer, FSH: Follicle stimulating hormone; FET: Frozen embryo transfer, IQR: Interquartile range, IVF: *in vitro* fertilization, N: Number or patients, SD: Standard deviation

Appendix 9. Description of study variables by data source for clinic 5

	CARTR Plus						Reabstracted Data					
Variable	**N**	**%**	**Mean**	**SD**	**Median**	**IQR**	**N**	**%**	**Mean**	**SD**	**Median**	**IQR**
Intake of patient												
Patient age (years)	146		35.8	5.50	35.0	(33-39)	146		35.7	5.45	35	(33-39)
Oocyte provider age (years)	145		33.2	4.92	34.0	(29-37)	145		33.2	4.96	34	(29-37)
Reason for treatment cycle												
Diminished ovarian reserve												
Yes	33		22.6				33		22.6			
No	113		77.4				113		77.4			
Advanced female age												
Yes	15		10.3				32		21.9			
No	131		89.7				114		78.1			
FSH (IU/L)	91		7.03	3.47	6.30	(4.80-8.10)	53		7.12	2.63	6.50	(5.60-8.30)
AFC (# follicles)	69		12.0	5.70	11	(8-15)	18		11.8	5.54	12	(8-14)
AMH (ng/dL)*	36		2.60	2.33	1.90	(1.00-2.90)	26		2.60	2.02	1.85	(1.00-3.40)
Stimulation												
Cycle type												
IVF	89		61.0				89		61.0			
FET	52		35.6				52		35.6			
Frozen oocyte IVF	2		1.37				2		1.37			
Oocyte banking	3		2.05				3		2.05			
Cancelled cycle												
Yes	3		2.05				2		1.37			
No	143		98.0				144		98.6			
Reason for cancelled												
Low ovarian response	2		66.7				2		100			
Premature ovulation	0		0.00				0		0.00			
Other	1		33.3				0		0.00			
Missing	0		0.00									

Oocyte origin												
Fresh own oocytes	122		85.3				106		73.6	122		
Fresh donor oocytes	18		12.6				16		11.1	18		
Other	3		2.10				0		0.00	3		
Missing	0		0.00				22		15.3	0		
Embryo transfer												
Embryo transfer												
Yes	93		65.0				94		65.3			
No	50		35.0				47		32.6			
Missing	0		0.00				3		2.08			
ET day	93		4.39	1.02	5	(3-5)	93		1.76	2.07	0	(0-3)
Fresh cycles	43		3.84	1.13	3	(3-5)	43		3.84	1.13	3	(3-5)
2	4	9.30					4	9.30				
3	19	44.2					19	44.2				
5	20	46.5					20	46.5				
6	0	0.00					0	0.00				
Frozen cycles							50		4.86	0.61	5	(5-5)
2							0	0.00				
3							2	4.00				
4							7	14.0				
5							37	74.0				
6							4	8.00				
>6							0	0.00				
Missing							0	0.00				
# Embryos transferred												
1	67		72.0				67		71.3			
2	24		25.8				25		26.6			
3	1		1.08				1		1.06			
4	1		1.08				1		1.06			
eSET or eDET												
Yes	76		11.6				81		86.2			
No	17		18.3				13		13.8			
Missing	0		0.00									

Embryology										
Embryo cryopreservation										
IVF										
Vitrification	69	100				69	100			
Slow-freeze	0	0.00				0	0.00			
Mixed	0	0.00				0	0.00			
FET										
Vitrification	49	94.2				35	67.3			
Slow-freeze	2	3.85				0	0.00			
Mixed	0	0.00				0	0.00			
Missing	1	1.92				17	32.7			
# embryos thawed	52	1.38	0.93	1	(1-1.5)	52	1.42	0.91	1	(1-2)
# embryos utilizable after thaw	52	1.27	0.66	1	(1-1.5)	52	1.29	0.61	1	(1-2)
Pregnancy										
Pregnancy type										
Not pregnant	86	58.9				84	57.5			
Biochemical	6	4.11				5	3.42			
Clinical intrauterine	54	37.0				54	37.0			
Other	0	0.00				0	0.00			
Unknown	0	0.00				3	2.05			
# fetal sac										
1	49	90.7				47	87.0			
2	5	9.26				7	13.0			
3	0	0.00				0	0.00			
Missing	0	0.00								
# fetal heart										
0	6	11.1				4	7.41			
1	45	83.3				44	81.5			
2	3	5.56				6	11.1			
3	0	0.00				0	0.00			
Missing										

Chorionicity				
1	0	0.00	1	14.3
2	4	80.0	5	71.4
3	0	0.00	0	0.00
Missing	1	20.0	1	14.3

AFC: Antral follicle count, AMH: Antimüllerian hormone, CARTR Plus: Canadian Assisted Reproductive Technologies Register Plus, eDET: Elective double embryo transfer, eSET: Elective single embryo transfer, ET: Embryo transfer, FSH: Follicle stimulating hormone; FET: Frozen embryo transfer, IQR: Interquartile range, IVF: *in vitro* fertilization, N: Number or patients, SD: Standard deviation

Appendix 10. Description of study variables by data source for clinic 6

	CARTR Plus						Reabstracted Data					
Variable	**N**	**%**	**Mean**	**SD**	**Median**	**IQR**	**N**	**%**	**Mean**	**SD**	**Median**	**IQR**
Intake of patient												
Patient age (years)	146		34.5	4.10	35.00	(32-37)	146		34.5	4.12	35	(32-37)
Oocyte provider age (years)	146		3.9	4.14	34.00	(31-36)	146		33.8	4.15	34	(31-36)
Reason for treatment cycle												
Diminished ovarian reserve												
Yes	11	7.53					6	4.11				
No	135	92.5					140	95.9				
Advanced female age												
Yes	13		8.9				10		6.85			
No	133		91.1				136		93.20			
FSH (IU/L)	146		6.22	2.01	5.90	(5.00-7.00)	135		6.38	2.7	5.90	(4.90-7.10)
AFC (# follicles)	74		14.2	8.08	14	(8-19)	11		4.54	5.73	2	(2-2)
AMH (ng/dL)*	15		3.07	2.96	2.90	(0.70-4.10)	0					
Stimulation												
Cycle type												
IVF	89		61.0				89		61.0			
FET	52		35.6				52		35.6			
Frozen oocyte IVF	2		1.37				2		1.37			
Oocyte banking	3		2.05				3		2.05			
Cancelled cycle												
Yes	3		2.05				2		1.37			
No	143		98.0				144		98.6			
Reason for cancelled												
Low ovarian response	5		83.3				5		83.3			
Premature ovulation	0		0.00				0		0.00			
Other	1		16.7				1		16.7			
Missing	0		0.00									

Oocyte origin												
Fresh own oocytes	135		96.4				125		89.3			
Fresh donor oocytes	5		3.57				5		3.57			
Other	0		0.00				1		0.71			
Missing	0		0.00				9		6.43			
Embryo transfer												
Embryo transfer												
Yes	132		94.3				132		94.3			
No	8		5.71				8		5.71			
Missing	0		0.00				0		0.00			
ET day	132		4.95	0.37	5	(5-5)	132		3.54	2.27	5	(0-5)
Fresh cycles	94		4.97	0.31	5	(5-5)	94		4.97	0.31	5	(5-5)
2	0	0.00					0	0.00				
3	2	2.13					2	2.13				
5	91	96.8					91	96.8				
6	1	1.06					1	1.06				
Frozen cycles							38		4.92	0.49	5	(5-5)
2							1	2.63				
3							0	0.00				
4							0	0.00				
5							37	97.4				
6							0	0.00				
>6							0	0.00				
Missing												
# Embryos transferred												
1	61		46.2				61		46.2			
2	70		53.0				70		53.0			
3	1		0.76				1		0.76			
4	0		0.00				0		0.00			
eSET or eDET												
Yes	88		66.7				69		52.3			
No	44		33.3				63		47.7			
Missing	0		0.00									

Embryology										
Embryo cryopreservation										
IVF										
Vitrification	50	100				50	100			
Slow-freeze	0	0.00				0	0.00			
Mixed	0	0.00				0	0.00			
FET										
Vitrification	28	71.8				28	71.8			
Slow-freeze	11	28.2				2	5.13			
Mixed	0	0.00				0	0.00			
Missing	0	0.0				9	23.1			
# embryos thawed	39	2.03	0.96	2	(1-2)	40	2.00	0.96	2	(1-2)
# embryos utilizable after thaw	39	1.49	0.56	2	(1-2)	40	1.38	0.59	1	(1-2)
Pregnancy										
Pregnancy type										
Not pregnant	76	52.1				80	54.8			
Biochemical	6	4.11				2	1.37			
Clinical intrauterine	64	43.8				64	43.8			
Other	0	0.00				0	0.00			
Unknown	0	0.00				0	0.00			
# fetal sac										
1	42	65.6				43	67.2			
2	22	34.4				21	32.8			
3	0	0.00				0	0.00			
Missing	0	0.00								
# fetal heart										
0	6	9.38				6	9.38			
1	40	62.5				39	60.9			
2	18	28.1				19	26.7			
3	0	0.00				0	0.00			
Missing	0	0.00								

Chorionicity				
1	4	17.4	2	8.70
2	13	56.5	20	87.0
3	0	0.00	0	0.00
Missing	6	26.1	1	4.35

AFC: Antral follicle count, AMH: Antimüllerian hormone, CARTR Plus: Canadian Assisted Reproductive Technologies Register Plus, eDET: Elective double embryo transfer, eSET: Elective single embryo transfer, ET: Embryo transfer, FSH: Follicle stimulating hormone; FET: Frozen embryo transfer, IQR: Interquartile range, IVF: *in vitro* fertilization, N: Number or patients, SD: Standard deviation

Appendix 11-Description of missing charts from CARTR Plus

Variable	N	Mean/Percent	SD
Intake of patient			
Patient age (years)	12	35.4	2.94
Oocyte provider age (years)	12	35.1	3.09
Reason for treatment cycle			
Diminished ovarian reserve			
Yes	2	16.7	
No	10	83.3	
Advanced female age			
Yes	8	66.7	
No	4	33.3	
FSH (IU/L)	12	7.88	2.58
AFC (number of follicles)	11	12.0	5.53
AMH (ng/dL)	3	1.46	0.70
Stimulation			
Cycle type			
IVF	7	58.3	
FET	5	41.7	
Cancelled cycle			
Yes	2	16.7	
No	10	83.3	
Reason for cancelled			
Low ovarian response	1	50.0	
Premature ovulation	1	50.0	
Retrieval			
Oocyte origin			
Fresh own	10	100	
Fresh donor	0	0.00	
Other	0	0.00	
Embryo transfer			
Embryo transfer			
Yes	7	70.0	
No	3	30.0	
ET day			
Fresh cycles			
3	3	100	

Frozen cycles		
Missing	4	100
# Embryos transferred		
1	6	85.7
2	1	14.3
eSET or eDET		
Yes	2	28.6
No	5	71.4
Embryology		
Embryo cryopreservation		
IVF		
Vitrification	0	
Slow-freeze	0	
FET		
Vitrification	5	100
# embryos thawed		
1	4	80.0
2	0	0.00
3	1	20.0
# embryos utilizable after thaw		
0	1	
1	4	
Pregnancy		
Pregnancy type		
Not pregnant	8	66.7
Biochemical	2	16.7
Clinical intrauterine	2	16.7
# fetal sac		
1	2	100
Missing	0	0.00
# fetal heart		
0	1	50.0
1	1	50.0
Missing	0	0.00
Chorionicity	0	

AFC: Antral follicle count, AMH: Antimüllerian hormone, CARTR Plus: Canadian Assisted Reproductive Technologies Register Plus, eDET: Elective double embryo transfer, eSET: Elective single embryo transfer, ET: Embryo transfer, FET: Frozen embryo transfer, IVF: *in vitro* fertilization, N: Number or patients, SD: Standard deviation

Appendix 12. Recommendations for changes to CARTR Plus

Variable	Recommendation for change
Intake of patient	
Patient age (years)	None required
Oocyte provider age (years)	None required
Reason for treatment cycle	Urge physician to clearly document in chart on each cycle indication for treatment.
Diminished ovarian reserve	Eventually develop algorithm based on strict criteria from laboratory tests and ultrasound to reduce misclassification upon expert review
Advanced female age	Use strict criteria of women >35 years of age to meet criteria.
FSH (IU/L)	Urge clinics to report tests performed at site to reduce missingness. Eventually develop method of auto-upload from laboratory to database
AFC (number of follicles)	Urge clinics to report tests performed at site to reduce missingness.
AMH (ng/dL)	Urge clinics to report tests performed at site to reduce missingness.
Stimulation	
Cycle type	None required
Cancelled cycle	
Reason for cancelled	Clearly document reason by cancelling physician to reduce the risk of misclassification by data entry person.
Low ovarian response	Develop strict criteria for decision based on AFC and number of oocytes growing with stimulation
Premature ovulation	
Retrieval	
Oocyte origin	Change to mandatory variable when cycle is not cancelled When missing, create imputation strategy to represent majority fresh own then fresh donor
Embryo transfer	
Embryo transfer	None required
ET day	
Fresh cycles	None required
Frozen cycles	Encourage clinics to enter the day embryo frozen and the number of days since thaw prior to transfer (if grown from day 3 embryo to day 6)
# Embryos transferred	None
eSET or eDET	Eliminate eDET as no longer clinically relevant

	Ensure all clinics are using consistent definition of eSET (transfer of embryo when viable embryos are available regardless of ability to cryopreserve, stage or grade). Create a subsidiary variable called number of embryos available at time of transfer to avoid misclassification
Embryology	
Embryo cryopreservation	None required
# embryos thawed	None required
# embryos utilizable after thaw	None required
Pregnancy	
Pregnancy type	None required
Not pregnant	
Biochemical	
Clinical intrauterine	
# fetal sac	None required
# fetal heart	None required
Chorionicity	Encourage clinics to obtain early ultrasound from health care provider in pregnancy from which to obtain chorionicity or amnionicity. Strict definition based on either early ultrasound or pathology report of placenta

Appendix 13. Checklist of reporting criteria for studies validating health administrative data algorithms..

CHECKLIST ITEM	
TITLE, KEYWORDS, ABSTRACT	**REPORTED ON PAGE #**
Identify article as study of assessing diagnostic accuracy	55
Identify article as study of administrative data	56-57
INTRODUCTION:	
State disease identification & validation one of goals of study	59-60
METHODS:	
Participants in validation cohort:	
Describe the validation cohort (Cohort of patients to which reference standard was applied)	61
•Age	NA
• Disease	NA
• Severity	NA
• Location/Jurisdiction	61
Describe recruitment procedure of validation cohort	61
• Inclusion criteria	61
• Exclusion criteria	
Describe patient sampling (random, consecutive, all, etc.)	61
Describe data collection	62
• Who identified patients and did selection adhere to patient recruitment criteria	61
• Who collected data	62
• *A priori* data collection form	62
• Disease classification	NA
• Split sample (i.e. re-validation using a separate cohort)	NA
Test Methods:	
Describe number, training and expertise of persons reading reference standard	62
If >1 person reading reference standard, quote measure of consistency (e.g. kappa)	62
Blinding of interpreters of reference standard to results of classification by administrative data e.g. Chart abstractor blinded to how that chart was coded	62
Statistical Methods:	
Describe methods of calculating/comparing diagnostic accuracy	62-63
Participants:	
Report when study done, start/end dates of enrollment	60
Describe number of people who satisfied inclusion/exclusion criteria	61
Study flow diagram	NA
Test results:	
Report distribution of disease severity	NA

Report cross-tabulation of index tests by results of reference standard	NA
Estimates:	
Report at least 4 estimates of diagnostic accuracy	Table 2, Table 3
Diagnostic Accuracy Measures Reported:	
• Sensitivity	Table 3
• Specificity	Table 3
• PPV	Table 3
• NPV	Table 3
• Likelihood ratios	NA
• Kappa	Table 3
• Area under the ROC curve / c-statistic	NA
• Accuracy/agreement	Table 2, Table 3
• Other (specify)	
Report accuracy for subgroups (e.g. age, geography, different sex, etc.)	Table 4, Table 5, Table 6, Appendix 5, Appendix 6, Appendix 7, Appendix 8, Appendix 9, Appendix 10
If PPV/NPV reported, ratio of cases/controls of validation cohort approximate prevalence of condition in the population	Appendix 1
Report 95% confidence intervals for each diagnostic measure	Table 2, 3
DISCUSSION:	
Discuss the applicability of the validation findings	69-76

Adapted from Benchimol EI, et al (11)

Chapter 4. Discussion

4.1 Infertility burden

The World Health Organization recognizes the inability to have a healthy child after one year of attempting pregnancy as a disease or a disability (1). The psychosocial implications of infertility are vast, including depression, discrimination and ostracism, the latter being of particular importance in lower income countries (2). In the developing world, infertility has been estimated to affect 186 million women (3). The use of ART for the treatment of infertility has become widely available in many countries globally, though access varies greatly by country (4).

This field of medicine is evolving rapidly as biomedical techniques are improving (5). There have been many successes since the first IVF baby was born in 1978 (6). Among many advancements, intracytoplasmic semen injection, whereby sperm are injected directly into the oocyte, has provided a treatment option for couples affected by severe male factor infertility (7). Preimplantation genetic diagnosis can be achieved by obtaining a biopsy of an embryo and investigating for a specific genetic condition, after which selective transfer of unaffected embryos can be performed (8). Embryo and oocyte cryopreservation allows freezing and storage for use in a subsequent treatment cycle, preventing disposal of unused gametes and embryos (9,10). Cryopreservation of embryos also allows for a reduction in the number of embryos transferred (and reduction in the risk of multiple gestation), as none will be discarded. With these new treatments and techniques, it is essential we can adequately monitor both short-term and long-term outcomes.

4.2 ART monitoring

Important aspects in monitoring ART practices include patient demographics, method of embryo or oocyte cryopreservation, survival of embryos after thawing (11–13). Outcomes after treatment cycles are especially important to monitor as pregnancies conceived through ART have a higher association with adverse outcomes compared to those spontaneously conceived, including preterm delivery, hypertensive disorders of pregnancies and neonatal intensive care unit admissions (14–17). Multifetal gestation pregnancies, which are strongly associated with both ovulation induction and IVF due to multiple embryo transfer, account for a large proportion of these complications (18–20). The Canadian Fertility and Andrology Society (CFAS), the American Society of Reproductive Medicine (ASRM) and National Institute for Health and Care Excellence (NICE) have published recommendations to reduce the risk of multiple pregnancies, which should be monitored to assess adherence to target guidelines (21–23).

Determining the prevalence and burden of infertility, as well as performing regular surveillance on ART treatments and outcomes are essential to both inform policy, conduct research and counsel patients (2,24). For example, the International Committee for Monitoring Assisted Reproductive Technologies depends on large-population data from regional and national ART registries around the world (25). With these data, they are able to provide reports depicting trends in practice, utilization of health care, and pregnancy outcomes after treatment (25). Accurate and robust data are paramount to providing such reports. While these reports are reliant on administrative databases, our systematic review demonstrated that the quality assurance practices to establish accurate and reliable data are lacking in the literature. Moreover, where reports were published, adherence to reporting guidelines for studies using administrative data was also insufficient. We have provided a comprehensive review of the current literature

describing current practices, various strategies, and guidelines for which a validation study should adhere to in order to ensure accurate data.

Despite specifically searching for multifetal gestation pregnancies in our systematic search strategy, we were unable to identify any studies that validated this variable in an ART context. In the CARTR Plus validation study, there was excellent agreement with the number of embryos transferred as well as the number of fetal sacs and heart beats on ultrasound. The kappa coefficient for chorionicity, as well as the sensitivity for 1 chorion were quite poor, either due to small sample size or lack of reporting to the database. Due to the excellent quality of data in the ultrasound and embryology elements, these variables can safely be used for reporting to government agencies and performing surveillance studies of adherence to guidelines. There is still room for improvement, and urging clinics to report extensive pregnancy information to CARTR Plus should be the next step.

4.3 Reference standards

There are three commonly cited validation study designs ecological studies, reabstraction studies and gold standard studies (26). Ecological studies compare statistics of the code to those obtained from more reliable methods. Reabstraction studies compare the code to the medical record. Finally, gold standard studies compare the code to a case definition, either based on clinical or laboratory values or clinical consensus (26).

Hemminki and Gissler used national statistics to identify how valid their aggregate data are. Firstly, these reference standards rely on the accuracy of the national statistics, which were not established and should not be implicitly assumed (27,28). Secondly, the comparison is based on aggregate data rather than at the patient level, therefore identifying specific differences and agreements is impossible.

In our national validation study, we compared database records to the medical charts. The validity was excellent for many variables. Areas requiring attention are those reliant on clinical judgment, namely reason for treatment and elective embryo transfer. Molinaro et al attempted to validate diagnosis variables in SART using case definitions based on clinical values in the patients' charts rather than relying on the expertise of clinicians. They did not report their measures of validity making it challenging to determine if this method is superior (29). Using objective measures, like laboratory tests and strict diagnostic criteria for validation compared to documentation may be more reliable, though studies are lacking in an ART population.

4.4 Population selection

Ideally, a validation study would sample from the reference standard to determine how well the database is capturing the population and identify the prevalence of false negatives. Certainly, the validity of CARTR Plus depends on the quality and completeness of data entry. Determining the degree of completeness can be achieved by sampling from the reference standard to assess whether cycles are not recorded in the database.

Unfortunately, due to both financial and time constraints, sampling from our reference standard of the medical charts was not possible, as this would require the clinics to randomly sample charts from their clinics for us to verify if their record exists in the database. Consequently, we are unable to determine the prevalence of missing ART cycles in the CARTR Plus database. One way researchers have overcome this issue was to link birth registry data, containing a variable indicating method of conception (natural or ART), to the ART registry (30-37). This method, however, assumes that all patient underwent ART cycles in the state or country similar to that of delivery, which is often not the case. Again, this method assumes that

the reference standard has been adequately validated, which as demonstrated from our systematic review, is not always the case.

4.5 Measures of validity

The most commonly cited measures of validation in an ART population were sensitivity, specificity and PPV. Sensitivity reflects the ability of the code to identify those with the condition specified as compared to the reference standard, representing the completeness of reporting (38). The positive predictive value denotes the proportion of those reported to have the condition in the database who actually have the condition, representing the accuracy of the database. The kappa coefficient represents the agreement between the two data sources beyond chance. Interestingly, none of the studies included in the systematic review reported cut-offs for an appropriate level of completeness, accuracy or agreement. This is likely a result of the utility of the measures. For example, in screening tests for disease, it may be safer to have a higher false positive rate to maximize sensitivity; the specificity and positive predictive value diminish. Consequently, further testing due to abnormally elevated estimates of disease would be warranted. In the context of code validity, when assessing the validity of serious or adverse outcomes, (for example, ovarian hyperstimulation syndrome or multiple gestation), it would be reasonable to sacrifice the specificity and PPV to optimize sensitivity and ensure all cases are captured. If the prevalence of these conditions is higher than expected, investigation into determining etiology and practice changes would be warranted.

In our systematic review, many studies validated whether ART was used in the conception of a pregnancy, with the intention of identifying a high-risk obstetrical population. For example, Cohen et al found a sensitivity of 41% for ART use with a PPV of 55% in Massachusetts (30). The low sensitivity indicates that the study population captures less than half

of women who underwent ART, whereas the positive predictive value indicates that 55% women who were coded as having undergone ART actually underwent ART. Buck Louis et al demonstrated a 91% sensitivity for abnormal fetal conditions when comparing maternal report to the reference standard of the ART registry, with a PPV of 85%. The excellent sensitivity will pick up most infants with a condition and signal an elevated rate when present. The PPV indicates that 15% of infants will be mislabeled as having an abnormal condition, when they were in fact normal (39).

The majority of the included studies in the systematic review presented fewer than four estimates of validation and only eight studies reported confidence intervals for their estimates. By presenting multiple estimates, as we have done in the Canadian validation study, the reader is able to appreciate not only the degree of validity, but also where further investigation is warranted. For example, elective single embryo transfer is used as a surrogate for poor cycle outcome (indicating that only one embryo was available for replacement after stimulation). With a sensitivity of 76%, almost one quarter of cycles were mislabeled as being non-elective when they were elective. The resulting flag that cycle outcomes were worse than they truly were may lead clinicians to investigate how they are providing hormonal stimulation to patients, or embryologists to assess the culture media of the growing embryos.

4.6 Implications for future use

The importance of ensuring validated data when using routinely-collected data cannot be understated. When developing policy, like reducing rates of multiple gestation after ART by transferring the minimum embryos required to achieve pregnancy, we pull statistics from registry and administrative health databases. A low sensitivity in multiple fetal hearts or in chorionicity (with a high false negative rate) would mislead the reader to believe that the health care

practitioner is adhering to current practice guidelines. In our study, we found that variables reliant on clinical judgement had lower sensitivity, PPV and kappa estimates. Areas of human error, or clerical errors, were less common.

For oocyte source, the majority of error was based on missingness, which can largely be attributed to clerical error. Importantly, researchers developing study protocols using ART populations often exclude participants based on the oocyte source, especially oocyte donor cycles. In our study, if participants were excluded if they were not autologous oocyte 13% of the population, who are autologous oocyte providers would be inappropriately excluded. We, therefore, recommend researchers to use an imputation strategy to avoid excluding a large fraction of the population.

The variables of greater agreement were those that were based on lab values and discrete events (whether an embryo transfer was performed, number transferred, type of cycle initiated). Based on these findings, CARTR Plus users can rely on similarly structured data elements. Areas to use caution would be diagnosis variables until case definitions can be better described. For example, rather than relying on the diagnosis of "diminished ovarian reserve" as a reason for treatment, developing an algorithm incorporating markers of ovarian reserve including antral follicle count, serum FSH level and anti-mullerian hormone may be superior. These algorithms would need to be validated against a reference standard prior to their use.

4.7 Final conclusions

By performing the first validation study of CARTR Plus, we were able to investigate the validity of commonly used variables in the database. Overall, we found very high data quality for many of the variables that were analyzed. As such, these elements can be used in surveillance reports as well as research studies; researchers can now reliably document utilization of validated

data when using CARTR Plus variables investigated in this study. Furthermore, specific areas of concerns will be addressed to improve accuracy in the data collection process.

Both our systematic review and the validation study provide the first assessment of the quality of national ART registries and administrative databases using ART data, as well as the CARTR Plus database. The rigorous methodology used can serve as a guide for future validation projects.

4.8 References

1. World Health Organization. Meeting to develop a global consensus on preconception care to reduce maternal and childhood mortality and morbidity [Internet]. Geneva: 2012 [cited 2018 Apr 12]. Available from: http://apps.who.int/iris/bitstream/handle/10665/78067/9789241505000_eng.pdf?sequence=1&isAllowed=y
2. Mascarenhas MN, Flaxman SR, Boerma T, Vanderpoel S, Stevens GA. National, regional, and global trends in infertility prevalence since 1990: A systematic analysis of 277 health surveys. PLoS Med. 2012;9(12):e1001356.
3. Rutstein S, Iqbal H. Infecunditiy, infertility, and childlessness in developing countries. DHS Comp Reports No 9 [Internet]. 2004 [cited 2018 Apr 12];Available from: http://www.measuredhs.com
4. Adamson GD. Global cultural and socioeconomic factors that influence access to assisted reproductive technologies. Women's Health. 2009;5(4):351–8.
5. Kamel RM. Assisted reproductive technology after the birth of Louise Brown. J Reprod Infertil. 2013;14(3):96–109.
6. Steptoe PC, Edwards RG. Birth after the reimplantation of a human embryo. Lancet. 1978;2(8085):366.
7. Palermo G, Joris H, Devroey P, Van Steirteghem AC. Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyte. Lancet. 1992;340(8810):17–8.
8. Xu K, Rosenwaks Z, Beaverson K, Cholst I, Veeck L, Abramson DH. Preimplantation genetic diagnosis for retinoblastoma: the first reported liveborn. Am J Ophthalmol. 2004;137(1):18–23.
9. Gordts S, Roziers P, Campo R, Noto V. Survival and pregnancy outcome after ultrarapid freezing of human embryos. Fertil Steril. 1990;53(3)`:469–72.
10. Donnez J, Dolmans MM. Fertility preservation in women. N Engl J Med. 2017;377(17):1657–65.
11. European Society of Human Reproduction and Embryology. Revised guidelines for good practice in IVF laboratories (2015). Hum Reprod. 2016;31(4):685–6.
12. Magli MC, Van den Abbeel E, Lundin K, Royere D, Van der Elst J, Gianaroli L, et al. Revised guidelines for good practice in IVF laboratories. Hum Reprod.2008;23(6):1253–62.
13. Cohen MA, Lindheim SR, Sauer M V. Donor age is paramount to success in oocyte donation. Hum Reprod. 1999;14(11):2755–8.
14. McGovern PG, Llorens AJ, Skurnick JH, Weiss G, Goldsmith LT. Increased risk of preterm birth in singleton pregnancies resulting from in vitro fertilization–embryo transfer or gamete intrafallopian transfer: A meta-analysis. Fertil Steril 2004;82(6):1514–20.
15. Pandey S, Shetty A, Hamilton M, Bhattacharya S, Maheshwari A. Obstetric and perinatal outcomes in singleton pregnancies resulting from IVF/ICSI: a systematic review and meta-analysis. Hum Reprod Update. 2012;18(5):485–503.
16. Jackson RA, Gibson KA, Wu YW, Croughan MS. Perinatal Outcomes in Singletons Following In Vitro Fertilization: A Meta-Analysis. Obstet Gynecol. 2004;103(3):551–63.
17. Sazonova A, Källen K, Thurin-Kjellberg A, Wennerholm U-B, Bergh C. Obstetric outcome after in vitro fertilization with single or double embryo transfer. Hum Reprod [Internet] 2011 [cited 2016 Jul 25];26(2):442–50.

18. Ombelet W, De Sutter P, Van der Elst J, Martens G. Multiple gestation and infertility treatment: registration, reflection and reaction--the Belgian project. Hum Reprod Update. 2005;11(1):3–14.
19. Chambers GM, Chapman MG, Grayson N, Shanahan M, Sullivan EA. Babies born after ART treatment cost more than non-ART babies: a cost analysis of inpatient birth-admission costs of singleton and multiple gestation pregnancies. Hum Reprod. 2007;22(12):3108–15.
20. Ledger WL, Anumba D, Marlow N, Thomas CM, Wilson EC. Fertility and assisted reproduction: The costs to the NHS of multiple births after IVF treatment in the UK. BJOG. 2006;113(1):21–5.
21. National Institute for Health and Care Excellence. Fertility problems: assessment and treatment [Internet]. Natl. Inst. Heal. Care Excell. 2013 [cited 2016 Jul 27];Available from: https://www.nice.org.uk/guidance/cg156/resources/fertility-problems-assessment-and-treatment-35109634660549
22. Committee of the Society for Assisted Reproductive Technology P, Committee of the American Society for Reproductive Medicine P. Elective single-embryo transfer. Fertil Steril 2012;97:835–42.
23. Min J, Sylvestre C. Guidelines on the number of embryos transferred [Internet]. Canadian Fertilitiy Andrology Society. 2013 [cited 2016 Jul 25];Available from: http://www.cfas.ca/images/stories/pdf/cfas_cpg_embryo_transfer_2013.pdf
24. Okun N, Sierra S. Pregnancy outcomes after assisted human reproduction. J Obstet Gynaecol Can. 2014;36(1):64–83.
25. 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.
26.Van Walraven C, Austin P. Administrative database research has unique characteristics that can risk biased results. J Clin Epidemiol. 2012;65(2):126–131.
27. Gissler M, Klemetti R, Sevón T, Hemminki E. Monitoring of IVF birth outcomes in Finland: a data quality study. BMC Med Inform Decis Mak 2004;4.
28. Hemminki E, Klemetti R, Rinta-Paavola M, Martikainen J. Identifying exposures of in vitro fertilization from drug reimbursement files: a case study from Finland. Med Inform Internet Med 2003;28:279–289.
29. Cohen B, Bernson D, Sappenfield W, Kirby RS, Kissin D, Zhang Y, Copeland G, Zhang Z, Macaluso M, States Monitoring Assisted Reproductive Technology Collaborative. Accuracy of assisted reproductive technology information on birth certificates: Florida and Massachusetts, 2004-06. Paediatr Perinat Epidemiol 2014;28:181–190.
30. Luke B, Brown MB, Spector LG. Validation of infertility treatment and assisted reproductive technology use on the birth certificate in eight states. Am J Obstet Gynecol 2016;215:126–127.
31. Sunderam S, Schieve LA, Cohen B, Zhang Z, Jeng G, Reynolds M, Wright V, Johnson C, Macaluso M, Massachusetts Consortium for Assisted Reproductive Technology Epidemiologic Research (MCARTER). Linking birth and infant death records with assisted reproductive technology data: Massachusetts, 1997–1998. Matern Child Health J 2006;10:115–125.
32. Zhang Z, Macaluso M, Cohen B, Schieve L, Nannini A, Chen M, Wright V, Research MC

for ARTE. Accuracy of assisted reproductive technology information on the Massachusetts birth certificate, 1997-2000. Fertil Steril 2010;94:1657–1661.
33. Williams CL, Bunch KJ, Stiller CA, Murphy MFG, Botting BJ, Wallace WH, Davies M, Sutcliffe AG. Cancer risk among children born after assisted conception. N Engl J Med 2013;369:1819–1827.
34. Stern JE, Gopal D, Liberman RF, Anderka M, Kotelchuck M, Luke B. Validation of birth outcomes from the Society for Assisted Reproductive Technology Clinic Outcome Reporting System (SART CORS): Population-based analysis from the Massachusetts Outcome Study of Assisted Reproductive Technology (MOSART). Fertil Steril 2016;106:717–722.
36. . Kotelchuck M, Hoang L, Stern J, Diop H, Belanoff C, Declercq E. The MOSART Database: Linking the SART CORS Clinical Database to the Population-Based Massachusetts PELL Reproductive Public Health Data System. Matern Child Heal J 2014;18:2167–2178.
37. Zhang Y, Cohen B, Macaluso M, Zhang Z, Durant T, Nannini A. Probabilistic linkage of assisted reproductive technology information with vital records, Massachusetts 1997-2000. Matern Child Heal J 2012;16:1703–1708.
38. Lain SJ, Hadfield RM, Raynes-Greenow CH, Ford JB, Mealing NM, Algert CS, Roberts CL. Quality of data in perinatal population health databases: a systematic review. Med Care 2012;50:e7-e20.
39. Buck Louis GM, Druschel C, Bell E, Stern JE, Luke B, McLain A, Sundaram R, Yeung E. Use of assisted reproductive technology treatment as reported by mothers in comparison with registry data: the Upstate KIDS Study. Fertil Steril 2015;103:1461–1468.

Appendix C: Certificates of ethical approval

TOH approval

Ottawa Hospital Research Institute
Institut de recherche de l'Hôpital d'Ottawa

uOttawa

University of Ottawa Heart Institute
Institut de cardiologie de l'Université d'Ottawa

Ottawa Health Science Network Research Ethics Board/ Conseil d'éthique de la recherche du Réseau de science de la santé d'Ottawa

Civic Box 411 725 Parkdale Avenue, Ottawa, Ontario K1Y 4E9 613-798-5555 ext. 14902 Fax : 613-761-4311
http://www.ohri.ca/ohsn-reb

December 13, 2016

Dr. Mark Walker
Attn: Ruth Rennicks White
Ottawa Hospital - General Campus
Department of Obstetrics/Gynecology/Newborn Care
Center for Practice-Changing Research, Room L1241, Box 241
501 Smyth Road, Ottawa, ON K1H 8L6

Dear Dr. Walker:

Re: Protocol # 20160862-01H **An evaluation of the validity of the Canadian Assisted Reproductive Technologies Register (CARTR) Plus database**

Protocol approval valid until - December 12, 2017

I am pleased to inform you that this protocol underwent delegated review by the Ottawa Health Science Network Research Ethics Board (OHSN-REB) and is approved. No changes, amendments or addenda may be made to the protocol or the consent form without the OHSN-REB's review and approval.

Approval is for the following:
- Protocol (version 1) dated September 15, 2016
- Case Report Form dated November 17, 2016
- English Information Letter dated December 12, 2016
- English Letter of Support dated December 6, 2016
- English Reminder Letter dated December 12, 2016

Your request for a French exemption is approved; the study may proceed in English only.

The REB no longer requires a 'valid until' date at the bottom of all approved informed consent forms. The consent forms currently approved for use by the REB are listed above.

If the study is to continue beyond the expiry date noted above, a Renewal Form should be submitted to the REB approximately six weeks prior to the current expiry date. If the study has been completed by this date, a Termination Report should be submitted.

The OHSN-REB complies with the membership requirements and operates in compliance with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans; the International Conference on Harmonization - Good Clinical Practice: Consolidated Guideline, and the provisions of the Personal Health Information Protection Act 2004.

Yours sincerely,

[redacted]

Francine F-A Sarazin, Ph.D., C.Psych.
Vice-Chairperson
Ottawa Health Science Network Research Ethics Board

FFAS/kd

The Ottawa Hospital | L'Hôpital d'Ottawa
RESEARCH INSTITUTE | INSTITUT DE RECHERCHE
uOttawa
UNIVERSITY OF OTTAWA HEART INSTITUTE
INSTITUT DE CARDIOLOGIE DE L'UNIVERSITÉ D'OTTAWA

Ottawa Health Science Network Research Ethics Board/ Conseil d'éthique de la recherche du réseau de science de la santé d'Ottawa

Civic Box 675, 725 Parkdale Avenue, Ottawa, Ontario K1Y 4E9 613-798-5555 ext. 16719 Fax : 613-761-4311
http://www.ohri.ca/ohsn-reb

December 5, 2016

To Whom It May Concern:

This letter is to confirm that Amy Geertsma, Manager, Ottawa Health Science Network Research Ethics Board, is hereby authorized to sign ethics correspondence in my absence.

I will be away from December 8, 2016 thru December 18, 2016 inclusive.

Yours sincerely,

[redacted]

Raphael Saginur, M.D.
Chairperson
Ottawa Health Science Network Research Ethics Board

RS/km

Ottawa Hospital Research Institute
Institut de recherche de l'Hôpital d'Ottawa

uOttawa

UNIVERSITY OF OTTAWA HEART INSTITUTE
INSTITUT DE CARDIOLOGIE DE L'UNIVERSITÉ D'OTTAWA

Ottawa Health Science Network Research Ethics Board/ Conseil d'éthique de la recherche du Réseau de science de la santé d'Ottawa
Civic Box 411 725 Parkdale Avenue, Ottawa, Ontario K1Y 4E9 613-798-5555 ext 14902 Fax : 613-761-4311
http://www.ohri.ca/ohsn-reb

November 2, 2017

Dr. Mark Walker
Attn: Ruth Rennicks White
Ottawa Hospital - General Campus
Department of Obstetrics/Gynecology/Newborn Care
Center for Practice-Changing Research, Room L1241, Box 241
501 Smyth Road, Ottawa, ON K1H 8L6

Dear Dr. Walker:

Re: Protocol # 20160862-01H An evaluation of the validity of the Canadian Assisted Reproductive Technologies Register (CARTR) Plus database

The Protocol Amendment Report dated August 28, 2017 and revised Protocol (version 2) dated August 28, 2017 are approved.

Date of approval: November 2, 2017

Ethical approval remains in effect until December 12, 2017.

The OHSN-REB complies with the membership requirements and operates in compliance with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans; the International Conference on Harmonization - Good Clinical Practice: Consolidated Guideline and the provisions of the Personal Health Information Protection Act 2004.

Yours sincerely,

[redacted]

Raphael Saginur, M.D.
Chairperson
Ottawa Health Science Network Research Ethics Board

/kd

Ottawa Hospital Research Institute
Institut de recherche de l'Hôpital d'Ottawa

uOttawa

UNIVERSITY OF OTTAWA HEART INSTITUTE
INSTITUT DE CARDIOLOGIE DE L'UNIVERSITÉ D'OTTAWA

Ottawa Health Science Network Research Ethics Board/ Conseil d'éthique de la recherche du Réseau de science de la santé d'Ottawa

Civic Box 411 725 Parkdale Avenue, Ottawa, Ontario K1Y 4E9 613-798-5555 ext. 14902 Fax : 613-761-4311
http://www.ohri.ca/ohsn-reb

Monday, November 20, 2017

Dr. Mark Walker
Attn: Ruth Rennicks White
Ottawa Hospital - General Campus
Department of Obstetrics/Gynecology/Newborn Care
Center for Practice-Changing Research, Room L1241, Box 241
501 Smyth Road, Ottawa, ON K1H 8L6

Dear Dr. Walker:

RE: Protocol# - 20160862-01H **An evaluation of the validity of the Canadian Assisted Reproductive Technologies Register (CARTR) Plus database**

Renewal Expiry Date - Friday, November 16, 2018

Thank you for your Annual Renewal submission undated (received October 20, 2017). I am pleased to inform you that your Annual Renewal Request was reviewed by the Ottawa Health Science Network Research Ethics Board (OHSN-REB) and is approved. No changes, amendments or addenda may be made in the protocol or the consent form without the OHSN-REB's review and approval.

This renewal is approved as of: November 17, 2017.

The projected date of study completion has been extended to December 2018.

The file has been updated to reflect the closing of recruitment.

Renewal is valid for a period of one year. . If the study is to continue beyond the expiry date noted above, a Renewal Form should be submitted to the REB, in hardcopy. All Annual Renewal Reports, regardless of review type (i.e., full board or delegated), must now be submitted according to the full board meeting submission deadlines AND at least 30 days prior to the expiry date of the study to prevent a lapse in approval. If the study is completed by this date, a Termination Report should be submitted.

The OHSN-REB complies with the membership requirements and operates in compliance with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans; the International Conference on Harmonization - Good Clinical Practice: Consolidated Guideline; and the provisions of the Personal Health Information Protection Act 2004.

Yours sincerely,

[redacted]

Raphael Saginur, M.D.
Chairperson
Ottawa Health Science Network Research Ethics Board

MV/gb

CHEO exemption

02 December 2016

Dr. Deshayne Fell
CHEO INTRA
Clinical Research\ Epidemiology

Romeo# 20160484
REB# 16/141X
Title: **An evaluation of the validity of the Canadian Assisted Reproductive Technologies Register (CARTR) Plus database**

Dear Dr. Fell,

Thank you for your recent submission of an investigator response. The investigator response indicated that this study was a research endeavor. Upon review, the Board completed the ARECCI Ethics Screening Tool with the information provided in the REB application. The screening tool indicated that the submission is probably a quality improvement project. As such this study is exempt from REB approval.

Although the project is exempt from REB approval, the investigator is responsible for complying with all TCPS 2 guidance relevant to their project as it pertains to the respect for the intrinsic value of human beings, the concern for welfare and the obligation to treat people fairly and equitably.

The investigator should ensure that ethics approval or exemption is secured from each of the sites that agree to participate in this project. It should be noted that the CHEO REB cannot act as Board of Record for the participating clinics.

Kindly refer to the above-mentioned proposal number in all future correspondence. Should you have any questions or wish to discuss this further, please contact the office of the REB at extension 2128.

Regards,

[redacted]

Franco Momoli, *M.Sc., Ph.D.*
Interim Chair, CHEO Research Ethics Board