



Published in final edited form as:

*J Mod Appl Stat Methods*. 2017 ; 16(1): 744–752. doi:10.22237/jmasm/1493599140.

## Using multiple imputation to address the inconsistent distribution of a controlling variable when modeling an infrequent outcome

Yujia Zhang<sup>a</sup>, Sara Crawford<sup>a</sup>, Sheree L. Boulet<sup>a</sup>, Michael Monsour<sup>a</sup>, Bruce Cohen<sup>b</sup>, Patricia McKane<sup>c</sup>, and Karen Freeman<sup>d</sup>

<sup>a</sup>Division of Reproductive Health, Centers for Disease Control and Prevention, Atlanta, GA

<sup>b</sup>Massachusetts Department of Public Health, Boston

<sup>c</sup>Michigan Department of Community Health, Lansing

<sup>d</sup>Florida Department of Health, Tallahassee

### Abstract

Temporal changes in methods for collecting longitudinal data can generate inconsistent distributions of affected variables, but effects on parameter estimates have not been well described. We examined differences in Apgar scores of infants born in 2000-2006 to women with ovulatory dysfunction (risk) or tubal obstruction (reference) who underwent assisted reproductive technology (ART), using Florida, Massachusetts, and Michigan birth certificate data linked to the Centers for Disease Control and Prevention's National ART Surveillance System database. Florida had inconsistent information on induction of labor (a control variable) from a 2004 change in birth certificate format. Because we wanted to control for bias that may be introduced by the inconsistent distribution of labor induction in analysis, we used multiple imputation data in analysis. We used Cox-Iannacchione weighted sequential hot deck method to conduct multiple imputation for the labor induction values in Florida data collected before this change, and missing values in Florida data collected after the change and overall Massachusetts and Michigan data. The adjusted odds ratios for low Apgar score were 1.94 (95% confidence interval [CI] 1.32-2.85) using imputed induction of labor and 1.83 (95% CI 1.20-2.80) using not imputed induction of labor. Compared with the estimate from multiple imputation, the estimate obtained using not imputed induction of labor was biased towards the null with inflated standard errors, but the magnitude of differences was small.

---

Corresponding Author: Dr Yujia Zhang, 4770 Buford Highway NE, MS F74, Atlanta, GA 30341-3717, USA Telephone: 770-488-6342, Fax: 770-488-6354, E-mail: YZhang2@cdc.gov.

**Disclosures:** Y.Z. has nothing to disclose. S.C. has nothing to disclose. S.B. has nothing to disclose. M.M. has nothing to disclose. B.C. has nothing to disclose., P.M. has nothing to disclose. K.F. has nothing to disclose.

**Disclaimer:** The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

## Keywords

assisted reproductive technology; inconsistent data distribution; multiple imputation; weighted sequential hot deck

---

## Introduction

Maternal and child health research often uses population-based longitudinal data; however, some information collected may not be consistent over the entire study period due to changes in data collection procedures. Affected variables may experience a change in the measured prevalence before and after the changes in data collection. The impact of inconsistent information in analytic data on parameter estimation has not been well described. Traditionally, researchers simply use values for the problematic variable(s) as collected, essentially ignoring the change in data collection. However, inconsistent data distributions may impact parameter estimates and variance estimation. To control for bias that may be introduced by the inconsistent data distribution, an alternative strategy is to use imputation to correct the inconsistent information in the data collected before or after the change. One method of imputation fills each missing entry with an imputed value, such that standard complete-data methods can be used for analysis. However, this method ignores the variability contributed by the lack of information on the missing values, leading to variance underestimation. A second method, multiple imputation, replaces each missing entry with two or more values and draws inferences by combining the results of several complete-data analyses to address within and between-imputation variability in variance estimation [1-3].

The aim of the present study is to compare the use of the traditional method of ignoring the change in data distribution with multiple imputation when inconsistent information is found for a controlling variable due to temporal changes in data collection. The results obtained from the logistic regression model using the multiply imputed data were used as a benchmark when assessing bias and changes in variability.

## Materials and methods

### Data source

We used a population-based, historical dataset collected from multiple sources for the analyses. The Division of Reproductive Health at the Centers for Disease Control and Prevention (CDC) maintains a nationwide registry of Assisted Reproductive Technology (ART) cycles, the National ART Surveillance System (NASS), for procedures performed annually in the U.S [4]. The NASS contains detailed information on patient obstetric history, reasons for using ART, and the ART procedure itself, but only limited information on patient demographics and resultant births. More detailed information on maternal characteristics, pregnancy and delivery complications, and pregnancy outcomes can be found in state-based vital records systems.

The Division of Reproductive Health and the state public health departments (the members at the time the study was conducted were Connecticut, Florida, Massachusetts, and Michigan) have created a collaborative known as States Monitoring Assisted Reproductive

Technology (SMART) [5] to establish, evaluate, improve, and promote state-based surveillance of ART. The collaborative links the NASS data to state vital records using a probabilistic linkage algorithm [6] creating a multistate dataset which provides rich information on ART and maternal and infant health outcomes. Currently, CDC has received data from three states (as Connecticut is a new member). The linked data can be used to monitor and examine ART pregnancy outcomes and to compare ART conceived infants and their mothers to the general population. Because the linked datasets have a large sample size, they allow researchers to address issues with relatively infrequent outcomes.

### Study population

Linked NASS data and live birth records from Florida, Massachusetts, and Michigan from 2000 to 2006 were used for analyses. Only the first live born infant of the first live birth was included if a woman was identified as having more than one birth in the study time period to eliminate the potential impact of subsequent treatments on maternal complications and pregnancy outcomes, yielding 16,876 eligible infants. Detailed information on the demographics of the study population was previously published [7]. The Apgar score at five minutes post-delivery is recorded on state live birth files; the main outcome of interest for our analysis was an Apgar score (at five minutes) of less than seven. Apgar score at five minutes is a score of the very first test given to a newborn to quickly evaluate a newborn's physical condition with a range one to ten. Score values of seven and above generally are considered normal. Since 43 infants had an unknown Apgar score at five minutes, 16,833 infants were used for analyses. The primary risk factor of interest was infertility diagnosis (ovulatory dysfunction only, tubal obstruction only, and other infertility reasons), which was identified using the NASS variable reason for ART. We were primarily interested in comparing women with ovulatory dysfunction only to women with tubal obstruction only, but we also controlled for maternal age, race/ethnicity, education, adequacy of prenatal care, co-morbid conditions, delivery method, induction of labor, gestational age, and newborn gender plus a composite variable of gestational age and birth weight. There were 97 cases of infants with an Apgar score of less than seven found in ovulatory dysfunction only and tubal obstruction only in all three states, of which 35 cases were in Florida. A more detailed description of the controlling variables can be found in the previously published paper [7].

### Analytic datasets

Both Michigan and Massachusetts used the U.S. Standard Certificate of Live Birth Revised 1/1989 format to collect live birth records for the entire study period. Florida used the U.S. Standard Certificate of Live Birth Revised 1/1989 through February of 2004, but then switched to U.S. Standard Certificate of Live Birth Revised 11/2003 from the National Center for Health Statistics. Of all the variables used for this analysis, only one variable, labor induction (indicating whether using artificially stimulating childbirth in delivery), was affected by this change in the birth certificate format. The old birth certificate, associated with more than 50.0% of Florida data, has checkboxes to indicate use of specific obstetric procedures, including induction of labor. There is also a checkbox to indicate that none of the procedures were used, but no checkboxes to indicate specifically which procedures had not been used. For purposes of coding for vital statistics, if induction of labor was not indicated on the birth certificate it was assumed that labor was not induced, unless nothing

was marked in the section on obstetric procedures, in which case the data were considered not classifiable or missing. Under the new birth certificate format, induction of labor is coded as 'Yes' (induction of labor), 'No' (no induction of labor) and 'Unknown' (missing). This change resulted in different distributions for the induction of labor variable, since the new format allowed one to directly distinguish between missing and 'No' values, while the old format did not (if 'Yes' was not indicated, 'No' was assumed unless the entire section was left blank [missing]).

To address the impact of this change in the birth certificate on parameter estimates using traditional and multiple imputation approaches, we created a dataset with imputed values for the induction of labor variable in addition to the not imputed dataset, which retained all the induction of labor values in the Florida data as originally collected with the known change in distribution. The data to be imputed was separated into two subsets, Florida only and the other two states (Massachusetts and Michigan). For Florida we used multiple imputation to replace all values for induction of labor that were collected prior to the change in birth certificate format, regardless of whether induction of labor was indicated or not because keeping 'Yes' values of induction of labor obtained before the change would change the distribution used for imputation, resulting a uncorrected distribution in imputed data, and also to replace the missing values in the data collected with the new format. The imputed Massachusetts and Michigan subset retained the original non-missing data for induction of labor, but multiple imputation was applied to replace missing values.

## Imputation

The multiple imputation was performed using SUDAAN's HOTDECK procedure (SUDAAN Release 11, RTI International, Research Triangle Park, North Carolina). SUDAAN was developed to analyze survey data. However, SUDAAN is also able to analyze non-survey data by setting the strata value in the statement 'NEST' as `_ONE_` and the DESIGN option in PROC statement as WR (sampling with replacement for population data). SUDAAN's procedures for complex survey data structures can be used to address nested data distributions in non-survey data. In our data, ART patients were nested in fertility clinics and we imputed by clinic through setting the HOTDECK's 'IMPBY' statement to clinic.

Since there was a large variation in clinic size, each missing entry in a small fertility clinic represents a larger percent of total cycles within the clinic for missing than a missing entry in a large fertility clinic subject to the smaller denominator. This impact on multiple imputation can be addressed using a weighted approach, i.e., assigning a larger weight to a missing entry for a small fertility clinic and a smaller weight to a missing entry for a large fertility clinic. Weights for a missing entry within clinic can be computed using SUDAAN's WTADJUST procedure by setting the strata value in the statement 'NEST' as `_ONE_`, the next nest variable as clinic, the DESIGN option as WR and the ADJUST option in PROC WTADJUST as "NONRESPONSE." The dependent variable in the weight computing model is a binary variable with value one for non-missing entries and value zero for missing entries. The independent variables are those used in the analytic model. Shown below are the basic SUDAAN statements used to computing the weights:

```

PROC WTADJUST DATA=ONE DESIGN=WR ADJUST=NONRESPONSE;

WEIGHT _ONE_;

NEST _ONE_ CLINIC;

CLASS AGE REASON_FOR_ART SEX...;

MODEL RESPONSE=AGE REASON_FOR_ART SEX ...;

```

The HOTDECK procedure identifies all infants, by fertility clinic, with a missing value for induction of labor. For each infant with a missing value, a set of similar infants from the same clinic is collected, i.e., infants with characteristics similar to the infant with the missing value (as specified by the variables used to fit the adjusted model), but with an observed value for induction of labor. An infant is randomly drawn from this set of infants and the observed labor induction value for that infant is assigned in place of the missing value. The process is repeated until all missing values for the induction of labor variable within the clinic are imputed. SUDAAN's HOTDECK procedure uses a weighted sequential hot deck method proposed by Cox [8] and Iannacchione [9] (Cox-Iannacchione WSHD method) to perform imputation. For survey data, this method uses the sample weights in the imputation process to ensure the weighted distribution of imputation revised data over clusters is preserved. For our non-survey data, this method uses weights based on the number of missing values in imputation to address the impact of uneven missing value distributions between small and large fertility clinics. In addition, this method limits the number of times of using the same infant with an observed value, depending on the infant's non-respondent weight, to ensure the infant is not being used excessively in the imputation process.

For Florida data, all collected induction of labor values with the old birth certificate format were set to missing and we used the induction of labor values collected using the new birth certificate format to impute those collected using the old format, and also used this approach to impute the missing values that occurred after the change. We chose to impute the old values since the Florida data will continue to be collected in the new format, the new coding mechanism was more complete, and we expect additional states to transition to the new format. For the other two states, since they didn't adopt the new format in the study period we used the existing data distribution to impute missing values. Multiple imputation requires using at least two datasets with imputed values, and the parameter estimates then are averaged over a predictive distribution for missing data. We imputed five times [10] by setting the option, MULTIMP = 5 in the IMPVAR statement and SUDAAN outputted all five imputation results in one dataset. Shown below are the basic SUDAAN statements used to impute missing values of induction of labor for Florida:

```

PROC HOTDECK DATA=DATA_FL SEED=3123845 NOTSORTED;

WEIGHT WTFINAL;

IMPBY SITE;

```

```

IMPVAR INDUCTION_LABOR AGE REASON_FOR_ART .../MULTIMP=5;

IMPID INFANT_ID;

IMPNAME IND_LABOR="IND_LABOR_IMP" AGE="AGE_IMP" REASON_FOR_ART =
"REASON_FOR_ART _IMP" ...;

IDVAR INDUCTION_LABOR AGE REASON_FOR_ART ...;

OUTPUT /IMPUTE=default FILENAME=OUTDATAFL REPLACE;

```

We found that the imputed dataset had 114 fewer observations (< 1.0% of the total eligible infants) due to imputation failure resulting from missing labor induction values for all infants in a few clinics.

### Statistical analysis

We examined the distribution of labor induction for the not imputed and imputed datasets in order to verify that the distribution for the imputed values was comparable to the distribution of labor induction for the not imputed dataset. Also, we fit a random effects logistic regression model for the outcome of a five minute Apgar score less than seven to both datasets, where the primary risk factor of interest was the infertility diagnosis. We adjusted for maternal age, race/ethnicity, education, adequacy of prenatal care, co-morbid conditions, delivery method, induction of labor, plurality, gestational age, newborn gender, and the composite variable of gestational age and birth weight as fixed effects in the model, and included clinic as a random effect to account for clustering. The parameter estimates, odds ratios, standard errors, and confidence intervals for the risk factor of interest obtained using the imputed data were compared to those obtained using the not imputed data in order to assess bias and variability. Since very limited cases (35) were found in the interested levels of the risk factor for Florida, we didn't perform a stratified analysis restricting to Florida only because the estimates are more likely biased subject to the small sample size in the logistic modeling [11-12]. Institutional Review Boards of CDC, Florida Department of Health, Massachusetts Department of Public Health, and Michigan Department of Community Health approved the original project [7].

### Results

Table 1 shows the distribution of the induction of labor variable in Florida before and after the adoption of the latest national standard birth certificate. The percent of all infants born to mothers with an induced labor increased from 19.4% before the change (January 2000 - February 2004) to 24.0% after the change (March 2004 - December 2006), demonstrating an inconsistency in the distribution. A year-by-year exploration indicated that from 2000 to 2003 the percent of infants born via induction was consistent over time (19.1% for 2000; 19.6% for 2001; 19.8% for 2002; 19.1% for 2003), but there was a notable increase from 19.4% in January-February of 2004 to 24.4% in March-December of 2004 and consistently higher in successive years (24.2% for 2005; 23.7 for 2006). A similar pattern was seen when restricting the data to ART-conceived, live-born, first-delivery, first-born infants, with the

percent increasing from 16.5% before the change to 21.9% after the change. The distribution of ART-conceived infants born after induction of labor, based on imputed values for the period prior to the birth certificate change, had a distribution similar to that of the observed data after the change (20.4% and 21.9%, respectively). We also examined the distribution of the induction of labor variable for the eligible ART conceived infants in the other two states within the same period in the original data. The prevalence was consistent across years with an average of 17.8% for Massachusetts and 12.2% for Michigan. Table 2 shows the odds ratios and confidence intervals (CI) for the unadjusted and adjusted models. The crude odds ratio (OR) of the ovulatory dysfunction only group as compared to the tubal obstruction only group was 1.86 (95% CI: 1.31-2.63, p-value < 0.001). The adjusted OR (aOR) was 1.83 (95% CI: 1.20-2.80, p-value = 0.005) using the not imputed dataset; and 1.94 (95% CI: 1.32-2.85, p-value = 0.001) using the imputed dataset. Finally, we repeated the analysis using multiply imputed data, but retained original Florida values that indicated induction of labor was used; results were identical to the prior analysis with imputed data.

## Discussion

When working with longitudinal, multi-state data, inconsistent information for one or more variables is likely to occur due to changes in variable collection methods over time. In our study, Florida adopted a change to their birth certificate in the middle of the study period, resulting in an inconsistent distribution for a control variable, induction of labor, over time. We examined the distribution of labor induction in Florida birth certificate data, and saw an appreciable increase in measured prevalence after the change. However, we observed a consistent prevalence of induction of labor during the entire study period as measured by Massachusetts and Michigan birth certificate data, indicating that the increase in prevalence detected for Florida birth certificate data was likely an artifact of the change in birth certificate format. Because we wanted to control for bias that may be introduced by the inconsistent distribution of labor induction in analysis, in addition to the collected data we tested multiple imputed data in analysis.

Of the two methods used, we believe the regression results obtained using the imputed dataset are more appropriate since they are derived from a dataset with statewide consistent distributions of labor induction over time and negligible missing values (<1.0% missing). While the approach using the not imputed data does not have a large number of missing values (3.1% missing), it ignores the change in data collection, and maintains an inconsistent distribution in labor induction over time caused by the possibly misclassified non-inductions in the Florida data collected before the change. The inconsistency of the distribution likely biases parameter estimation. Because we determined that the imputation approach was more appropriate, we compared the multivariable regression results from the non-imputed data to the results using multiple imputation, treating the multiple imputation approach as a benchmark when assessing bias and changes in variability.

The parameter estimate for ovulatory dysfunction in the not imputed dataset was smaller than that for the imputed dataset, and the standard error was larger, yielding a wider confidence interval for the aOR and a larger p-value. This change may reflect the impact of the inconsistency in the induction of labor distribution before and after the change in Florida



data. Induction of labor is often used in the research of obstetrical outcomes [13-14]. However in this study, while the estimate obtained using multiple imputed data differed from that obtained using the original data, the magnitude of differences was small and confidence intervals obtained using two approaches were mostly overlapped thus it does not seem that any claim about significant differences from the results with or without imputation can be made, though the multiple imputation was correctly performed and the distribution of induction of labor was statewide consistent.

This study has several strengths. First, the dataset used for analysis was a long term, multistate, population-based large-sample dataset, ensuring enough cases to analyze an infrequent outcome. Second, we accounted for the clustered nature of the design when performing the imputation and fitting the logistic model. Especially in this study, we addressed the impact of large size variations among fertility clinics in performing multiple imputation using a weighted approach. However, this was not done for the previous published study such that the presented estimates of the primary risk factor in this paper are slightly different from those previously published (aOR=1.94, 95% CI: 1.32-2.85 versus aOR=1.90, 95% CI=1.30-2.77).

Nevertheless, the study also has limitations. First, a gold standard is lacking in this study because the induction of labor values comparable to the new birth certificate format are unknown for the earlier years used in our analyses. We investigated the possibility of obtaining a gold standard from Florida hospital discharge data, but found that induction of labor was coded in a way similar to the old birth certificate form; therefore no gold standard was available for comparison. Second, both Massachusetts and Michigan did not adopt the new birth certificate form in the study period. Including their data in the analysis may have altered the presence or direction of bias or changes in variability. However, performing a sub-analysis with only Florida data would provide very limited cases within interested categories of infertility diagnosis, and likely introduce small sample bias in fitting the logistic model. The third limitation resulted from the imputation algorithm employed. Because some clusters contained no responders (i.e., all values for induction of labor were missing), the imputation failed for those clusters, resulting in the imputed data having a small portion of missing values (<1.0%).

In summary, we used multiple imputation to address an inconsistency in the distribution of a control variable for a multivariable model. In comparison, results from the traditional method appear to introduce bias toward the null value and increase variability. As more states adopt the latest national standard birth certificate, we expect inconsistent information for one or more variables to occur in other states, and propose using multiple imputation to address this issue when it occurs.

## References

1. Rubin DB. Multiple imputation for nonresponse in surveys. J Wiley and Sons; New York: 1987.
2. Schafer JL. Multiple imputation: a primer. Statistical Methods in Medical Research. 1999; 8:3–15. [PubMed: 10347857]
3. Rubin DB. Multiple imputation after 18+ years. Journal of the American Statistical Association. 1996; 91:473–489.



4. Centers for Disease Control and Prevention. 2010 Assisted Reproductive Technology Success Rate. Atlanta, GA: Centers for Disease Control and Prevention; 2012. American Society for Reproductive Medicine, Society for Assisted Reproductive Technology.
5. Mneimneh A, Boulet S, Sunderam S, et al. Report from the CDC, states monitoring assisted reproductive technology (SMART) collaborative: data collection, linkage, dissemination, and use. *Journal of Women's Health*. 2013; 22:571–577.
6. Zhang Y, Cohen B, Macaluso M, et al. Probabilistic linkage of assisted reproductive technology information with vital records, Massachusetts 1997–2000. *Maternal and Child Health Journal*. 2012; 16:1703–1708. [PubMed: 21909704]
7. Grigorescu V, Zhang Y, Kissin D, et al. Maternal characteristics and pregnancy outcomes after assisted reproductive technology (ART) by infertility diagnosis: ovulatory dysfunction (OD) versus tubal obstruction (TO). *Fertil Steril*. 2014; 101(4):1019–25. [PubMed: 24484993]
8. Cox B. The Weighted Sequential Hot Deck Imputation Procedure. [www.amstat.org/sections/srms/proceedings/papers/1980\\_152.pdf](http://www.amstat.org/sections/srms/proceedings/papers/1980_152.pdf)
9. Iannacchione V. Weighted Sequential Hot Deck Imputation Macros. Seventh Annual SAS User's Group International Conference; San Francisco CA. February 1982; 1982.
10. Sinharay S, Stern H, Russell D. The Use of Multiple Imputation for the Analysis of Missing Data. *Psychological Methods*. 2001; 6:317–329.
11. Rotnitzky A, Wypij D. A Note on the Bias of Estimators with Missing Data. *Biometrics*. 1994; 50:1163–1170. [PubMed: 7786998]
12. King G, Zeng L. Logistic regression in rare events data. *Political Analysis*. 2001; 9:137–163.
13. Aghideh F, Mullin P, Ingles S, et al. A comparison of obstetrical outcomes with labor induction agents used at term. *J Matern Fetal Neonatal Med*. 2014; 27(6):592–596. [PubMed: 23919802]
14. Cheng Y, Kaimal A, Snowden J, et al. Induction of labor compared to expectant management in low-risk women and associated perinatal outcomes. *Am J Obstet Gynecol*. 2012; 207(6):502.e1–502.e8. [PubMed: 23063017]

Table 1

The number and percent of Florida infants (overall and ART conceived) born in 2000 – 2006 after induction of labor, as measured by Florida birth certificate data and imputed data before and after adopting the new birth certificate format.

|                     | All Infants    | ART-conceived Infants <sup>a</sup> | ART-conceived Infants <sup>a</sup> (imputed induction of labor) |
|---------------------|----------------|------------------------------------|---|
| Year                | N (%)          | N (%)                              | N (%)   |
| Jan 2000-Feb 2004 * | 167,944 (19.4) | 450 (16.5)                         | 575 (20.5)  |
| Mar 2004-Dec 2006   | 156,582 (24.0) | 486 (21.9)                         | 486 (21.9)  |

\* Florida adopted the new format of birth certificate on 3/01/2004.

<sup>a</sup> Index birth includes all ART-conceived, live-born, first-delivery, first-born infants.

Crude (unadjusted) odds ratio and adjusted odds ratio (aOR) derived from logistic regression models assessing the association between five minute Apgar scores less than seven and mother's infertility diagnosis: Florida, Massachusetts, and Michigan births, 2000-2006..

Table 2

| Infertility diagnosis      | Crude OR (95% CI) P value | Not Imputed Data <sup>a</sup> aOR (95% CI) P value | Imputed Data <sup>b</sup> aOR (95% CI) P value |
|----------------------------|---------------------------|--|--|
| Tubal Obstruction Only     | Ref                       | Ref  | Ref  |
| Ovulatory Dysfunction Only | 1.86 (1.31-2.63) <0.001   | 1.83 (1.20-2.80) 0.005                             | 1.94 (1.32-2.85) 0.001                         |
| Other                      | 1.20 (0.85-1.69) 0.297    | 1.30 (0.89-1.89) 0.180                             | 1.35 (0.91-1.99) 0.131                         |

<sup>a</sup>The not imputed dataset contains all observations from 2000-2006 for all three states as originally collected, ignoring the change in the data collection for the Florida induction of labor data that occurred in 2004.

<sup>b</sup>The imputed dataset contains all observations from 2000-2006 for all three states; however the labor of induction values for Florida from 2000-February 2004 and other missing values for induction of labor from Florida, Massachusetts, and Michigan have been imputed.