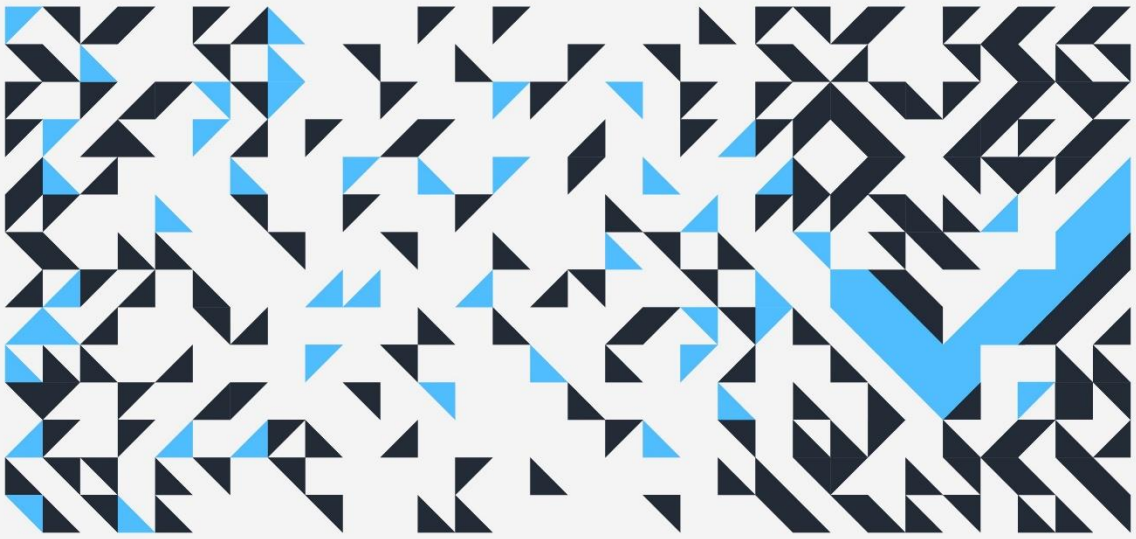
MILLIMAN REPORT

Methodology Review of Stork Club's Assisted Reproductive Technology Performance Measure Comparison Study

MARCH 15, 2024

Harsha Mirchandani, FSA, MAAA
Brittany Ryan, MHP
Carol Bazell, MD, MPH

Commissioned by Stork Club



Milliman

Executive Summary

SCOPE AND PURPOSE

Stork Club is a company that provides reproductive healthcare-related programs for employers with self-funded health plans. Its programs and services address reproductive wellness, fertility cryopreservation, conception, pregnancy, birth, postpartum recovery, newborn care, and back-to-work following pregnancy. Stork Club's services integrate with an employer's health plan as either a core part of the health benefits package or as an optional supplemental benefit. For contracted employers, employees (broadly-termed Stork Club members) may interact with Stork Club in a variety of ways around questions and concerns regarding reproductive health, including through technology platforms and trained care navigators. In some cases, this may result in a referral to Stork Club's contracted fertility clinics (i.e., in-network fertility clinics). After clinic consultation and employee acceptance of the clinic's fertility treatment plan, some individuals undergo assisted reproductive technology (ART) with embryo transfer at these in-network fertility clinics. ART encompasses all fertility treatments where eggs or embryos are handled and does not include treatments in which only sperm are handled or procedures in which a woman takes medicine only to stimulate egg production without the intention of having eggs retrieved.[1] Stork Club collects information from in-network fertility clinics about embryo transfers provided to Stork Club members as part of the Stork Club Outcome Initiative program.

Stork Club's ART performance measure comparison study (the "Stork Club Case Study") methodology analyzes ART performance measures for Stork Club members who underwent embryo transfers and compares them to national metrics for individuals receiving embryo transfers. Milliman was commissioned by Stork Club to assess the appropriateness of the following Stork Club Case Study methodological elements:

- Selection of ART performance measures
- Data collection, sample size, and time periods used in the analysis of ART performance measures
- Selection of populations for comparison and age-adjustment of measure results
- Consistency in the methodology used to calculate the measures for each population

The Stork Club Case Study reported on common performance measures for assessing ART[2]:

- Pregnancy rate per transfer
- Miscarriage rate per pregnancy
- Transfers per live birth
- Live birth rate per transfer
- Single Embryo Transfer (SET) rate per transfer
- Multiples rate per transfer

For our review, we focused on five measures, and have excluded 'Transfers per Live Birth' from our discussion and overall findings as it is the inverse of another included measure, 'Live Birth Rate (per transfer)', and is therefore duplicative.

The Stork Club Case Study compares the following two populations, using data from the sources described below:

1. Stork Club members who receive ART services at Stork Club in-network fertility clinics. Data is submitted by Stork Club's in-network fertility centers.
2. Information about all individuals who receive ART services at fertility clinics is collected in the Centers for Disease Control and Prevention (CDC) ART data and Society for Assisted Reproductive Technology (SART) data. These data are drawn from national publicly available CDC and SART sources.

Our review was limited to the Stork Club Case Study methodology, and we have not reviewed Stork Club data or calculations or otherwise checked the results of the study. We are not offering any opinion on the accuracy or the validity of results now or in any future use of the methodology that we reviewed. The scope of our work does not include review of any other measures or studies Stork Club uses to assess its programs, such as its clinical or

[1] Centers for Disease Control and Prevention. What is Assisted Reproductive Technology? [Internet]. [cited 2024 Mar 8]. Available from: https://www.cdc.gov/art/whatis.html

[2] Pirtea P, de Ziegler D, Poulain M, Ayoubi JM. Which key performance indicators are optimal to assess clinical management of assisted reproduction cycles? Fertil Steril. 2020 Jul;114(1):24-30. doi: 10.1016/j.fertnstert.2020.04.055. Epub 2020 Jun 10. PMID: 32532485.

financial performance. Nor does it consider or comment on the potential relationship of Stork Club's ART performance measure results to the features of its product offering.

SUMMARY OF FINDINGS

From our review carried out in 2024, based on the current distribution of Stork Club members by age group and composition of in-network clinics, we conclude that Stork Club's ART performance measure comparison methodology is appropriate for three ART performance measures – pregnancy rate per transfer, live birth rate per transfer, and SET rate per transfer. The methodology for two measures – miscarriage rate per pregnancy and multiples rate per transfer – relies on having a larger sample size to produce reliable results because the rates for these measures approach extreme values. The absence of sufficient sample size increases the probability of favorable results due to random chance instead of meaningful differences in Stork Club performance compared to national averages.

The selected measures, formulas, and calculations used to develop Stork Club and national averages are appropriate for developing and comparing ART performance measure results for these populations. They are also consistent with typical industry practice. The methodology may no longer be appropriate if there are substantial changes in the age distribution of Stork Club members undergoing embryo transfers, composition of in-network clinics, or in the percentage of Stork Club embryo transfers with complete information reported. As the Stork Club Case Study is limited to comparing Stork Club's in-network fertility clinic embryo transfers to national averages, the methodology does not account for network selection, which would isolate the impact of Stork Club's program design on the ART performance measure results.

Our review considered Stork Club data collection process and time periods used in the analysis, the selection of populations and normalization of performance measure results, and the consistency in the methodology used to compare ART performance measure results for Stork Club member and national populations undergoing embryo transfers. Stork Club provided us with descriptive information about its services and the formulas and calculations used in the Stork Club Case Study, as well as statistical testing for comparisons of measure results.

The purpose of this report is to summarize our review of the Stork Club Case Study methodology. It should not be used for any other purpose. This report contains important caveats and limitations of our work and should be read in its entirety by qualified professionals.

This information does not constitute an endorsement or recommendation of Stork Club's services in aggregate or for any specific group or individual, historically or in the future. Any reader of this report must possess a certain level of expertise in areas relevant to this analysis to evaluate the significance and reasonability of assumptions and the impact of these assumptions on Stork Club's methodology and results. Milliman recommends that third parties be aided by qualified professionals when reviewing this report.

Professional standards, limitations, distribution, and use

Harsha Mirchandani is a senior consulting actuary with the firm of Milliman, Inc., a member of the American Academy of Actuaries, and meets the Qualification Standards of the Academy to render the opinion contained herein. To the best of our knowledge and belief, this report is complete and accurate and has been prepared in accordance with generally recognized and accepted actuarial principles and practices. The three listed co-authors of this report are employed by Milliman and collaborated on the methodology review. Plural pronouns ("we" and "our") are used throughout the report to reflect our collective conclusions.

In performing our review, we relied on calculations and other information provided by Stork Club. We have not reviewed, audited, or verified the data used in these calculations, or any other information provided to us by Stork Club. Such a review was beyond the scope of our engagement. If the underlying data or information is inaccurate or incomplete, the conclusions presented herein may likewise be inaccurate or incomplete.

No reliance should be placed on any advice not given in writing, or on draft versions of our figures, reports, or other forms of written communication.

Our review was limited to the Stork Club Case Study methodology that was provided to us in 2024. We are not commenting on ART performance measure results that may be achieved by any specific fertility clinic contracting with Stork Club, either historically or in the future.

Future experience will differ from the ART performance measure results we reviewed for many reasons, including, but not limited to member mix characteristics, national reporting standards and trends, and Stork Club data collection practices, as well as other random and non-random factors. It is important that actual experience be monitored and that appropriate adjustments be made to the methodologies on a regular basis to ensure they remain appropriate. It is certain that actual experience will vary from expected, perhaps materially.

This report should be reviewed in its entirety.

Background

OVERVIEW OF THE STORK CLUB PROGRAM

Stork Club is a company that provides reproductive healthcare-related programs for employers with self-funded health plans. Its programs and services address reproductive wellness, fertility cryopreservation, conception, pregnancy, birth, postpartum recovery, newborn care, and back-to-work following pregnancy. Stork Club provides a range of solutions such as personalized prenatal care, educational resources, and technology platforms that connect individuals with care coordinators and healthcare professionals. Its services integrate with an employer's health plan as either a core part of the health benefits package or as an optional supplemental benefit.

For contracted employers, employees (broadly termed Stork Club members) may interact with Stork Club in a variety of ways around questions and concerns regarding reproductive health. In some cases, engagement with Stork Club may result in a referral to Stork Club's contracted fertility clinics (i.e., in-network fertility clinics). As of February 2024, Stork Club had 40 in-network clinics across 23 states. After clinic consultation and employee acceptance of the clinic's fertility treatment plan, some individuals undergo assisted reproductive technology (ART) services with embryo transfer at these in-network fertility clinics. ART encompasses all fertility treatments where eggs or embryos are handled and does not include treatments in which only sperm are handled or procedures in which a woman takes medicine only to stimulate egg production without the intention of having eggs retrieved.[3] Stork Club collects information from in-network fertility clinics about embryo transfers provided to Stork Club members as part of the Stork Club Outcome Initiative program.

PURPOSE

Stork Club's ART performance measure comparison study (the "Stork Club Case Study") methodology analyzes ART performance measures for Stork Club members who underwent embryo transfers and compares them to national metrics for individuals receiving embryo transfers. Milliman was commissioned by Stork Club to review the Stork Club Case Study methodology, which includes assessing the appropriateness of measure selection, data collection,

[3] Centers for Disease Control and Prevention. What is Assisted Reproductive Technology? [Internet]. [cited 2024 Mar 8]. Available from: https://www.cdc.gov/art/whatis.html

sample size, and time periods, population selection and age-adjustment, and individual measure components. The Stork Club Case Study reported on common performance measures for assessing ART.[4]

- Pregnancy rate per transfer
- Miscarriage rate per pregnancy
- Transfers per live birth
- Live birth rate per transfer
- Single Embryo Transfer (SET) rate per transfer
- Multiples rate per transfer

For our review, we focused on five measures, and have excluded 'Transfers per Live Birth' from our discussion and overall findings as it is the inverse of another included measure, 'Live Birth Rate (per transfer)', and is therefore duplicative.

The Stork Club Case Study methodology compares specific ART performance measures for the following two populations, using data from the sources described below:

1. Stork Club members who receive ART services at Stork Club in-network fertility clinics. Data is submitted by Stork Club's in-network fertility centers.
2. Information about all individuals who receive ART services at fertility clinics is collected in the Centers for Disease Control and Prevention (CDC) ART data and Society for Assisted Reproductive Technology (SART) data. These data are drawn from national publicly available CDC and SART sources.

Objective

Our review assesses the Stork Club Case Study methodology's appropriateness for the purposes of calculating and comparing ART performance measure results across populations and identifies methodological limitations to consider when interpreting the results. This review was limited to the Stork Club Case Study methodology, which Stork Club used to calculate ART performance measure results for two populations. Readers of the Milliman report can review the Stork Club Case Study results alongside the Milliman report to better understand the results in the context of our findings. We have not reviewed Stork Club data or calculations or otherwise checked the results of the study. We are not offering any opinion on the accuracy or the validity of results now or in any future use of the methodology that we reviewed. The scope of our work does not include review of any other measures or studies Stork Club uses to assess its programs, such as its clinical or financial performance. Our work also does not consider or comment on the potential relationship of Stork Club's ART performance measure results to the features of its product offering.

Data and Methodology

Stork Club engaged Milliman to provide an independent review of its ART performance measure study (the "Stork Club Case Study"), which compares ART performance measures for Stork Club members receiving embryo transfers at in-network fertility clinics to national benchmarks for embryo transfers represented in CDC and SART data.

This independent review of Stork Club's ART performance measure study design methodology relies on the reports and calculations shared by Stork Club in 2024. Stork Club also provided responses to our questions on the services provided by Stork Club and further information about how each measure was calculated.

This review focuses on the following Stork Club Case Study methodological elements:

- **Measure selection**: the selection of ART performance measures.
- **Data collection and time periods**: the data collection, sample size, and time periods used in the analysis of ART performance measures.
- **Population selection and age-adjustment**: the selection of populations for comparison and age-adjustment of measure results.
- **Measure components**: consistency in the methodology used to calculate the measures for each population.

[4] Pirtea P, de Ziegler D, Poulain M, Ayoubi JM. Which key performance indicators are optimal to assess clinical management of assisted reproduction cycles? Fertil Steril. 2020 Jul;114(1):24-30. doi: 10.1016/j.fertnstert.2020.04.055. Epub 2020 Jun 10. PMID: 32532485.

MEASURE SELECTION

The six ART performance measures chosen by Stork Club for its case study encompass key stages and milestones in the embryo transfer process. These range from the decision to implant a single embryo, through to pregnancy and potential pregnancy loss, culminating in live births and the occurrence of multiple births.

DATA COLLECTION AND TIME PERIODS

Stork Club data

As part of the Stork Club Outcome Initiative program, in-network clinics submit ART performance measures data to Stork Club within three months of submitting a claim post-treatment completion. Stork Club conducts quarterly audits to confirm the receipt of all necessary data and to address any gaps in the information provided.

Stork Club reports that its members received a total of 182 embryo transfers at in-network fertility clinics across the calendar years 2021 to 2023, but only 155 transfers (85%) have complete data reported for all six ART performance measures included in the Study Club Case Study. Therefore, only these 155 transfers are included in the case study. No data for the ART performance measures was received for eight transfers, and an additional 19 transfers are not used in the case study due to data quality issues.

National benchmark data

The national ART performance measure results used in the Stork Club Case Study rely upon two national datasets: SART preliminary 2021 National Summary Report and the CDC 2021 National Summary. These data sets became available in 2023 and were the most recent and complete data from SART and the CDC available at the time of Stork Club's Case Study. The CDC estimates that they collect data and report on more than 95% of all ART cycles in the United States, because most non-reporting clinics are small.[5] The embryo transfers represent a subset of ART cycles as reported by SART and CDC as total cycles reported also include egg retrieval and egg freezing that are stored for potential future use. In its 2021 National Summary, the CDC reported on services provided by 453 clinics, 80% of which were SART members.[6] SART's preliminary 2021 National Report accounted for 86.6% of transfers reported by the CDC. The CDC notes that any discrepancies in published statistics between CDC and SART success rates may be due to treatments performed at non-SART member clinics as well as differences in data submission deadlines, data processing procedures, and statistical methods.[7]

TABLE 1. NUMBER OF IVF TRANSFERS BY SOURCE

Source	Time Period	Number of transfers
CDC	2021 Final	107,945
SART	2021 Preliminary	93,699
Stork Club	2021-2023	155

Table 2 specifies the national data sources used for each ART performance measure. The more comprehensive CDC data serves as the source for all measures with available published information. For the remaining measures, SART data has been used due to the absence of necessary information from the CDC, such as pregnancies, miscarriages, and SETs by age group.

TABLE 2. STORK CLUB CASE STUDY SOURCES FOR NATIONAL BENCHMARK RESULTS BY ART PERFORMANCE MEASURE

MEASURE	SOURCE
Pregnancy Rate (per transfer)	SART 2021 Preliminary
Miscarriage Rate (per pregnancy)	SART 2021 Preliminary
Transfers per Live Birth	CDC 2021 Final
Live Birth Rate (per transfer)	CDC 2021 Final
SET Rate (per transfer)	SART 2021 Preliminary
Multiples Rate (per transfer)	CDC 2021 Final

[5] Centers for Disease Control and Prevention. How Do We Monitor Assisted Reproductive Technology? US Dept of Health and Human Services; Available from: https://www.cdc.gov/art/nass/index.html#:~:text=Because%20most%20non%2Dreporting%20clinics,cycles%20in%20the%20United%20States
[6] Centers for Disease Control and Prevention. 2021 Assisted Reproductive Technology Fertility Clinic and National Summary Report. p.4-6. US Dept of Health and Human Services; 2023. Available from: https://www.cdc.gov/art/reports/2021/pdf/Report-ART-Fertility-Clinic-National-Summary-H.pdf
[7] Centers for Disease Control and Prevention. 2021 Assisted Reproductive Technology Fertility Clinic and National Summary Report. p.6. US Dept of Health and Human Services; 2023. Available from: https://www.cdc.gov/art/reports/2021/pdf/Report-ART-Fertility-Clinic-National-Summary-H.pdf

The SART and CDC measures only include ART performance measures results for individuals who used their own eggs. All Stork Club transfers were for individuals who used their own eggs, which is aligned with both national data sources.

The Stork Club Case Study uses CDC's cumulative success rates, which includes all embryo transfers and its ART performance measures from 2020 and 2021 based on 2020 egg retrievals.[8] The lag in CDC data is due to the live birth delivery data that requires data through October of the following year to account for the gestational period.[9] Further data lags occur due to the time clinics need to report data to the CDC, and the time it takes the CDC to validate and analyze the data.[10]

SART includes ART cycles started within a single calendar year, but the ART performance measure may not be realized until up to two years later.[11] The SART report is preliminary because a small portion of cycles have delayed outcomes past the cycle finalization deadline of 2023, but once these are finalized, the SART's report status will transition from preliminary to final.[12] As of the date of our review, final 2021 SART data was not publicly available.

POPULATION SELECTION AND AGE-ADJUSTMENT

The Stork Club Case Study evaluates ART performance measure results for Stork Club members who received ART embryo transfer at in-network fertility clinics compared to the national population who received embryo transfers as reported by CDC or SART, as applicable to the measure. National age distribution percentages are applied to Stork Club measure results by age group to normalize for differences in age between Stork Club members and the national population. The national reported age groups are: '<35', '35 to 37', '38 to 40', '>40' for the CDC, and '<35', '35 to 37', '38 to 40', '40 to 42', '>42' for SART. The age groups used by the Stork Club Case Study are aligned to those in the CDC national dataset, where age groups '40 to 42' and '>42' are combined into '>40' when age-adjusting against SART data.

MEASURE COMPONENTS

Tables 3A and 3B display ART performance measure definitions for the national benchmark data and Stork Club data, respectively. Based on information provided by Stork Club, we have specified the numerator and denominator terminology presented in the Stork Club Case Study to emphasize the similarities and differences in numerators and denominators between the national and Stork Club populations. Table 3C highlights the differences in methods, where differences between Stork Club and national averages are color-coded for reference.

TABLE 3A. STORK CLUB CASE STUDY ART PERFORMANCE MEASURE DEFINITIONS – NATIONAL BENCHMARK DATA

MEASURE	NUMERATOR	DENOMINATOR
Pregnancy Rate (per transfer)*	Total pregnancies	Total transfers
Miscarriage Rate (per pregnancy)*	Confirmed pregnancies resulting in miscarriage	Total pregnancies
Transfers per Live Birth**	Total transfers	Total live births
Live Birth Rate (per transfer)**	Total live births	Total transfers
SET Rate (per transfer)*	Total elected SETs	Total transfers
Multiples Rate (per transfer)**	Total live births – total singleton live births	Total transfers

*National Benchmark Data: *2021 SART preliminary and **2021 CDC final.*

[8] Centers for Disease Control and Prevention. 2021 Assisted Reproductive Technology Fertility Clinic and National Summary Report. p.3-4. US Dept of Health and Human Services; 2023. Available from: https://www.cdc.gov/art/reports/2021/pdf/Report-ART-Fertility-Clinic-National-Summary-H.pdf
[9] Centers for Disease Control and Prevention. 2021 Assisted Reproductive Technology Fertility Clinic and National Summary Report. p.3-4. US Dept of Health and Human Services; 2023. Available from: https://www.cdc.gov/art/reports/2021/pdf/Report-ART-Fertility-Clinic-National-Summary-H.pdf
[10] Centers for Disease Control and Prevention. 2021 Assisted Reproductive Technology Fertility Clinic and National Summary Report. p.3-4. US Dept of Health and Human Services; 2023. Available from: https://www.cdc.gov/art/reports/2021/pdf/Report-ART-Fertility-Clinic-National-Summary-H.pdf
[11] Society for Assisted Reproductive Technology. Understanding the SART Clinic Report. [Internet]. [cited 2024 Mar 6]. Available from: https://www.sart.org/patients/fyi-videos/understanding-the-sart-clinic-report/
[12] Society for Assisted Reproductive Technology. Understanding the SART Clinic Report. [Internet]. [cited 2024 Mar 6]. Available from: https://www.sart.org/patients/fyi-videos/understanding-the-sart-clinic-report/

TABLE 3B. STORK CLUB CASE STUDY ART PERFORMANCE MEASURE DEFINITIONS – STORK CLUB DATA

MEASURE	NUMERATOR	DENOMINATOR
Pregnancy Rate (per transfer)	Total pregnancies	Total transfers
Miscarriage Rate (per pregnancy)	Total reported miscarriages	Total pregnancies
Transfers per Live Birth	Total transfers	Total reported pregnancies – total reported miscarriages
Live Birth Rate (per transfer)	Total reported pregnancies – total reported miscarriages	Total transfers
SET Rate (per transfer)	Total elected SETs	Total transfers
Multiples Rate (per transfer)	Total number of multiple embryo transfers that resulted in non-miscarriage pregnancy with more than one heartbeat	Total transfers

TABLE 3C. DIFFERENCES IN ART PERFORMANCE MEASURE CALCULATIONS – NATIONAL BENCHMARK DATA VS. STORK CLUB DATA

MEASURE COMPONENT	DEFINITION DIFFERENCE
Pregnancies	Stork Club pregnancy is determined by a positive beta human chorionic gonadotropin (hCG) test and ultrasound, which is comparable to SART's clinical pregnancies, which SART defines as an 'intrauterine pregnancy confirmed by ultrasound'.[13]
Miscarriages	Stork Club clinics report 'Total reported miscarriages' around 8 weeks, whereas SART captures all miscarriages through 20 weeks of pregnancy.
Live Births	Stork Club does not collect live birth data. Measures that rely on live birth data, 'Transfers per Live Birth' and 'Live Birth Rate (per transfer)', instead use 'reported pregnancies – reported miscarriages' as a proxy for live births. This is in contrast with the CDC, which collects actual live birth data.
Multiple Live Births	Unlike SART, which collects live birth data and can distinguish between single and multiple births, Stork Club does not collect this information. Therefore, Stork Club uses the 'total number of multiple embryo transfers resulting in non-miscarriage pregnancies with more than one heartbeat' as a substitute numerator in the 'Multiples Rate (per transfer)' measure.

Results and Discussion

We reviewed the following key elements of the Stork Club Case Study methodology. We describe our review findings and identify and discuss the potential impact of the limitations of the methodological elements.

MEASURE SELECTION

The common measures used to assess ART performance focus on all the pivotal process and outcome milestones in the embryo transfer journey, such as achievement of pregnancy, live birth rate, and rate of singleton live births. The measures used in the Stork Club Case Study are commonly used in the industry to measure ART performance and provide an adequate overview of the continuum of stages in the embryo transfer journey through live birth. We did not identify any other ART performance measures that would need to be incorporated for the Stork Club Case Study methodology to be considered appropriate. However, the methodology includes two measures that are the inverse of each other – 'Transfers per Live Birth' and 'Live Birth Rate (per transfer)'. Consequently, we would not expect them to produce different results, rendering the inclusion of both as duplicative. For our review, we focused on five measures, deliberately omitting 'Transfers per Live Birth' from our discussion and overall findings.

DATA COLLECTION AND TIME PERIODS

Data collection

Stork Club requires in-network fertility clinics to submit information relevant to the calculation of ART performance measures within 30 days of a claim, and a quarterly review is undertaken to ensure data completeness and quality. However, while 95% of embryo transfers have the necessary data reported, which is in line with CDC reporting, another 10% of transfers are removed by Stork Club for data quality reasons. Having information for the calculation of

[13] Society for Assisted Reproductive Technology. Preliminary National Summary Report for 2021. [Internet]. [cited 2024 Mar 6]. Available from: https://www.sartcorsonline.com/Csr/Public?ClinicPKID=0&reportingYear=2021&newReport=True#

ART performance measures for only 85% of embryo transfers introduces the possibility of selection bias. For example, in-network fertility clinics with a lower pregnancy or live birth rate per transfer may disproportionately not report information to Stork Club or may provide information in a way that cannot be used, whereby the data are removed. In-network fertility clinics also may disproportionately not report on embryo transfers that resulted in miscarriage for a variety of reasons, such as members lost to follow up. It is unknown what effect the inclusion of data on the 15% of embryo transfers excluded from the Stork Club population would have on the ART performance measure results reported in the Stork Club Case Study. Using the 'Pregnancy rate (per transfer)' measure as an example: assume there are 100 total transfers, but usable data is only collected on 85 transfers. If there are 50 pregnancies that result from those 85 transfers, the pregnancy rate per transfer is 50/85, or 58.8%. If there were no pregnancies that resulted from those 15 transfers where data was not collected, the actual pregnancy rate per transfer would be 50/100, or 50%. This example illustrates the potential impact that omitted data could have on Stork Club ART performance measure results. Nevertheless, the Stork Club Case Study methodology that uses data from the large majority (85%) of all embryo transfers for Stork Club members at in-network fertility clinics in 2021-2023 is sufficiently complete that we do not view incomplete data as being the sole explanation for any observed differences in ART performance measure results between the national and Stork Club populations.

Sample size and statistical testing

Stork Club's embryo transfer volume from 2021-2023 is small. This small sample size increases uncertainty and underscores the need for an unbiased sample. Notably, the age distribution of key measure denominators – specifically, transfers and pregnancies – is similar between Stork Club and national data. This suggests that Stork Club members are not meaningfully different from the national population on this dimension in 2021-2023. Therefore, age-adjustment is unlikely to be skewing the overall results or introducing bias. This similarity in age distribution between Stork Club and national data allows for t-tests on aggregated age-adjusted results to assess the statistical significance of differences in ART performance measure results between Stork Club and national populations.

Stork Club's statistical testing shows that three measure results have a p-value less than 0.05, which demonstrates that the measure results are statistically significantly different according to standard statistical cut-offs.[14] This means that any observed differences in measure results between Stork Club and national population are likely to represent true differences in performance as opposed to resulting from random variation alone. However, Stork Club's statistical testing on the ART performance measures 'Miscarriage Rate (per pregnancy)' and 'Multiples Rate (per transfer)' do not show a statistically significant difference between Stork Club and national populations. Therefore, despite any observed differences in measure results, random variation means that Stork Club's true performance could be better, worse, or similar to national results for these two measures.

For the three measures that Stork Club currently shows are statistically significant, any future shifts in age distribution of Stork Club members undergoing embryo transfer could introduce large enough bias such that the Stork Case Study methodology might no longer be appropriate. For instance, if the sample included a substantially larger share of individuals over age 40 than the national population, statistical testing could need to be conducted on individual age groups to determine if comparisons of an aggregated age-adjusted performance measure remain appropriate.

Time periods

We reviewed CDC and SART's national data from 2017 to 2020 for the five ART performance measures to detect any significant historical trends that could impact a comparison of 2021 national data to 2021-2023 Stork Club data that covered a more recent time period. The national pregnancy per transfer rate improved slightly from about 53% in 2017 to around 55% in 2019-2021, while the national miscarriage rate remained steady at around 15% from 2017-2021. National live births per transfer rose consistently until 2020, with a slight decrease from 2020 to 2021. The national SET rate per transfer increased from 59% to 74% from 2017-2019 but plateaued at 78% to 79% in 2020-2021. This trend was inversely mirrored by a decrease in national multiple births per transfer from 7% to 4% in 2017-2019 which plateaued at 3% in 2020-2021. The upward trend in 2017-2019 reflects an industry shift towards an overall SET rate that is consistent with evidence-based clinical guidelines which suggests that these rates have reached a steady state and are unlikely to substantially change in the future.[15] The impact of the COVID-19 pandemic on 2020 and 2021 national averages is unclear. Given current clinical guidelines and the stable ART performance measure results between 2020 and 2021, we expect no major trends in these national rates following 2021 once

[14] Andrade C. (2019). The P Value and Statistical Significance: Misunderstandings, Explanations, Challenges, and Alternatives. Indian journal of psychological medicine, 41(3), 210–215. https://doi.org/10.4103/IJPSYM.IJPSYM_193_19

[15] Reimundo, P., Gutiérrez Romero, J. M., Rodríguez Pérez, T., & Veiga, E. (2021). Single-embryo transfer: a key strategy to reduce the risk for multiple pregnancy in assisted human reproduction. Advances in laboratory medicine, 2(2), 179–198. Available from: https://doi.org/10.1515/almed-2021-0013

national data becomes available through 2023.[16] While alignment between Stork Club and national data time periods would be preferred, we believe it is reasonable for the methodology to ignore trend adjustments when considering the difference in these time periods.

POPULATION SELECTION AND AGE ADJUSTMENT

Population selection

The methodology in the Stork Club Case Study is applied to two populations: (1) Stork Club members receiving ART embryo transfers at in-network fertility clinics; and (2) the national population receiving ART embryo transfers at fertility clinics represented in CDC or SART data, depending on the ART performance measure. The study methodology does not allow the isolation of the impact of network selection from the impact of Stork Club's program interventions on ART performance measures.

Age adjustment and other normalization factors

The Stork Club Case Study methodology normalizes for age by applying weights to Stork Club's age-specific rates for four age groups ('<35', '35 to 37', '38 to 40', '>40') based on the respective national age distribution, which varies across measures depending on the measure's denominator and the national data source. For instance, the 'Pregnancy Rate (per transfer)' measure is weighted according to the SART's distribution of transfers by age group. This is an appropriate methodology to normalize Stork Club measures by age, given the level of age stratification available in the benchmark data.

Numerous factors influence ART performance measure results. Some are clinic-related, like service quality or willingness to accept patients with complex infertility issues.[17] Others are patient-specific, such as age, egg and sperm quality, infertility cause, and genetic factors.[18] Processes in the embryo transfer journey, such as the use of preimplantation genetic testing (PGT) and intracytoplasmic sperm injection (ICSI), that may be indicated for certain patients have the potential to impact ART performance measure results. Other sociodemographic factors, such as race, may also play contributory roles to differences in ART performance measure results. In the U.S., racial disparities in ART performance measure results, particularly for Black people, have remained persistent over the past 15 years.[19] Race alone is an influencing factor in both live birth and cumulative live birth rates, with Black people experiencing lower live and cumulative live birth rates compared to white people.[20] Employed status can also be a relevant factor in ART performance measure results, and all Stork Club members have employer-provided fertility benefits, unlike the national mix of employer-funded and self-pay members. The high ART costs may pose financial challenges for self-pay members, which could influence decisions that affect ART performance measure results. While age is one of the stronger predictors in determining ART performance measure results as fertility generally decreases with age, normalizing for other factors would enable more robust assessment of true differences in ART performance measures results between populations.[21] Specifically, while differences in individual factors alone between comparison populations may not have substantial impacts on ART performance measure results for those populations, multiple factors in aggregate could more meaningfully impact these results. However, Stork Club methodology is consistent with industry practice in comparing ART performance measure result differences by adjusting only for age because complete information on other relevant factors is not available in the national comparison data. The national datasets collect information on a limited number of additional factors by age group that have the potential to impact ART performance measure results. Stork Club does not currently collect information on these factors. Two processes in the embryo transfer journey, PGT and ICSI, are reported in the SART database, where they are available by age group. However, normalizing for these two factors would not be expected to meaningfully change ART performance measure results in the general population receiving embryo transfers.[22] Other

[16] Reimundo, P., Gutiérrez Romero, J. M., Rodríguez Pérez, T., & Veiga, E. (2021). Single-embryo transfer: a key strategy to reduce the risk for multiple pregnancy in assisted human reproduction. Advances in laboratory medicine, 2(2), 179–198. Available from: https://doi.org/10.1515/almed-2021-0013
[17] Centers for Disease Control and Prevention. 2021 Assisted Reproductive Technology Fertility Clinic and National Summary Report. p.9-10. US Dept of Health and Human Services; 2023. Available from: https://www.cdc.gov/art/reports/2021/pdf/Report-ART-Fertility-Clinic-National-Summary-H.pdf
[18] Centers for Disease Control and Prevention. 2021 Assisted Reproductive Technology Fertility Clinic and National Summary Report. p.9-10. US Dept of Health and Human Services; 2023. Available from: https://www.cdc.gov/art/reports/2021/pdf/Report-ART-Fertility-Clinic-National-Summary-H.pdf
[19] Seifer, D.B., Simsek, B., Wantman, E. et al. (2020). Status of racial disparities between black and white women undergoing assisted reproductive technology in the US. Reprod Biol Endocrinol 18, 113. Available from: https://doi.org/10.1186/s12958-020-00662-4https://rbej.biomedcentral.com/articles/10.1186/s12958-020-00662-4
[20] Seifer, D.B., Simsek, B., Wantman, E. et al. (2020). Status of racial disparities between black and white women undergoing assisted reproductive technology in the US. Reprod Biol Endocrinol 18, 113. Available from: https://doi.org/10.1186/s12958-020-00662-4
[21] Fertility and ageing. (2005). Human Reproduction Update, 11(3), 261–276. Available from: https://doi.org/10.1093/humupd/dmi006
[22] Practice Committees of the American Society for Reproductive Medicine and the Society for Assisted Reproductive Technology. Electronic address: asrm@asrm.org (2020). Intracytoplasmic sperm injection (ICSI) for non-male factor indications: a committee opinion. Fertility and sterility, 114(2), 239–245. https://doi.org/10.1016/j.fertnstert.2020.05.032 / Simopoulou, M., Sfakianoudis, K., Maziotis, E., Tsioulou, P., Grigoriadis, S., Rapani, A., Giannelou, P., Asimakopoulou, M., Kokkali, G., Pantou, A., Nikolettos, K., Vlahos, N., & Pantos, K. (2021). PGT-A: who and when? A systematic review and network meta-analysis of RCTs. Journal of assisted reproduction and genetics, 38(8), 1939–1957. https://doi.org/10.1007/s10815-021-02227-9

factors mentioned above, including the reason for infertility, are not currently available in national datasets by age group, so even if Stork Club were to collect this information, normalizing for these factors would not be possible.

MEASURE COMPONENTS

Live birth data and miscarriages

Stork Club does not collect live birth information but instead uses the formula "reported pregnancies – reported miscarriages" as a proxy, with miscarriages reported after a minimum of 8 weeks. This proxy presents potential misalignment with actual live birth data due to the following:

- CDC's 2021 data reveals that about 82% of pregnancies resulted in live births, while 16% led to pregnancy loss, including 15.5% miscarriages and 0.5% stillbirths.[23] The remaining 2% had other or unknown outcomes. Therefore, subtracting only miscarriages from total pregnancies may not fully account for the gap between pregnancies and live births.[24]
- Research on miscarriage rates by week is limited due to the lack of a reliable tracking method aside from self-reported data. However, available research suggests that miscarriage rates after 8 weeks range from 1.5% to 5.2%, varying by source.[25]

These factors indicate the possibility that Stork Club live births may be mildly overstated when compared to national benchmarks that use actual reported live births. Because Stork Club's live birth estimates rely on miscarriage data, any understatement in reported miscarriages could result in lower results for two ART performance measures - Miscarriage Rate (per pregnancy) and Multiples Rate (per transfer) - and a higher rate for one measure, Live Birth Rate (per transfer). Of these three measures, only Live Birth Rate (per transfer) is statistically significantly different under the Stork Club Case Study methodology, and we do not expect the potential small overstatement of Stork Club live births could solely explain any observed differences between Stork Club and the national population for this measure.

REVIEW SUMMARY

Our review of the Stork Club Case Study methodology found that it included industry-standard ART performance measures to track the embryo transfer journey through live birth and has a sufficiently complete data collection process. Despite Stork Club's small number of transfers, the similar age distribution between Stork Club and national data allows for age-adjusted comparisons on an overall level, rather than by age group for three of the selected ART performance measures. The methodology for the two measures - miscarriage rate per pregnancy and multiples rate per transfer - relies on having a larger sample size to produce reliable results for these rates that approach extreme values (e.g., 0% or 100%). Future shifts in age distribution, composition of in-network clinics or the percentage of Stork Club embryo transfers with complete information reported could impact the methodology's appropriateness when applied to future experience.

As the Stork Club Case Study is limited to comparing Stork Club's in-network fertility clinic embryo transfers to national averages, the methodology does not account for network selection which would isolate the impact of Stork Club's program design on the ART performance measure results. While the methodology only adjusts for age, has a misalignment of data collection periods for the two populations and uses a proxy to estimate number of live births for Stork Club results, it is largely consistent with industry practice and is appropriate for comparing ART performance in 2021-2023 between Stork Club and national populations for three ART performance measures - pregnancy rate per transfer, live birth rate per transfer, and SET rate per transfer.

[23] Centers for Disease Control and Prevention. 2021 Assisted Reproductive Technology Fertility Clinic and National Summary Report. p.15. US Dept of Health and Human Services; 2023. Available from: https://www.cdc.gov/art/reports/2021/pdf/Report-ART-Fertility-Clinic-National-Summary-H.pdf
[24] Centers for Disease Control and Prevention. 2021 Assisted Reproductive Technology Fertility Clinic and National Summary Report. p.15. US Dept of Health and Human Services; 2023. Available from: https://www.cdc.gov/art/reports/2021/pdf/Report-ART-Fertility-Clinic-National-Summary-H.pdf
[25] Tong, S., Kaur, A., Walker, S. P., Bryant, V., Onwude, J. L., & Permezel, M. (2008). Miscarriage risk for asymptomatic women after a normal first-trimester prenatal visit. Obstetrics and gynecology, 111(3), 710–714. Available from: https://doi.org/10.1097/AOG.0b013e318163747c; Datayze. Miscarriage Probability Chart. [Internet]. [cited 2024 Mar 6]. Available from: https://datayze.com/miscarriage-chart.php

Milliman

Milliman is among the world's largest providers of actuarial, risk management, and technology solutions. Our consulting and advanced analytics capabilities encompass healthcare, property & casualty insurance, life insurance and financial services, and employee benefits. Founded in 1947, Milliman is an independent firm with offices in major cities around the globe.

milliman.com

CONTACT

Harsha Mirchandani
harsha.mirchandani@milliman.com

Brittany Ryan
brittany.ryan@milliman.com

Carol Bazell
carol.bazell@milliman.com