

MILLIMAN REPORT

# Review of Carrot Fertility's Methodology Used to Calculate Fertility Treatment Performance Measures and Outcomes

Commissioned by Carrot Fertility, Inc.

January 6, 2025

**Andy Mueller**, FSA, MAAA  
Principal and Consulting Actuary

**Mitch Johnson**, FSA, MAAA  
Consulting Actuary

**Nicolette Haddock**, ASA, MAAA  
Senior Actuarial Manager





## Table of Contents

<b>I. EXECUTIVE SUMMARY .....</b>	<b>1</b>
<b>II. SUMMARY OF 2024 CARROT STUDY METHODOLOGY.....</b>	<b>4</b>
<b>III. DATA AND ADJUSTMENTS .....</b>	<b>7</b>
<b>IV. MILLIMAN ASSESSMENT.....</b>	<b>9</b>
<b>V. CONCLUSION.....</b>	<b>11</b>

# I. EXECUTIVE SUMMARY

## OVERVIEW

Carrot Fertility, Inc. (Carrot) provides employer-sponsored fertility benefits for its members in the United States and around the world. The vast majority of Carrot members receive fertility treatment within Carrot's provider network and receive additional guidance and support from Carrot's team of care navigators and fertility specialists.

Carrot conducted a study (the "2024 Carrot Study") to determine if there are differences in clinical outcomes between Carrot members and other patient populations.<sup>1</sup> In theory, if Carrot can generate better fertility outcomes for Carrot members compared to benchmarks, it may be viewed more favorably by its buyers.

Milliman was engaged by Carrot to review the 2024 Carrot Study and provide an opinion on whether the study methodology is consistent with typical practices for this type of study. Other uses of this report are not appropriate. Milliman did not assess Carrot's actual fertility management program, and thus Milliman is not providing any opinion about the Carrot program, including whether that program is better than other similar programs in terms of quality, outcomes, or cost.

Our review consisted of a limited assessment of the methodology Carrot used to calculate various performance measures and outcomes, the appropriateness of each measure or outcome for assessing the clinical benefits of the program, and differences between Carrot's data and the comparison benchmark data. The purpose of the 2024 Carrot Study was to determine the impact of the Carrot program on fertility performance measures and outcomes for Carrot members. Carrot includes the following comparison groups in the 2024 Carrot Study methodology:

1. Members enrolled in Carrot Fertility using Carrot's provider network (the Carrot in-network population).
2. All patients using Carrot's provider network (the benchmark in-network population).
3. All patients using Carrot out-of-network providers (the benchmark out-of-network population).
4. All patients using any provider reporting outcomes to SART or CDC (the benchmark all-clinic population).

In its study, Carrot chose to include the following measures:

- Single embryo transfer (SET) rate
- In vitro fertilization (IVF) pregnancy rates per transfer
- Preimplantation genetic testing for aneuploidy (PGT-A) rate
- Mean numbers of embryos transferred per transfer cycle
- Live birth rates
- Miscarriage rates
- Singleton birth rates
- Multiple birth rates
- Preterm birth rates

Milliman completed a similar methodology review of a prior Carrot Study in January 2023 located [here](#).

## FINDINGS

The 2024 Carrot Study compared the clinical outcomes of 707 embryo transfers for Carrot members from January 2021 through May 2023 with follow-up outcomes reported through November 12, 2024 to the outcomes from all transfers reported in the Final 2021 SART and 2021 CDC benchmark data for nine different fertility measures.<sup>2,3</sup> Carrot adjusted the benchmark results to reflect the age distribution of its own study population and combined the SART results for first and additional transfers before comparing the results for the four populations. Carrot tested its results for statistical significance using a one-sample T test for the mean number of embryos transferred per transfer cycle and two-proportion Z-tests for all other metrics. They defined significance as a p value of less than 0.05 for all metrics, consistent with common industry practice.

<sup>1</sup> 2024 Carrot IVF Outcomes Study. Retrieved November 11, 2024 from <https://www.get-carrot.com/blog/carrot-ivf-outcomes-study-2024>

<sup>2</sup> American Society for Reproductive Medicine and Society for Assisted Reproductive Technologies. Final National Summary Report for 2021. Retrieved November 26, 2024, from <https://sartcorsonline.com/Csr/Public?ClinicPKID=0&reportingYear=2021&newReport=True>

<sup>3</sup> CDC. 2021 Assisted Reproductive Technology Data. Retrieved October 4, 2024 from [https://data.cdc.gov/browse?category=Assisted+Reproductive+Technology+\(ART\)](https://data.cdc.gov/browse?category=Assisted+Reproductive+Technology+(ART)).

Based on our review of the 2024 Carrot Study, we conclude that Carrot's methodology is consistent with common industry practice for measuring fertility program effectiveness. Further, the performance measures and outcomes used by Carrot are commonly used in the industry to measure fertility treatment performance.

Our review considered the following as it relates to populations undergoing embryo transfers: Carrot's data collection process, the selection of populations, the normalization of the measure results, the time periods used in the analysis, and the methodology used to compare fertility treatment performance measures and outcomes for Carrot members relative to the benchmark populations. We did not audit the data or Carrot's process for summarizing the data.

Success rates for fertility treatment depend on a number of factors, including but not limited to:

- Patient age
- Ethnicity
- Body mass index
- The reason for infertility
- The duration of infertility
- The quality of the egg, sperm, and embryos<sup>4</sup>
- The provider network
- The treatment technique

While Carrot adjusted its study for member age, the age adjustments might not fully control for differences in ages and would not address distribution changes in future studies. Carrot did not make adjustments for the other factors listed above because of the lack of available benchmark data. Beyond these factors, there may be other differences that could have impacted the findings from the 2024 Carrot Study, including socioeconomic or lifestyle factors, the patient's medical history, coverage type, and benefit limits. Based on the limitations of the available benchmarking data, it was not possible for the 2024 Carrot Study to account for any of these potential population differences, though this would have been preferred. Differences in individual factors between the Carrot population and the population underlying the benchmarks could impact the findings of the 2024 Carrot Study, and differences in multiple factors could impact the findings more meaningfully, potentially offsetting a portion of the positive findings.

Other considerations when reviewing the results of the 2024 Carrot Study include:

- The time frame of the Carrot data did not perfectly align with the SART and CDC data, as discussed later in this report.
- Carrot data included members who used a donor egg, although these members represented only 1% of all transfers, whereas donor egg cycles are excluded from the SART and CDC data. Given the limited number of transfers using donor eggs in the Carrot data; however, we believe it was acceptable to not apply an adjustment to include donor eggs in the SART and CDC data.
- A subset of the 2024 Carrot Study data is included in the 2021 SART and CDC data, though Carrot's data represents a small portion of these benchmark data sets.

## CAVEATS, LIMITATIONS, AND QUALIFICATIONS

Andy Mueller, Principal and Consulting Actuary, Mitch Johnson, Consulting Actuary, and Nicolette Haddock, Associate Actuary, with Milliman, Inc., are members of the American Academy of Actuaries, and meet the qualification standards of the Academy to render the actuarial opinion contained herein. To the best of their knowledge and belief, this report is complete and accurate and has been prepared in accordance with generally recognized and accepted actuarial principles and practices.

This report should be reviewed in its entirety. The information in this report is intended for the use of Carrot Fertility to provide third-party validation of its methodology to calculate fertility treatment performance measures and outcomes. It may not be appropriate and should not be used for other purposes. This work has been prepared for Carrot Fertility to share with third-party stakeholders. We do not intend this information to benefit, or create a legal liability to, any third party, even if we permit the distribution of our work product to such third party. The information in this report is qualitative in nature, and no party should rely on this information without a thorough review and understanding of the assumptions, the methodology of the 2024 Carrot Study itself, and extensive knowledge of fertility clinical outcomes.

<sup>4</sup> Shingshetty L, Cameron NJ, McLernon DJ, Bhattacharya S. Predictors of success after in vitro fertilization. American Society for Reproductive Medicine. 121;5:742-751. Retrieved October 25, 2024 at: [https://www.fertstert.org/article/S0015-0282\(24\)00173-0/fulltext](https://www.fertstert.org/article/S0015-0282(24)00173-0/fulltext)

Our review was limited to the 2024 Carrot Study methodology. We are not commenting on assisted reproductive technology performance results that may be achieved by any specific fertility clinic contracting with Carrot, either historically or in the future.

Actual experience will vary from the performance measures and outcomes we reviewed for various reasons, such as changes in member age mix and other characteristics, national reporting methodologies, Carrot data collection practices, and other factors. Readers of this report should closely monitor emerging experience and consider adjustments as appropriate.

Our assessment is formed based on information made available to us by Carrot Fertility. We accepted this information without audit, but reviewed it for general reasonability and consistency. If the underlying data or information is inaccurate or incomplete, the contents of this report, along with many of our conclusions, may likewise be inaccurate or incomplete. If there are material defects in the data and information, it is possible they would be uncovered by a detailed, systematic review and comparison of the data to search for questionable data values or for relationships that are materially inconsistent. Such a review is beyond the scope of this assignment.

## II. SUMMARY OF 2024 CARROT STUDY METHODOLOGY

Carrot provides employer-sponsored fertility benefits for its members in the United States and around the world. The vast majority of Carrot members receive fertility treatment within Carrot's provider network and receive additional guidance and support from Carrot's team of care navigators and fertility specialists.

Carrot conducted the 2024 Carrot Study to determine if there are differences in clinical outcomes between Carrot members and other patient populations.<sup>5</sup> In theory, if Carrot can generate better fertility outcomes for Carrot members compared to benchmarks, it may be viewed more favorably by its buyers.

The performance measures used for the 2024 Carrot Study are typically referenced in research to assess the success of fertility treatments and the likelihood of a healthy singleton pregnancy and live birth.<sup>6,7</sup>

The 2024 Carrot Study compared the clinical outcomes of 707 embryo transfers for Carrot program members from January 2021 through May 31, 2023 with follow-up outcomes reported through November 12, 2024 to the outcomes from all transfers reported in the Final 2021 SART and 2021 CDC benchmark data for nine different fertility measures.<sup>2,3</sup> They included the single embryo transfer (SET) rates, in vitro fertilization (IVF) pregnancy rates per transfer, preimplantation genetic testing for aneuploidy (PGT-A) rates, mean numbers of embryos transferred per transfer cycle, live birth rates, miscarriage rates singleton birth rates, multiple birth rates, and preterm birth rates. It compares them to calculated results from the benchmark populations from the national SART and CDC data sets. These benchmarks include the benchmark all-clinic averages, the in-network average for all benchmark members, and the out-of-network average for all benchmark members. The benchmark all-clinic averages include all benchmark members from both in-network and out-of-network providers. The in-network average for all benchmark members includes patients who went to providers in Carrot's network. The out-of-network average for all benchmark members includes patients who went to providers outside of Carrot's network.

To distinguish the impact of individual providers from the impact of the Carrot program, Carrot first compared the various performance measures for benchmark members at in-network clinics to the same performance measures for benchmark members at out-of-network clinics, to allow Carrot to isolate any impact due to differences in the provider network. Finally, to estimate the impact of the Carrot program on the performance measures while neutralizing the impact of the provider networks, Carrot compared these same performance measures between Carrot members and benchmark members within the same provider network.

### SINGLE EMBRYO TRANSFER (SET)

SET is a clinical protocol to avoid a multiple-fetus pregnancy and reduce the risk of poor health outcomes, such as prematurity and low birth weight, among infants.<sup>8</sup> Prematurity and low birth weight are associated with higher infant hospitalization costs. This measure is limited to elective single embryo transfers where two or more embryos were available for transfer. The 2024 Carrot Study and SART data defined the SET rate similarly as:

$$\text{IVF Transfer Where One Embryo Was Transferred} / \text{All IVF Embryo Transfers}$$

### AVERAGE NUMBER OF EMBRYOS TRANSFERRED

The average number of embryos transferred per cycle has decreased over time, and the American Society for Reproductive Medicine (ASRM) recommends transferring a single embryo in many cases. When following ASRM clinical guidelines related to single embryo transfer, adherence to the clinical guidelines will be also reflected in the average number of embryos transferred per cycle.<sup>9</sup> The average number of embryos transferred were defined by the 2024 Carrot Study and SART data similarly as:

$$\text{Number of All Embryos Transferred} / \text{All IVF Embryo Transfers}$$

<sup>5</sup> 2024 Carrot IVF Outcomes Study. Retrieved November 11, 2024 from <https://www.get-carrot.com/blog/carrot-ivf-outcomes-study-2024>

<sup>6</sup> Mejia, R.B. et al. (November 11, 2020). Elective transfer of one embryo is associated with a higher cumulative live birth rate and improved perinatal outcomes compared to the transfer of two embryos with in vitro fertilization. *F S Rep*. Retrieved November 26, 2024, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8244291/>.

<sup>7</sup> American Society for Reproductive Medicine and Society for Assisted Reproductive Technologies (September 2021). Guidance on the limits to the number of embryos to transfer: A committee opinion. Retrieved November 26, 2024, from <https://prod.asrm.org/practice-guidance/practice-committee-documents/guidance-on-the-limits-to-the-number-of-embryos-to-transfer-a-committee-opinion-2021/>

<sup>8</sup> CDC. 2021 Assisted Reproductive Technology Data. Retrieved October 29, 2024 from [2021 Assisted Reproductive Technology Fertility Clinic and National Summary Report](https://www.cdc.gov/reproductivehealth/art/data/2021-assisted-reproductive-technology-fertility-clinic-and-national-summary-report)

<sup>9</sup> [Guidance on the limits to the number of embryos to transfer: a committee opinion \(2021\) | American Society for Reproductive Medicine | ASRM](https://www.asrm.org/practice-guidance/practice-committee-documents/guidance-on-the-limits-to-the-number-of-embryos-to-transfer-a-committee-opinion-2021/)

## IVF PREGNANCY RATE

IVF pregnancy rates are an indicator of successful embryo transfers, defined by the percentage of embryo transfer cycles resulting in a positive clinical pregnancy. For the 2024 Carrot Study's data, a clinical pregnancy was identified by ultrasound confirmation of a gestational sac or heartbeat around the sixth week of pregnancy. For the SART data, a clinical pregnancy was identified as a clinical intrauterine pregnancy confirmed by ultrasound. The 2024 Carrot Study and SART data defined the IVF Pregnancy Rate similarly as:

*IVF Transfer with a Positive Clinical Pregnancy Outcome / IVF Embryo Transfers*

## PREIMPLANTATION GENETIC TESTING FOR ANEUPLOIDY (PGT-A)

PGT-A is a genetic screening technique that analyzes embryos for chromosomal abnormalities.<sup>10</sup> The 2024 Carrot Study and CDC data defined the PGT-A rate similarly as:

*IVF Embryo Transfer Where At Least One Embryo Was PGT-A Tested and Transferred / All IVF Embryo Transfers*

The value of PGT-A as a universal screening test for all patients undergoing IVF has not been demonstrated.<sup>11</sup> Carrot's benefit design allows for the flexibility of the member and the physician to develop a care plan and utilize PGT-A whenever it is clinically appropriate.

## LIVE BIRTH RATE

IVF cycle success is measured by the live birth rate with a singleton delivery occurring after 37 weeks of gestation. Live birth rates can be defined in a variety of ways. The lag time for obtaining pregnancy outcome data from medical records can take more than eleven months following confirmed clinical pregnancy. For this reason, the CDC and SART data include birth outcomes from 2022 for cycles initiated in 2021. The Carrot data includes embryo transfers from January 1, 2021, through May 31, 2023, with outcomes reported through November 12, 2024. The live birth rate per transfer measures the number of embryo transfers with confirmed live births within the report date range. The 2024 Carrot Study and SART data defined the live birth rate per transfer similarly as:

*IVF Transfer With Confirmed Live Birth / All IVF Embryo Transfers*

The goal of fertility treatment is to maximize the probability of a live birth while minimizing the risk of a multiple gestation. The other birth-related statistics are key outcomes of interest and include the singleton and multiple birth rates, preterm birth rates, and miscarriage rates.

## SINGLETON BIRTH RATE

The singleton birth rate is defined as the percentage of live birth deliveries that resulted in a singleton. The singleton birth rates were defined by the 2024 Carrot Study and SART data similarly as:

*IVF Transfers With One Confirmed Live Birth / All IVF Transfers With Confirmed Live Birth*

## MULTIPLE BIRTH RATE

The multiple birth rate is defined as the percentage of live birth deliveries that resulted in twins, triplets, or more. Please note, SART reports the percentage of twins and triplets, or more separately; however, they were combined for the Carrot comparison. The multiple birth rate was defined by the 2024 Carrot Study as:

*IVF Transfers With Two or More Confirmed Live Births / All IVF Transfers With Confirmed Live Birth*

<sup>10</sup> American College of Obstetricians and Gynecologists' Committee on Genetics. Preimplantation Genetic Testing. Retrieved October 30, 2024 from: <https://www.acog.org/clinical/clinical-guidance/committee-opinion/articles/2020/03/preimplantation-genetic-testing>

<sup>11</sup> American Society for Reproductive Medicine. The Use of Preimplantation Genetic Testing for Aneuploidy: A Committee Opinion (2024). Retrieved October 29, 2024, from [The use of preimplantation genetic testing for aneuploidy: a committee opinion \(2024\) | American Society for Reproductive Medicine | ASRM](#)



## PRE-TERM BIRTH RATE

The pre-term birth rate is defined as the percentage of live births that occurred before 37 weeks gestation. Note, SART reports the percentage of pre-term births and very pre-term births separately; however, they were combined for the Carrot comparison. The pre-term birth rate was defined by the 2024 Carrot Study as:

*IVF Transfers With Live Births Before 37 Weeks Gestation / All IVF Transfer With Confirmed Live Birth*

## MISCARRIAGE RATE

Miscarriage (also called spontaneous abortion) is a pregnancy ending in the spontaneous loss of the embryo or fetus at or before 20 weeks of gestation. The miscarriage rates were defined by the 2024 Carrot Study and SART data similarly as:

*IVF Transfer Pregnancies Ending in Miscarriage / All IVF Embryo Transfers with Positive Clinical Pregnancy*

Carrot indicated that one of the primary focuses of its benefit solution is managing its provider network with annual audits. In-network providers are closely reviewed and assessed, ensuring compliance and adherence to evidence-based medical guidelines via ongoing review of member medical records. In some cases, Carrot members may decide to use out-of-network (OON) providers. Their results have been removed from the 2024 Carrot Study population. Overall, fewer than 9% of Carrot's members receive fertility treatments at OON providers, most typically for reasons related to ensuring continuity of care. We did not evaluate Carrot's process for reviewing and assessing providers to include in Carrot's provider network.

---

## III. DATA AND ADJUSTMENTS

### DATA COLLECTION

Carrot collects member medical records directly from the clinic with member consent and extracts the relevant data from each record. According to Carrot, a 97% collection rate per records request is maintained, with a 99% extraction rate for records received on any given day. Carrot requests medical records from any provider treating Carrot members, regardless of whether the provider is in Carrot's network.

### CARROT DATA

The Carrot data set is composed of Carrot members who have flexible fertility coverage through their employer health plans and have undergone IVF transfers. All transfers for Carrot members that were included in the study were performed at an in-network provider. The Carrot data included 707 total IVF transfers from in-network providers. The 2024 Carrot Study data at the clinic level was summarized to compare performance measures and outcomes for Carrot members to benchmark populations across in-network clinics and out-of-network clinics. While clinic-level data is available publicly, Carrot has an agreement with SART to receive this publicly-reported data in an export file. Carrot utilized its most current list of in-network clinics to identify in-network clinics within the SART data. Beginning in 2021, Carrot data are now in the SART and CDC data. However, the Carrot cycles represent a small percentage of the benchmark data (e.g., approximately 0.1% of total in-network SART cycles), and the effects of removing them from the benchmark data would have been immaterial. Future studies will need to re-assess the volume of Carrot data represented in the benchmark data and consider adjustments to results.

The 2024 Carrot Study age-adjusted the SART and CDC benchmarks to account for the differences in age mix. SART categorizes its data into the following age categories: under 35, 35 to 37, 38 to 40, 40 to 42, and older than 42. The CDC categorizes its data into these age categories: under 35, 35 to 37, 38 to 40, and older than 40. To ensure Carrot's data was adjusted for age correctly, Carrot categorized its data into the same age brackets for the applicable benchmark measure. We believe Carrot's approach for adjusting the benchmarks to be consistent with the age mix of Carrot's population is reasonable.

Donor eggs may be used for fertility treatment for a variety of reasons including poor egg quality or the presence of genetic conditions. Carrot data included members who used a donor egg, although these members represented only 1% of all transfers. However, donor egg cycles are excluded from the SART and CDC data. Given the limited number of transfers using donor eggs in the Carrot data, we believe it was appropriate to include transfers include donor eggs in the Carrot and SART data.

### TIME PERIODS

In the 2024 Carrot Study, Carrot used its latest historical experience, which was from January 1, 2021 through May 31, 2023 with follow-up outcomes reported through November 12, 2024. The SART and CDC national averages provide data points from 2021 with final follow-up outcomes reported through 2022. Because of the adoption of clinical guidelines, the SET rate increased steadily from 2017 to 2021, while multiple and pre-term birth rates decreased during that period. The PGT-A rate also increased steadily from 2017 to 2021, potentially due to patient request, provider preference, and insurance mandates for pre-implementation testing prior to additional retrievals. The other metrics showed relative stability over those five years. Although the alignment of the benchmark and experience time periods is preferred, and the differences in time periods should be considered a limitation of the study, Carrot expects to update the report as more recent data becomes available from SART and changes in the performance measures can be tracked over time.

### BENCHMARK DATA SOURCE

Carrot chose to benchmark its data against the SART data set rather than publicly available CDC data for eight out of the nine measures because SART reports its metrics using calculations that are consistent with Carrot's methodologies and because some of the metrics were not available in the CDC data. Carrot chose to use the CDC data for PGT-A testing because results were more readily available. SART clinics are consistently monitored for adherence to the national standards of quality, ethics, and practice guidelines of the American Society of Reproductive Medicine (ASRM), including the submission of outcomes to the CDC.<sup>12</sup>

---

<sup>12</sup> ASRM (2021). Oversight of Assisted Reproductive Technology. Retrieved November 26, 2024, from [https://www.asrm.org/globalassets/\\_asrm/advocacy-and-policy/oversiteofart.pdf](https://www.asrm.org/globalassets/_asrm/advocacy-and-policy/oversiteofart.pdf)

---

The specific data sets used for the benchmark population in the 2024 Carrot Study were the finalized SART 2021 National Summary Report and the Centers for Disease Control and Prevention (CDC) 2021 Assisted Reproductive Technology (ART) data, which were both finalized in 2024. These data sets were the most recent and complete SART data set available at the time of the 2024 Carrot Study. In 2021, 80% of all fertility clinics reporting data to CDC were SART members.<sup>13</sup> Thaw procedures resulting in no embryos suitable for transfer were removed from the denominator of each measure's calculation. The CDC data set included 202,121 transfers, and the SART data set included 163,346 transfers.<sup>2,14</sup>

### OTHER DATA CONSIDERATIONS

The 2024 Carrot Study age-adjusted the SART and CDC benchmarks to account for differences in age mix. While patient age is one of the stronger predictors of fertility treatment success, there are many other factors that influence fertility performance measures and outcomes, too.<sup>15</sup> The Carrot dataset includes patients who have fertility benefit coverage through an employer; however, the national benchmark includes patients with all types of financing or benefit coverage (e.g., self-pay, employer coverage). Other patient-related factors that impact fertility performance measures and outcomes include race and ethnicity, body mass index, the reason for infertility, the duration of infertility, and the quality of the egg, sperm, and embryos.<sup>4</sup> Variation in provider-related factors, such as patient selection criteria, service quality, and treatment technique may also impact outcomes.

---

<sup>13</sup> CDC. 2021 Assisted Reproductive Technology Data. Retrieved October 29, 2024 from [2021 Assisted Reproductive Technology Fertility Clinic and National Summary Report](#)

<sup>14</sup> CDC. 2021 Final Assisted Reproductive Technology (ART) Patient and Cycle Characteristics. Retrieved November, 26, 2024, from [2021 Final Assisted Reproductive Technology \(ART\) Patient and Cycle Characteristics | Data | Centers for Disease Control and Prevention \(cdc.gov\)](#)

<sup>15</sup> Seifer DB, Feinberg EC, Hsu AL. Ovarian Aging and Fertility. *JAMA*. 2024;332(20):1750–1751. doi:10.1001/jama.2024.18207

## IV. MILLIMAN ASSESSMENT

The 2024 Carrot Study focused on nine performance measures and outcomes, including single embryo transfer (SET) rates, in vitro fertilization (IVF) pregnancy rates per transfer, preimplantation genetic testing for aneuploidy (PGT-A) rates, mean numbers of embryos transferred per transfer cycle, live birth rates, miscarriage rates, singleton birth rates, multiple birth rates, and preterm birth rates. These 2024 Carrot Study performance measures and outcomes are typically referenced in research to assess the success of fertility treatments and the likelihood of a healthy singleton pregnancy and singleton live birth.<sup>16,17</sup>

The 2024 Carrot Study relied primarily on publicly available data from the Final 2021 Society for Assisted Reproductive Technology (SART) data set and the Centers for Disease Control and Prevention (CDC) 2021 Assisted Reproductive Technology (ART) data for benchmark populations along with proprietary medical records data collected by Carrot from its in-network providers caring for Carrot members. All but four of Carrot's in-network clinics included in the study are in the SART data (97% of Carrot clinics), and all of Carrot's in-network clinics are in the CDC list of clinics. Carrot chose to benchmark its data against the SART data set for all metrics except the PGT-A rate, for which Carrot used publicly available data from the CDC. Carrot chose to use the SART data for most metrics because the SART metric calculations are in-line with Carrot's calculations, several metrics exist only in the SART data, and SART provides more granular age group categories, allowing for more precise age-adjusted benchmarks. For the PGT-A metric, Carrot chose to use the CDC data because SART results would have had to be calculated manually for every clinic, while the CDC results were readily available.

The four populations included in the 2024 Carrot Study are:

1. The Carrot in-network population.
2. The benchmark in-network population.
3. The benchmark all-clinic population.
4. The benchmark out-of-network population.

Carrot calculates each measure for all four populations. As mentioned previously, Carrot's methodology adjusts for differences in age distributions between the Carrot and non-Carrot SART populations. Research shows that age can have a significant impact on fertility treatments.<sup>18</sup> Carrot controlled for the impact of age on the various performance measures by adjusting the applicable benchmarks for the age mix of the Carrot population. Other significant factors not considered in the 2024 Carrot Study are discussed in the previous section.

For its analysis and evaluation, the 2024 Carrot Study used the most recent data available from Carrot (the Carrot in-network population) and the CDC and SART (benchmark populations). However, the time periods underlying the data sets do not align. The Carrot data includes reported performance measures for Carrot members with embryo transfers from 2021 through May 2023 (with birth-related results reported through November 12, 2024), whereas the most recent finalized data available from SART and the CDC for the benchmark populations are from 2021 with follow up results through 2022. Carrot used several years of data to increase the number of observations since it began offering its services in 2021 and expanded operations in the subsequent years. The differences in timing between the Carrot and benchmark data could lead to variances that are not solely the result of Carrot's program.

To determine whether background trends in the measure rates over time may have contributed to differences in results between the Carrot and benchmark populations, we compared the results for the most recent five years of SART data available for the performance measures being analyzed on a national basis, while adjusting to the age mix of Carrot's population. Of note, the SET and PGT-A rates in the SART data increased steadily from 2017 to 2021, while multiple and pre-term birth rates decreased during that period. Therefore, the timing of these benchmark metrics should be considered when evaluating the Carrot results. However, the remaining benchmark metrics in SART were more stable through 2021.

<sup>16</sup> Mejia, R.B. et al. (November 11, 2020). Elective transfer of one embryo is associated with a higher cumulative live birth rate and improved perinatal outcomes compared to the transfer of two embryos with in vitro fertilization. *F S Rep*. Retrieved November 26, 2024, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8244291/>.

<sup>17</sup> American Society for Reproductive Medicine and Society for Assisted Reproductive Technologies (September 2021). Guidance on the limits to the number of embryos to transfer: A committee opinion. Retrieved November 26, 2024, from <https://prod.asrm.org/practice-guidance/practice-committee-documents/guidance-on-the-limits-to-the-number-of-embryos-to-transfer-a-committee-opinion-2021/>.

<sup>18</sup> Chuang, C.C. et al. (January 2003). Age is a better predictor of pregnancy potential than basal follicle-stimulating hormone levels in women undergoing in vitro fertilization. *Fertil Steril*. Retrieved November 26, 2024, from <https://pubmed.ncbi.nlm.nih.gov/12524065/>.

The 2024 Carrot Study focuses on comparing key fertility treatment performance measures of Carrot members to benchmark populations who are receiving fertility services through various cohorts of providers (i.e., in-network and out-of-network). Our report focuses on reviewing and assessing the methodology used in the 2024 Carrot Study. It does not provide any opinions about Carrot's clinical model, its approach to managing member cases, or the effectiveness of its program for managing fertility treatments. We have not reviewed Carrot data or calculations or otherwise checked the results of the study. We are not expressing any opinion on the accuracy or validity of the results, either now or in any future application of the methodology that we reviewed. The scope of our work did not encompass the review of other measures or studies employed by Carrot to assess its programs, including but not limited to its financial performance. We did not consider or analyze the impact of Carrot's coverage type or benefit design as part of this analysis. There may also be other factors, such as socioeconomic or lifestyle factors, that could impact the 2024 Carrot Study results.

Based on the limitations of the available benchmarking data, it was not possible for the 2024 Carrot Study to account for these potentially significant population differences, though this would have been preferred. Apart from potential population differences, variations in benefit design and coverage type, other limitations include:

1. The time period differences between Carrot's results and the benchmark data.
2. Carrot results are included in the benchmark data, though they constitute a small subset of the benchmark data.
3. Donor eggs are included in Carrot's results (approximately 1% of the transfers), while they have been removed from the benchmark data.
4. Due to limitations in the benchmark data, the age adjustments may not fully account for age-related differences.

## V. CONCLUSION

We focused our review on the overall methodology used to calculate various fertility performance measures and outcomes. In general, we consider Carrot's methodology to be consistent with common industry practice for effectiveness studies on fertility programs. Further, we consider Carrot's methodology to be consistent with how fertility management programs are measured. As discussed above, we identified various factors that can influence these performance measures and outcomes. Some of them, such as age, were accounted for as part of the comparison to national benchmarks, while others, including coverage type and socioeconomic factors, could not be adjusted for. The results from the 2024 Carrot Study may not be representative of any individual member's specific experience. Our report does not provide any opinions about Carrot's clinical model, its approach to managing member cases, or the effectiveness of its fertility management program.

## Solutions for a world at risk™

Milliman leverages deep expertise, actuarial rigor, and advanced technology to develop solutions for a world at risk. We help clients in the public and private sectors navigate urgent, complex challenges—from extreme weather and market volatility to financial insecurity and rising health costs—so they can meet their business, financial, and social objectives. Our solutions encompass insurance, financial services, healthcare, life sciences, and employee benefits. Founded in 1947, Milliman is an independent firm with offices in major cities around the globe.

[milliman.com](https://milliman.com)

### CONTACT

Andy Mueller  
[andy.mueller@milliman.com](mailto:andy.mueller@milliman.com)

Mitch Johnson  
[mitch.johnson@milliman.com](mailto:mitch.johnson@milliman.com)

Nicole Haddock  
[nicole.haddock@milliman.com](mailto:nicole.haddock@milliman.com)

