

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

V. Bacal^{1,2,3,*}, D.B. Fell^{2,4}, H. Shapiro^{6,7}, A. Lanes^{4,5}, A.E. Sprague^{4,5}, M. Johnson⁵, and M. Walker^{1,2,3,5}, and L.M. Gaudet^{2,3,8}

¹Department of Obstetrics and Gynecology, University of Ottawa, Ottawa, Canada ²School of Epidemiology and Public Health (SEPH), University of Ottawa, Ottawa, Canada ³Ottawa Hospital Research Institute, Ottawa, Canada ⁴Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada ⁵Children's Hospital of Eastern Ontario, Better Outcomes Registry and Network (BORN), Ontario, Ottawa, Canada ⁶Department of Obstetrics and Gynecology, University of Toronto, Toronto, Canada ⁷Mount Sinai Fertility, Toronto, Canada ⁸Department of Obstetrics and Gynecology, Queen's University, Kingston, Canada

*Correspondence address. Department of Obstetrics and Gynaecology, The Ottawa Hospital, 501 Smyth Road, Ottawa, ON K1H8L6, Canada. Tel: 613-798-5555; E-mail: vbaca078@uottawa.ca <http://orcid.org/0000-0002-6816-5616>

Submitted on October 7, 2019; resubmitted on January 7, 2020; editorial decision on January 22, 2020

STUDY QUESTION: Are data accurately documented in the Canadian Assisted Reproductive Technologies Register (CARTR) Plus database?

SUMMARY ANSWER: Measures of validity were strong for the majority of variables evaluated while 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 evaluated prior to utilization for research both to inform the study design and to determine the extent to which key study variables, such as patient characteristics or therapies provided, are accurately documented in the database. CARTR Plus is Canada's national register for collecting extensive information on IVF and corresponding pregnancy outcomes, and it has yet to be validated.

STUDY DESIGN, SIZE, DURATION: This study evaluating the data translation CARTR Plus database examined IVF cycles performed in 2015 using data directly from patient charts. 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. Only a single treatment cycle record from a unique patient at each clinic was considered during chart selection.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Twenty-five data elements (patient characteristics, treatments and outcomes) were reabstracted from patient charts, which were declared the reference standard. Data were reabstracted by two independent auditors with relevant clinical knowledge after confirming inter-rater reliability. These data elements from the chart were then compared to those in CARTR Plus. To determine the validity of these variables, we calculated kappa coefficients, sensitivity, specificity, positive predictive value and negative predictive value with 95% CI for categorical variables and calculated median differences and intraclass correlation coefficients (ICC) for continuous variables.

MAIN RESULTS AND THE ROLE OF CHANCE: Six clinics agreed to participate in this study representing five 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 were 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). Agreement was nearly perfect for pregnancy type, and number of fetal sacs and fetal hearts on ultrasound, all with kappa coefficients greater than 0.90.

LARGE-SCALE DATA: N/A

LIMITATIONS, REASONS FOR CAUTION: CARTR Plus contains over 200 variables, of which only 25 were assessed in this study. This foundational validation work should be extended to other CARTR Plus database variables in future studies.

WIDER IMPLICATIONS OF THE FINDINGS: This study provides the first assessment of the quality of the data translation process of the CARTR Plus database, and we found very high quality for the majority of the variables that were analyzed. We identified key data points that are either too often lacking or inconsistent with chart data, indicating that changes in the data entry process may be required.

STUDY FUNDING/COMPETING INTEREST(S): This study was funded by Canadian Institutes of Health Research (CIHR) (Grant Number FDN-148438) and by the Canadian Fertility and Andrology Society Research Seed Grant (Grant Number: N/A). The authors report no conflict of interest.

TRIAL REGISTRATION NUMBER Not applicable.

Key words: ART / infertility / database / quality assurance / validation / reproductive epidemiology

WHAT DOES THIS MEAN FOR PATIENTS?

Health care and registry databases are excellent sources of information that can be used in research and for quality assurance purposes. They are relatively inexpensive, easily accessible and collect information about a large number of people. This information includes patient characteristics, diagnoses and treatments. However, because this information is not collected specifically for a research project, it may contain significant inaccuracies. In order to use these databases for the best-quality research, they must be checked through a process called validation.

The CARTR Plus database is the only national database in Canada collecting extensive information on infertility cycles from over 35 IVF clinics in Canada. Since CARTR Plus began collecting patient cycle characteristics in 2013, the information it contains has not been validated. In this article, the authors looked at the data entry processes of CARTR Plus database. Six clinics spanning Canada were recruited to participate. Data were collected from patient charts and compared to those from the database.

Overall, they found very high data quality for most of the information they looked at. They concluded that this database can be used to report back to government bodies or health care professionals, as well as future use in research studies. They also provide guidance for specific ways in which the database can become more error-free.

Introduction

In 2001, the Royal Commission for New Reproductive Technology estimated that one-quarter of a million Canadian couples have difficulty conceiving (Norris, 2001). 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 (Bushnik et al., 2012). 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 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 3-fold (Connolly et al., 2010). In Canada, the incidence of preterm birth in treated infertile couples in 2014 was 24–28% (Canadian Assisted Reproductive Technologies Register (CARTR) Plus, 2015), three times higher than the general obstetrical population (Public Health Agency of Canada, 2017). Other complications that contribute to the indirect cost include ectopic pregnancy, placenta previa and preeclampsia, which also have a higher incidence among ART pregnancies (Maymon and Shulman, 1996; Romundstad et al., 2006; Pandey et al., 2012). Despite the elevated risks for these conditions, their absolute incidence remains low (Sazonova et al., 2011), 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 and easily accessible and are collected on a population scale (Iron and Manuel, 2007; Benchimol et al., 2011). These data can be used

for evaluating access to and quality of health care, health service planning, reporting to governing bodies and clinical research (Iron and Manuel, 2007). However, routinely collected data are generally not collected with the intent of performing research (Benchimol et al., 2011). As a result, studies reliant on these data are subject to misclassification, unmeasured confounding due to missing variables and missing data (Benchimol et al., 2015). The accuracy of routinely collected data is subject to errors from inter-observer discrepancies, documentation problems, illegible charts, missing data elements and timeliness of input into the database (Hierholzer Jr, 1991). In order to establish an adequate quality of data and reduce potential misclassification bias in research, studies to measure the accuracy of variables contained within health databases are highly recommended as per the REporting of studies Conducted using Observational Routinely-collected Data guideline (Benchimol et al., 2015).

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 for all patients undergoing IVF since 2013 from all 33 ART clinics across Canada. CARTR Plus is the only database in Canada to contain national IVF data, and the accuracy of the data has not yet been assessed. Because these data may be used to inform policymakers 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 evaluate a subset of clinically relevant variables from CARTR Plus to determine the extent to which key study variables are accurately documented in the database.

Materials and Methods

Study design

This data quality study evaluating the data translation process of CARTR Plus database examined IVF cycles from 1 January 1 2015 to 31 December 2015 using patient chart reabstraction as the gold reference standard.

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 third party not involved in the clinic selection, and its name was only revealed after the clinic was chosen. We selected our six clinics from five 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' category based on an 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 of data input (determined based on the degree of missingness of key data elements submitted monthly by clinics, as well as the clinics' final submission prior to the development of the annual report). 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 retrieved by the clinic.

Data extraction

We identified 25 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 [Supplementary Table S1](#) for the 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 they are likely to have high agreement with chart data but provide little insight into the mechanism behind the missingness ([Dunn et al., 2011](#)). 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 the 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. REDCap is a secure web-based application that encrypts input data ensuring that patient privacy is maintained ([Harris et al., 2009](#)). Each REDCap entry of chart data was double-checked for errors.

Statistical analyses

We analyzed characteristics of the sample groups using frequencies for categorical variables, and means and SDs or medians and interquartile ranges (IQR) for continuous variables, stratified by source of data (reabstracted versus database-derived). Reabstracted data from charts were considered the reference standard.

Sensitivity reflects the proportion of all records in which a diagnosis or procedure is documented on the medical chart that are also entered as such into the CARTR Plus database. Specificity reflects the proportion of all records in which a diagnosis is not documented in the medical chart and is also not entered into the database. The positive predictive value (PPV) denotes the proportion of diagnoses entered into CARTR Plus that were also entered as such in the medical chart, representing the accuracy of the database. The kappa coefficient represents the agreement between the two data sources while accounting for agreement or disagreement due to chance ([Sim and Wright, 2005](#)).

For categorical variables, we calculated kappa coefficients, sensitivities, specificities, PPV and negative predictive values (NPV) with 95% CI. For continuous variables, we computed the median absolute difference between the two data sources and performed Wilcoxon signed-rank tests to assess for statistically significant differences. We also calculated and intraclass correlation coefficients (ICC) with 95% CI. Kappa coefficients and ICCs were graded according to the levels described by [Landis and Koch \(1977\)](#). Percentage 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, USA).

Hypothesis

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

Sample size justification

We performed *a priori* sample size calculations to determine the number of records we would need to compare between the two sources (i.e. database and reabstracted medical charts). For continuous variables, this was based on being able to estimate an ICC of 0.90 with

a 95% CI yielding a margin of error ± 0.10 (Zou, 2012). 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 , using estimates of the prevalence for each of the categorical variables generated from historical CARTR Plus data (Donner and Eliasziw, 1987; Bartfay and Donner, 2001; Donner and Zou, 2002). These calculations generated a minimum total sample size of 726.

Finally, in order to account for potential missing data of up to 20% for some 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.

Results

Six clinics agreed to participate in this study representing five Canadian provinces. The cycle volume per clinic in 2015 ranged from 329 to 2212. We collected data from a total of 876 patient charts. There were 12 charts that were not retrievable at one clinic site, which were assumed to be missing 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. Among the 876 charts that were reabstracted, comparing data retrieved from patient records, variables with the greatest amount of missing data in CARTR Plus (as a result of not being entered) included Day 2–4 FSH (31% of reabstracted charts), antral follicle count (AFC) (38% of reabstracted charts), anti-Müllerian hormone (AMH) (62% of reabstracted charts) and oocyte origin (13% of reabstracted charts) (Supplementary Table SII).

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 I). The estimated ICCs for patient age and oocyte provider age were 0.99 and 0.86, respectively, indicating almost perfect agreement (Table II). Among the subset of records with complete information documented on AFC and AMH in both data sources, there was almost perfect agreement between CARTR Plus and the reabstracted data with ICCs greater than 0.90 (Table II). The ICC for FSH level was lower at 0.68 (95% CI: 0.64–0.72), though still in a range indicating strong agreement, and the median difference between the two sources was 0. The kappa coefficient for diminished ovarian reserve as a reason for treatment indicated strong agreement ($\kappa = 0.72$, 95% CI: 0.66, 0.78) and that for advanced female age was moderate ($\kappa = 0.60$, 95% CI: 0.53, 0.67) (Table III). See Supplementary Table SII for complete 2×2 contingency tables for categorical variables. Of note, the PPV for advanced female age was only 0.56 (95% CI: 0.48, 0.63) (Table III), indicating that if the patient was labeled as such in CARTR Plus, there is a 56% probability that she is actually ≥ 35 years of age.

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 I). 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 agreement of 87% (Table III). 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 (Supplementary Table SII).

Embryology

An embryo transfer was performed in 80% of cycles that were not cancelled in both data sources (Table I). Fifty-five percent of these transfers were single embryo transfers (SET) and 42% were double embryo transfers (DET) (Table I). 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 III). The kappa coefficient for elective SET or DET (eSET/eDET) was moderate at 0.55 (95% CI: 0.49, 0.61), with sensitivity, specificity and PPVs greater than 0.75. The NPV 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 II).

Over 90% of embryos were frozen using the vitrification technique after a fresh cycle. Determination of the cryopreservation technique for frozen cycles in CARTR Plus was derived from the data entered for the primary fresh cycle. Approximately 80% of embryos were cryopreserved by vitrification in the frozen embryo transfers (FETs). Eighty-three cycles were missing the method of cryopreservation in the CARTR Plus database (Supplementary Table SII). The weighted kappa coefficients for cryopreservation technique overall were quite strong for both frozen and fresh cycles. Among FET cycles, the percentage 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.

Pregnancy

Thirty-three percent of all initiated cycles and 44% of cycles with an embryo transfer (289 clinical intrauterine pregnancies/655 cycles with embryo transfer) resulted in a clinical intrauterine pregnancy (Table I) according to the reabstracted data. Among these clinical pregnancies, ultrasound assessment detected one fetal sac in 77% 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 III). 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).

Table 1 Description of study variables by data source.

| Variable | CARTR Plus | | | | | Reabstracted Data | | | | | | |
|-----------------------------|------------|------|------|------|--------|-------------------|-----|------|------|------|--------|-------------|
| | 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 | | | | |

(Continued)

Table 1 Continued.

| Variable | CARTR Plus | | | | | | Reabstracted Data | | | | | |
|------------------------------|------------|------|------|------|--------|-------------|-------------------|------|------|------|--------|-------------|
| | N | % | Mean | SD | Median | IQR | N | % | Mean | SD | Median | IQR |
| 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 | | | | |

(Continued)

Table 1 Continued.

| Variable | CARTR Plus | | | | | | Reabstracted Data | | | | | |
|---------------------------------|------------|------|------|------|--------|-------------|-------------------|------|------|------|--------|-------------|
| | N | % | Mean | SD | Median | IQR | N | % | Mean | SD | Median | IQR |
| 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: anti-Müllerian 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, IQR: interquartile range, N: Number or patients
 *AMH levels were converted from pmol/L to ng/dL using a conversion factor of 7.14 (Almog et al., 2011)
 †There was an error in data entry which was recoded as missing

Table II Measures of agreement for continuous variables.

| | N | ICC | 95% CI | Median difference | P value | IQR | Range | % agreement |
|---------------------------------|-----|------|--------------|-------------------|---------|--------|-------------|-------------|
| Dates | | | | | | | | |
| Patient date of birth | 876 | | | 0 days | <0.05 | (0, 0) | (0, 11 688) | 98.9 |
| Cycle start date | 876 | | | 0 days | <0.05 | (0, 0) | (0, 35) | 88.5 |
| Oocyte collection date | 798 | | | 0 days | <0.05 | (0, 0) | (0, 1386) | 95.8 |
| Intake | | | | | | | | |
| Patient age (years) | 876 | 0.99 | (0.99, 0.99) | 0 | <0.05 | (0, 0) | (0, 11) | 96.6 |
| Oocyte provider age (years) | 858 | 0.86 | (0.84, 0.88) | 0 | <0.05 | (0, 0) | (0, 78) | 92.5 |
| FSH (IU/L) | 503 | 0.68 | (0.64, 0.72) | 0 | <0.05 | (0, 0) | (0.0, 45.0) | 64.4 |
| AFC (# follicles) | 342 | 0.92 | (0.91, 0.94) | 0 | <0.05 | (0, 0) | (0, 24) | 62.1 |
| AMH (ng/dL) | 69 | 0.92 | (0.89, 0.95) | 0 | 0.03 | (0, 0) | (0.0, 8.1) | 83.2 |
| Embryo transfer | | | | | | | | |
| ET day | | | | | | | | |
| Fresh cycles (days) | 402 | 0.98 | (0.98, 0.99) | 0 | 0.25 | (0, 0) | (0, 3) | 99.5 |
| Frozen cycles (days) | 229 | 0.00 | | 5 | <0.05 | (5, 6) | (2, 8) | 3.27* |
| Embryology | | | | | | | | |
| # embryos thawed | 244 | 1.00 | (0.99, 1.00) | 0 | 0.25 | (0, 0) | (0, 2) | 99.4 |
| # embryos utilizable after thaw | 244 | 0.93 | (0.92, 0.95) | 0 | <0.05 | (0, 0) | (0, 4) | 96.2 |

ICC: intraclass correlation coefficient

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

Table III 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 |

(Continued)

Table III Continued.

| | κ | 95% CI | SN | 95% CI | SP | 95% CI | PPV | 95% CI | NPV | 95% CI | % agreement |
|-----------------------------|----------|--------------|------|--------------|------|--------------|------|--------------|------|--------------|-------------|
| 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 |

κ : kappa coefficient; NPV: negative predictive value; PPV: positive predictive value; SN: sensitivity; SP: specificity
*Weighted kappa coefficient

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 [Supplementary Table SIII](#) for full results).

Sensitivity analysis—assessment of clinic-specific results

Patient intake

For advanced female age, one of the six clinics (clinic 3) had a percentage agreement lower than 85% ([Table IV](#)). 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 age was reported in the reabstracted data, they were consistently documented as such in CARTR Plus.

FSH levels, AMH levels and AFC demonstrated particularly low agreement in Clinic 2. However, percentage agreement was generally poor for FSH and AFC across all clinics, with estimates ranging from 22.6 to 81.5%.

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 eSET and eDET ([Table V](#)). The percentage agreement for eSET and eDET was much lower at one particular clinic compared to the others ([Table VI](#)). 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.

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, fertility diagnoses and outcomes. Our study, which assessed the data quality within CARTR Plus compared with reabstracted patient chart data, demonstrated that for most of the data elements selected the measures of validity were quite strong. The areas with moderate agreement were FSH levels, reason for treatment cycle, reason for a cancelled cycle, oocyte origin, eSET or eDET and chorionicity.

Missing data

In our study, we identified a minimum of 120 more laboratory test results and ultrasound reports in the chart reabstraction for FSH, AMH and AFC than were recorded in the CARTR Plus database. There was one clinic that did not enter any data into CARTR Plus for any of these test results, despite results being available on the medical charts for many of the patients. However, there was no identifiable trend (geographic location, method of data entry, size of clinic) that

Table IV Sensitivity analysis: percentage 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% |

Table V Prevalence of eSET or eDET by data source.

| eSET or eDET* | CARTR Plus | | Reabstracted data | |
|---------------|------------|------|-------------------|------|
| | 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 |
| DET | n = 273 | | n = 273 | |
| Elective | 138 | 50.6 | 201 | 73.6 |
| Non-elective | 135 | 49.5 | 72 | 26.4 |

Table VI Sensitivity analysis: percentage agreement of eSET and eDET by clinic, cycle type and day of transfer.

| | eSET | eDET |
|---------------------|-------|-------|
| <i>Clinic</i> | | |
| 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% |
| <i>Cycle type</i> | | |
| Fresh | 97.3% | 88.6% |
| Frozen | 86.9% | 89.3% |
| <i>Transfer day</i> | | |
| Day 3 | 97.9% | 79.4% |
| Day 5 | 93.6% | 88.0% |

could account for the missing data among the other clinics. 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 FSH, AMH or AFC in future research projects until the data entry process into CARTR Plus can be clarified and improved.

Misclassification

Advanced female age

As a woman ages, the number of oocytes remaining in her ovaries decreases and the probability of conception diminishes (Broekmans

et al., 2009). The Society of Obstetricians and Gynaecologists of Canada recommends referral to an IVF clinic among women aged 35 years or older after 6 months of trying to conceive (Liu et al., 2011). 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 percentage agreement between the two sources was 82.1%. The discrepancy between the kappa coefficient and percentage 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 percentage 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 (Feinstein and Cicchetti, 1990). 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 (Byrt et al., 1993). Juurlink et al. (2006) argue that in certain cases, PPV and sensitivity are more valuable than the kappa coefficient (Juurlink et al., 2006; Benchimol et al., 2011). With no specific guideline indicating which measure is ideal for reporting and heterogeneity in the literature on the chosen measurement, Benchimol et al. (2011) encourage reporting a minimum of four different measures of validity with corresponding CIs (Benchimol et al., 2011).

Oocyte origin

The overall kappa for oocyte origin and specific kappa coefficients for fresh own oocytes were in the moderate range. However, the other measures of agreement, including percentage agreement, sensitivity, specificity and PPV, were more in keeping with strong agreement between the two data sources. Importantly, 106 treatment cycles (10% of charts) in CARTR Plus were missing information on this element. These missing data were evenly distributed among three clinics, two of which uploaded data directly from an electronic health record system and one that manually input data. The chart reabstraction data indicated that these missing values in the CARTR Plus database were predominantly fresh own oocytes, followed by 'other', which were

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.

eSET

Elective SET or DET 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 ([Committee of the Society for Assisted Reproductive Technology of the American Society for Reproductive Medicine, 2012](#)). The risk of multiple pregnancy after SET is significantly reduced for both cleavage- and blastocyst-stage embryos compared to DET ([Pandian et al., 2013](#)). 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 ([Min and Sylvestre, 2013](#); [Pandian et al., 2013](#); [Peeraer et al., 2014](#); [Greenblatt, 2015](#)).

The Canadian Fertility and Andrology Society, the American Society of Reproductive Medicine and the UK National Institute for Health and Care Excellence 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 ([Committee of the Society for Assisted Reproductive Technology of the American Society for Reproductive Medicine, 2012](#); [Min and Sylvestre, 2013](#); [National Institute for Health and Care Excellence, 2013](#)). 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 percentage 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. 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 birthweight in neonates lower for eSET compared to non-elective SET ([Styer et al., 2016](#); [Mersereau et al., 2017](#)). Based on the error trend we found, the risk of poor pregnancy outcomes with non-elective SET would likely be attenuated if such a study were carried out using the CARTR Plus database, assuming non-differential misclassification ([Armstrong, 1998](#)).

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 ([Canadian Assisted Reproductive Technologies Registry \(CARTR\) Plus, 2017](#)). 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 ([Altman and Bland, 1994](#)). 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 was based on an ultrasound performed in the clinic, at which point it may have been too early to definitively ascertain chorionicity. 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 compared to dichorionicity with respect to intrauterine fetal demise, preterm delivery and placental insufficiency ([Hack et al., 2007](#)). These complications, which are more prevalent in ART pregnancies ([Kanter et al., 2015](#); [Mateizel et al., 2016](#)), may be inappropriately described if not accurately reported in the registry.

Implications for future use

The importance of ensuring validated data when using routinely collected data cannot be understated. When developing policy, such as reducing the rates of multiple gestation after ART by transferring the minimum embryos required to achieve pregnancy, we utilize data from registry and administrative health databases. A low sensitivity in multiple fetal hearts or in chorionicity (with a high false negative rate) could 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 judgment had lower sensitivity, PPV and kappa estimates. Transcription or clerical errors from the chart into the database were less common.

In screening tests for disease, it may be safer to allow a higher false positive rate to maximize sensitivity; the specificity and PPV 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.

For oocyte source, the majority of error was based on missingness, which can largely be attributed to clerical error of failing to enter the information from the chart into the database. Importantly, researchers developing study protocols using ART populations often exclude participants based on the oocyte source, especially oocyte donor cycles. In our study, we found that 13% of patients with missing oocyte source actually were autologous. If these data are used in future research studies, 13% of the population 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 laboratory 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 the 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 AFC, serum FSH level and AMH, may be superior. These algorithms would need to be validated against a reference standard prior to their use.

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 the 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 of 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.

Many studies in ART are now focusing their primary outcome on live birth rates, as this is most relevant to the patient population. While treatment cycle information is collected primarily at the clinics and entered in the CARTR Plus database, the mechanism for ascertaining birth outcomes differs by province and the data in some sites are entered many months after the birth occurs. For example, some clinics contact the patients directly to obtain birth outcome data, while in others these data are obtained from a hospital EMR. In the Ontario clinics, birth outcomes data are automatically populated into the CARTR Plus record via a record linkage with the BORN Ontario provincial birth registry database. At the time when we were designing our study, the processes used nationally for obtaining birth outcome data following IVF treatment cycles were still under development and refinement in the new data system. We therefore opted to restrict our validation to treatment cycle information. Future research to assess the validity of birth outcome data in CARTR Plus should be performed as an extension to this study.

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 the data entry process for a national ART database performed in accordance with recommendations for reporting measures of both validation of administrative databases and diagnostic accuracy studies (Bossuyt *et al.*, 2003; Benchimol *et al.*, 2015).

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 in the USA 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 (Centers for Disease Control and Prevention *et al.*, 2016). As previously described, percentage 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 NPVs and PPVs, were not utilized, thereby making it difficult to interpret the accuracy of the presented information or to compare with our own results (Benchimol *et al.*, 2011, 2015).

Conclusion

In conclusion, our study provides the first assessment of the quality of the data translation process from the patient record to the registry for CARTR Plus. This is also the first evaluation of the validity of data entry 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 that 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.

Supplementary data

Supplementary data are available at *Human Reproduction Open* online.

Acknowledgements

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.

Authors' roles

Vanessa Bacal—primary author. Deshayne Fell—involved in developing the protocol; senior expert in epidemiology; involved in editing the manuscript. Ann Sprague—involved in developing the protocol; involved in editing the manuscript. Andrea Lanes—involved as an expert in CARTR Plus database and epidemiologist; Involved in

developing the protocol; involved in analysis and editing the manuscript. Heather Shapiro—involved as a senior clinical expert; involved in development of the protocol and editing the manuscript. Moya Johnson—involved as an expert in CARTR Plus database; involved with data collection and editing the manuscript. Mark Walker—involved as a senior expert in epidemiology; involved in development of the protocol and editing the manuscript. Laura Gaudet—involved as a senior expert in epidemiology; involved in development of the protocol and editing the manuscript.

Funding

Canadian Institutes of Health Research (CIHR) (FDN-148438); Canadian Fertility and Andrology Society Research Seed Grant (Grant Number: N/A).

Conflict of interest

The authors report no conflict of interest.

References

- Almog B, Shehata F, Suissa S, Holzer H, Shalom-Paz E, La Marca A, Muttukrishna S, Blazar A, Hackett R, Nelson SM *et al*. Age-related normograms of serum antimüllerian hormone levels in a population of infertile women: a multicenter study. *Fertil Steril* 2011;**95**:2359, e1–2363.
- Altman DG, Bland JM. Diagnostic tests 2: predictive values. *BMJ* 1994;**309**:102.
- Armstrong BG. Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup Environ Med* 1998;**55**:651–656.
- Bartfay E, Donner A. Statistical inferences for interobserver agreement studies with nominal outcome data. *J R Stat Soc Ser D Stat* 2001;**50**:135–146.
- Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttman A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol* 2011;**64**:821–829.
- Benchimol EI, Smeeth L, Guttman 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**.
- 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.
- Broekmans FJ, Soules MR, Fauser BC. Ovarian aging: mechanisms and clinical consequences. *Endocr Rev* 2009;**30**:565–593.
- Bushnik T, Cook JL, Yuzpe AA, Tough S, Collins J. Estimating the prevalence of infertility in Canada. *Hum Reprod* 2012;**27**:738–746.
- Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993;**46**:423–429.
- Canadian Assisted Reproductive Technologies Register (CARTR) Plus. Preliminary treatment cycle data for 2014. In: Ottawa, ON, 2015
- Canadian Assisted Reproductive Technologies Registry (CARTR) Plus. Final treatment cycle and pregnancy outcome data for 2015. In: Ottawa, ON, 2017
- Centers for Disease Control and Prevention, American Society for Reproductive Medicine, Society for Assisted Reproductive Technology. 2016 Assisted reproductive technology fertility clinic success rates report. Atlanta (GA):US Dept of Health and Human Services; 2018.
- 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.
- Connolly MP, Hoorens S, Chambers GM. The costs and consequences of assisted reproductive technology: an economic perspective. *Hum Reprod Update* 2010;**16**:603–613.
- Donner A, Eliasziw M. Sample size requirements for reliability studies. *Stat Med* 1987;**6**:441–448.
- Donner A, Zou G. Interval estimation for a difference between intraclass kappa statistics. *Biometrics* 2002;**58**:209–215.
- 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.
- Feinstein AR, Cicchetti DV. High agreement but low kappa: I. the problems of two paradoxes. *J Clin Epidemiol* 1990;**43**:543–549.
- Greenblatt E. *Advisory Process for Infertility Services Key Recommendations Report*. 2015. Available at: http://www.health.gov.on.ca/en/public/programs/ivf/docs/ivf_report.pdf. (26 July 2016, date last accessed)
- Hack K, Derks J, Elias S, Franx A, Roos E, Voerman S, Bode C, Koopman-Esseboom C, Visser G. Increased perinatal mortality and morbidity in monochorionic versus dichorionic twin pregnancies: clinical implications of a large Dutch cohort study. *BJOG* 2007;**115**:58–67.
- 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**:377–381.
- Hierholzer WJ Jr. Health care data, the epidemiologist's sand: comments on the quantity and quality of data. *Am J Med* 1991;**91**:215–265.
- Iron K, Manuel D. Quality assessment of administrative data (QuAAD): an opportunity for enhancing Ontario's health data. *Inst Clin Eval Stud* 2007.
- Juurink D, Preyra C, Croxford R, Chong A, Austin P, Tu J, Laupacis A. Canadian Institute for Health Information discharge abstract database: a validation study. *Inst Clin Eval Stud* 2006.
- Kanter JR, Boulet SL, Kawwass JF, Jamieson DJ, Kissin DM. Trends and correlates of monozygotic twinning after single embryo transfer. *Obstet Gynecol* 2015;**125**:1111–1117.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–174.
- Liu K, Case A, Cheung AP, Sierra S, AlAsiri S, Carranza-Mamane B, Case A, Dwyer C, Graham J, Havelock J *et al*. Advanced reproductive age and fertility. *J Obstet Gynaecol Canada* 2011;**33**:1165–1175.
- Mateizel I, Santos-Ribeiro S, Done E, Van Landuyt L, Van de Velde H, Tournaye H, Verheyen G. Do ARTs affect the incidence of monozygotic twinning? *Hum Reprod* 2016;**31**:2435–2441.

- Maymon R, Shulman A. Controversies and problems in the current management of tubal pregnancy. *Hum Reprod Update* 1996;**2**:541–551.
- 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**:750–756.
- Min J, Sylvestre C. Guidelines on the number of embryos transferred. *Can Fertil Androl Soc* 2013. Available at: http://www.cfas.ca/images/stories/pdf/cfas_cpg_embryo_transfer_2013.pdf (25 July 2016, date last accessed)
- 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> (27 July 2016, date last accessed)
- Norris S. Reproductive infertility: prevalence, causes, trends and treatments. *Parliam Res Branch Libr Parliam* 2001. Available at: <http://publications.gc.ca/collections/Collection-R/LoPBdP/EB-e/prb0032-e.pdf> (20 December 2016, date last accessed)
- 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**:485–503.
- 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.
- Peeraer K, Debrock S, Laenen A, De Loecker P, Spiessens C, De Neubourg D, D'Hooghe TM. The impact of legally restricted embryo transfer and reimbursement policy on cumulative delivery rate after treatment with assisted reproduction technology. *Hum Reprod* 2014;**29**:267–275.
- Public Health Agency of Canada. Perinatal health indicators for Canada 2017. 2017.
- 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.
- 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.
- Sim J, Wright C. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy* 2005;**85**:257–268.
- Styer AK, Luke B, Vitek W, Christianson MS, Baker VL, Christy AY, Polotsky AJ. Factors associated with the use of elective single-embryo transfer and pregnancy outcomes in the United States, 2004–2012. *Fertil Steril* 2016;**106**:80–89.
- Zou G. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. *Stat Med* 2012;**31**:3972–3981.