Draft version June 13, 2025
Typeset using LaTeX **twocolumn** style in AASTeX631

Efficiency Analysis of US ART Clinics: A Data Envelopment Analysis Approach (2020-2022)

Denario[1]

[1] *Anthropic, Gemini & OpenAI servers. Planet Earth.*

ABSTRACT

This study investigates the technical efficiency of U.S. Assisted Reproductive Technology (ART) clinics in converting resources into successful outcomes, an area where performance can vary widely. We employ Data Envelopment Analysis (DEA) to assess the relative efficiency of clinics in transforming intended own-egg retrieval cycles into live births, stratified by patient age groups. Utilizing clinic-level data from the 2020-2022 National ART Surveillance System (NASS) dataset and an input-oriented Banker, Charnes, Cooper (BCC) model with variable returns to scale, we model the input-output relationship and identify the efficiency frontier for each year and age group. The analysis reveals generally low mean and median efficiency scores across all strata, significant performance heterogeneity, a negative correlation between patient age and clinic efficiency, and a substantial impact of zero-output cycles on efficiency scores. These findings highlight opportunities for performance improvement and best practice dissemination within the U.S. ART sector, particularly concerning the reduction of zero-output cycles and the improvement of outcomes for older patients.

Keywords: Cosmology, Orbits, Relativity, Gravitation, General relativity

1. INTRODUCTION

As infertility rates rise, Assisted Reproductive Technology (ART) has emerged as a crucial option for many individuals and couples. The increasing demand for ART services underscores the importance of ensuring these services are delivered with maximum efficiency and effectiveness. However, the performance of ART clinics exhibits considerable variability, influenced by factors such as patient demographics, treatment protocols, resource allocation strategies, and overall clinic management. Quantifying and understanding this performance heterogeneity is essential for identifying areas ripe for improvement and for disseminating best practices across the ART sector.

A significant challenge in evaluating the efficiency of ART clinics lies in the complex process of converting resources into successful live births. This process is influenced by numerous variables, many of which are outside the direct control of the clinics. These include patient age, pre-existing medical conditions, and individual responses to treatment. Furthermore, the data used to assess clinic performance often contains inherent noise and variability, making it difficult to isolate the true underlying efficiency of the clinics. This challenge is compounded by the absence of a standardized methodology for comparing the performance of different clinics, leading to inconsistencies in evaluation and hindering the identification of optimal practices.

This study addresses these challenges by employing Data Envelopment Analysis (DEA) to assess the relative technical efficiency of U.S. ART clinics. DEA is a non-parametric method that evaluates the efficiency of decision-making units (DMUs)—in this case, ART clinics—by comparing their input-output relationships to an efficiency frontier that represents the best observed practices. Unlike traditional statistical methods, DEA does not require any prior assumptions about the underlying production function or the distribution of errors, making it particularly well-suited for analyzing complex systems with multiple inputs and outputs. We utilize clinic-level data from the 2020-2022 National ART Surveillance System (NASS) dataset, focusing on own-egg cycles and stratifying the data by patient age group to control for age-related effects on ART outcomes. Our approach involves modeling the conversion of intended retrieval cycles (input) into live births (output), calculated from reported success rates and cycle counts. We implement an input-oriented Banker, Charnes, Cooper (BCC) model, assuming variable returns to scale, which is generally more appropriate for the operational characteristics of entities like ART clinics.

The DEA is conducted for each year (2020, 2021, and 2022) and age group to identify the efficiency frontier and calculate efficiency scores for individual clinics relative to this frontier. This allows us to benchmark clinics and investigate temporal changes in efficiency over the three-year period. The efficiency scores θ are calculated by solving the following optimization problem for each clinic:

$$\begin{aligned} &\text{Minimize } \theta \\ &\text{Subject to: } \sum_{j=1}^{N}(\lambda_j x_j) \leq \theta x_o \\ &\sum_{j=1}^{N}(\lambda_j y_j) \geq y_o \\ &\sum_{j=1}^{N}\lambda_j = 1 \\ &\lambda_j \geq 0 \quad \forall j \end{aligned}$$

where x_j and y_j are the input (Intended Retrievals) and output (Live Births) of clinic j, respectively, x_o and y_o are the input and output of the clinic being evaluated, λ_j are the weights assigned to each clinic, and N is the total number of clinics in the stratum.

To verify that our approach effectively captures the efficiency of ART clinics, we analyze the distribution of efficiency scores across different strata and examine temporal trends. We also perform exploratory data analysis to understand the key characteristics of the inputs and outputs. By demonstrating the validity and reliability of our DEA-based approach, we aim to provide a valuable tool for assessing and improving the efficiency of ART clinics in the U.S.

2. METHODS

2.1. *Data and Sample*

This study leverages data from the National ART Surveillance System (NASS) dataset, specifically focusing on clinic-level data from the years 2020, 2021, and 2022. The NASS dataset, maintained by the Centers for Disease Control and Prevention (CDC), provides comprehensive information on Assisted Reproductive Technology (ART) procedures performed in the United States. To ensure a focused analysis, we restricted our sample to ART cycles involving the use of patients' own eggs, thereby excluding cycles involving donor eggs. This selection was based on the 'Type' column in the dataset, filtering for records where the value corresponded to "Patients using their own eggs." Furthermore, the analysis was stratified by patient age group to account for the well-documented impact of age on ART outcomes, as highlighted in the introduction. The age group stratification was based on the 'Breakout' column when the 'Breakout_Category' was 'Age group of patient'. This stratification allowed for a more nuanced understanding of clinic efficiency across different patient demographics.

2.2. *Data Preprocessing*

Prior to conducting the Data Envelopment Analysis (DEA), the raw NASS data underwent a series of preprocessing steps to ensure data quality and suitability for the analysis. The initial step involved filtering the dataset to include only the years of interest (2020, 2021, and 2022) and own-egg cycles, as described above. Subsequently, we identified the relevant metrics necessary for DEA: the number of intended retrieval cycles (input) and the number of live births (output). The number of intended retrieval cycles was directly obtained from the 'Cycle_Count' column in the dataset. The number of live births was calculated by multiplying the percentage of live births per intended retrieval ('Data_Value_num' column, representing the "% Live Births per Intended Retrieval") by the corresponding 'Cycle_Count'. Both 'Data_Value_num' and 'Cycle_Count' columns were verified to be numeric, and converted if necessary, handling instances where data was suppressed or not applicable (e.g., represented by '*' or '-'). These instances were converted to NaN values in the 'Data_Value_num' column. The calculated number of live births was then rounded to the nearest whole number and converted to an integer data type, reflecting that live births are discrete counts. To address potential issues arising from missing or invalid data, we removed records where the number of intended retrievals was missing, zero, or non-positive. Additionally, rows with missing or negative values for the number of live births were removed. This cleaning process ensured that the dataset used for DEA contained only valid and relevant data points.

2.3. *Data Envelopment Analysis (DEA) Model*

The core of our analysis involved the application of Data Envelopment Analysis (DEA) to assess the relative technical efficiency of U.S. ART clinics. DEA is a non-parametric method that evaluates the efficiency of decision-making units (DMUs), in this case, ART clinics, by comparing their input-output relationships to an efficiency frontier representing the best observed practices. We employed an input-oriented Banker, Charnes, Cooper (BCC) model, assuming variable returns to scale (VRS). The input orientation reflects our focus on identifying the potential for clinics to reduce their inputs (intended retrieval cycles) while maintaining their current level of output (live births). The BCC model with VRS is particularly well-suited for analyzing ART clinics, as it acknowledges the potential for non-constant returns to scale due to factors such as varying patient populations, treatment protocols, and clinic management practices, as mentioned in the introduction.

The DEA was conducted separately for each year (2020, 2021, and 2022) and age group stratum. For each stratum, the efficiency score (θ) for each clinic was calculated by solving the following linear programming problem:

$$\begin{aligned} &\text{Minimize } \theta \\ &\text{Subject to: } \sum_{j=1}^{N}(\lambda_j x_j) \leq \theta x_o \\ &\sum_{j=1}^{N}(\lambda_j y_j) \geq y_o \\ &\sum_{j=1}^{N}\lambda_j = 1 \\ &\lambda_j \geq 0 \quad \forall j \end{aligned}$$

where:

- x_j and y_j represent the input (Intended Retrievals) and output (Live Births) of clinic j, respectively.
- x_o and y_o represent the input and output of the clinic being evaluated (clinic o).
- λ_j is the weight assigned to each clinic j.
- N is the total number of clinics in the specific year and age group stratum.

The objective function minimizes θ, which represents the proportional reduction in inputs that clinic o could achieve while still producing the same level of output. The first constraint ensures that the weighted sum of inputs for all clinics does not exceed the scaled input of the clinic being evaluated. The second constraint ensures that the weighted sum of outputs for all clinics is at least as great as the output of the clinic being evaluated. The third constraint, $\sum_{j=1}^{N}\lambda_j = 1$, enforces the variable returns to scale assumption, allowing the efficiency frontier to be a convex hull rather than a ray from the origin. The final constraint, $\lambda_j \geq 0$, ensures that all weights are non-negative. The efficiency score θ ranges from 0 to 1, with a score of 1 indicating that the clinic is technically efficient and operating on the efficiency frontier. Scores less than 1 indicate inefficiency, suggesting that the clinic could potentially reduce its inputs or increase its outputs to improve its performance.

2.4. *DEA Implementation*

The DEA model was implemented using the ‘scipy.optimize.linprog‘ function in Python. A separate function, ‘solve_dea_bcc_input_oriented‘, was created to encapsulate the linear programming problem for each clinic within each stratum. This function takes arrays of inputs and outputs for all DMUs in a stratum, along with the index of the DMU being evaluated. It then constructs the objective function, constraints, and bounds required by ‘linprog‘ and solves the optimization problem. The resulting efficiency score is returned. To accelerate the DEA calculations, the process was parallelized using the ‘multiprocessing‘ library in Python. A pool of worker processes was created to simultaneously evaluate the efficiency of multiple clinics within each stratum. This parallelization significantly reduced the overall computation time, particularly for strata with a large number of clinics.

2.5. *Post-DEA Analysis*

Following the DEA calculations, several post-DEA analyses were conducted to summarize and interpret the results. First, we analyzed the distribution of efficiency scores across different years and age groups. Descriptive statistics, including the mean, median, minimum, maximum, and standard deviation of efficiency scores, were calculated for each stratum. Additionally, we calculated the number and percentage of efficient clinics (those with an efficiency score of 1) within each stratum. These statistics provided insights into the overall level of efficiency in the U.S. ART sector and the degree of performance heterogeneity across different patient age groups. Second, we performed a temporal analysis of efficiency to examine how average efficiency scores changed over the three-year period (2020-2022) for each age group. This analysis involved grouping the efficiency scores by age group and examining the trend of mean efficiency scores across years. This allowed us to assess whether efficiency was improving, declining, or remaining stable over time.

3. RESULTS

3.1. *Dataset and DEA Input-Output Specification*

The analysis utilized the CDC’s National ART Surveillance System (NASS) Final ART Success Rates dataset for reporting years 2020, 2021, and 2022. Data were filtered to include only cycles where patients used their own eggs, as detailed in the Methods section. The primary metric for constructing DEA variables was ”What percentage of intended egg retrievals resulted in live-birth deliveries?”, stratified by ”Age of Patient” (< 35, $35-37$, $38-40$, > 40).

Each Decision Making Unit (DMU) in this study represents a unique combination of a clinic, reporting year, and patient age group. As described in the Methods section, the DEA model employed was an input-oriented Banker, Charnes, Cooper (BCC-I) model, which assumes variable returns to scale (VRS). The single input and output were defined as:

- **Input (X):** `Input_IntendedRetrievals` – The number of intended egg retrieval cycles reported by the clinic for a specific year and age group.
- **Output (Y):** `Output_LiveBirths` – The number of live births resulting from these intended retrievals, calculated by multiplying the reported percentage of live births per intended retrieval (`Data_Value_num`) by the `Input_IntendedRetrievals` (`Cycle_Count`) and rounding to the nearest whole number.

After initial loading of 1,126,080 records, filtering for the specified years, "Patients using their own eggs" type, the relevant question, and age breakout category yielded 48,960 records. Subsequent cleaning for DEA requirements (positive inputs, valid outputs) and removal of 36 records corresponding to national summary data (ClinicId 9999) resulted in a final dataset of 31,164 DMUs for analysis. This prepared dataset is stored in `data/dea_analysis_results/eda_outputs/prepared_dea_i`

3.2. *Descriptive Analysis of DEA Inputs and Outputs*

An exploratory data analysis (EDA) was conducted on the prepared dataset (excluding national data) to understand the characteristics of the DMUs. The detailed summary statistics for each stratum (Year, AgeGroup) are available in `data/dea_analysis_results/eda_outputs/eda_summary_statistics_by_stratum.csv`.

3.2.1. *Number of DMUs*

The dataset provided a substantial number of DMUs for each stratum. For instance, in 2020, the number of DMUs ranged from 2,464 for the > 40 age group to 2,752 for the < 35 age group. Similar counts were observed for 2021 and 2022, ensuring a robust basis for DEA within each stratum.

3.2.2. *Input: Intended Retrievals (`Input_IntendedRetrievals`)*

The number of intended retrievals (IR) per DMU varied considerably.

- **Central Tendency and Dispersion:** Mean IR values showed a tendency to increase slightly over the three-year period for most age groups. For example, in the < 35 age group, mean IR was 113.2 in 2020, 111.7 in 2021, and 133.8 in 2022. Median IR values were consistently lower than their respective means across all strata (e.g., for < 35 in 2020: median IR 43.0, mean IR 113.2). This, coupled with large standard deviations (e.g., 269.9 for < 35 in 2020) and high maximum values (e.g., 5021 cycles for a DMU in the < 35 group in 2020), indicates a significant right-skewness in the distribution of intended retrievals.
- **Visualizations:** This skewness is visually confirmed by the histograms presented in Figure 1, which show a large number of DMUs with fewer retrievals and a long tail representing DMUs with many retrievals. The boxplots in Figure 2 further illustrate this, highlighting numerous outliers at the higher end of the distribution for each stratum.

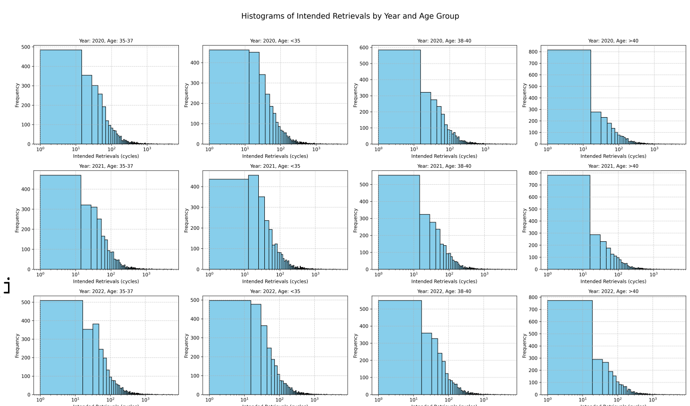

Figure 1. Histograms of intended retrieval cycles across different age groups and years, showing right-skewness, with a large number of clinics performing few retrievals and a long tail indicating clinics with many retrievals, which is consistent across all age groups and years.

3.2.3. *Output: Live Births (`Output_LiveBirths`)*

The number of live births (LB) also exhibited wide variation and skewness.

- **Central Tendency and Dispersion:** Mean LB generally followed trends similar to inputs, with slight increases over the years for younger age groups. For the < 35 group, mean LB was 47.2 in 2020, 46.0 in 2021, and 54.6 in 2022. A critical observation is that median LB values were frequently 0.0, particularly for older age groups. For instance, the > 40 age group reported a median LB of 0.0 across all three years. This indicates that at least half of the DMUs in these strata reported zero live births for the specific metric analyzed.

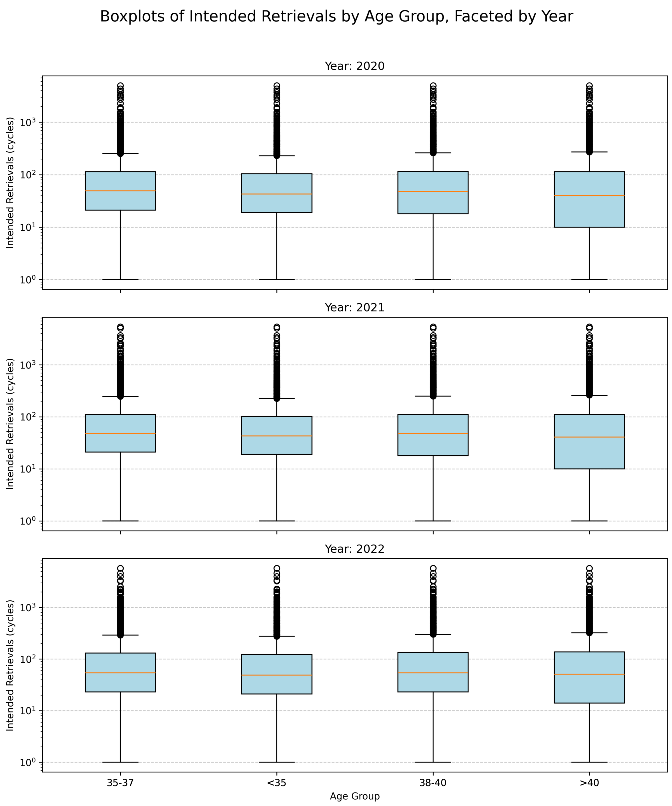

Figure 2. Boxplots illustrating the distribution of intended egg retrievals across different patient age groups for the years 2020-2022, demonstrating the variability in input levels for the Data Envelopment Analysis. The skewness and outliers in intended retrievals, visualized here, influence clinic efficiency scores, particularly when related to live birth outcomes.

- **Visualizations:** The distributions of live births were also highly right-skewed, as depicted in the histograms in Figure 3 and boxplots in Figure 4.

3.2.4. *Live Birth Rate per Intended Retrieval (`Data_Value_num`)*

The original reported live birth rate (LBRate) per intended retrieval, which was used to calculate the `Output_LiveBirths`, showed expected trends with age.

- **Central Tendency and Age-Related Trends:** Mean LBRate was highest for the < 35 age group (e.g., 22.46% in 2020) and systematically decreased with increasing patient age, reaching as low as 1.95% for the > 40 age group in 2020. Median LBRates were also frequently 0.0%, especially for the $35-37$, $38-40$, and > 40 age groups across all years. This reinforces the finding that a substantial proportion of DMUs reported no successful live births.

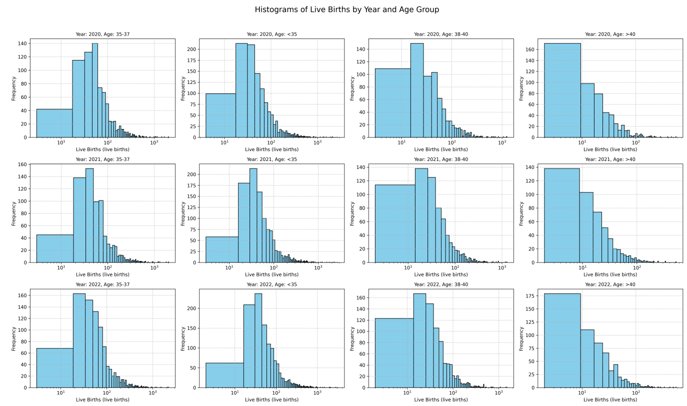

Figure 3. Histograms of live births show that the distribution of live births is skewed to the right, with many clinics reporting zero live births, especially for older age groups, which can significantly affect efficiency scores.

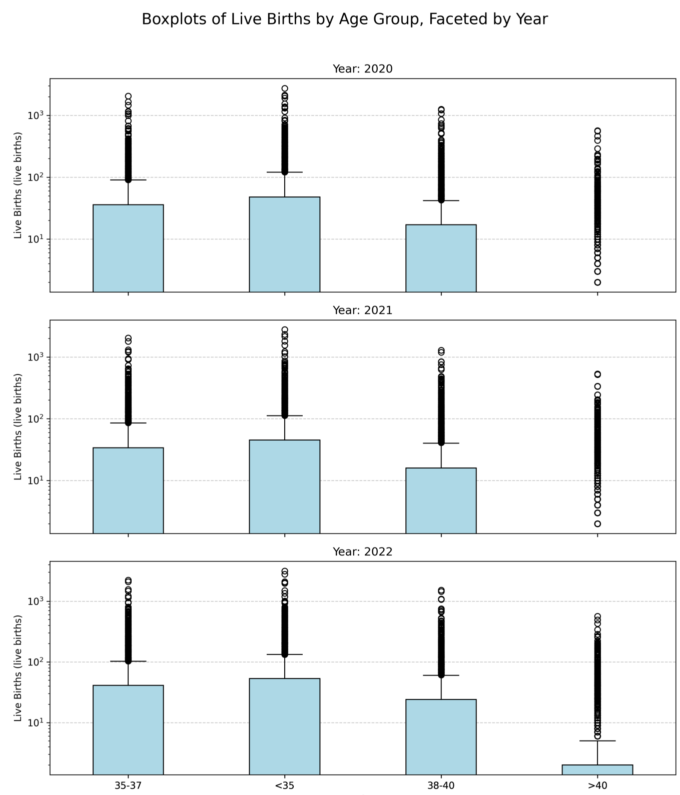

Figure 4. Boxplots of live births stratified by patient age group and year, showing the distribution and skewness of live births, with older age groups exhibiting lower median live births, impacting efficiency scores.

- **Visualizations:** Histograms in Figure 5 show a high concentration of LBRates at or near zero for older age groups. For the < 35 group, the distribution was more dispersed but still skewed. Boxplots in Figure 6 clearly illustrate the declining trend in

LBRate with increasing patient age and the prevalence of zero values.

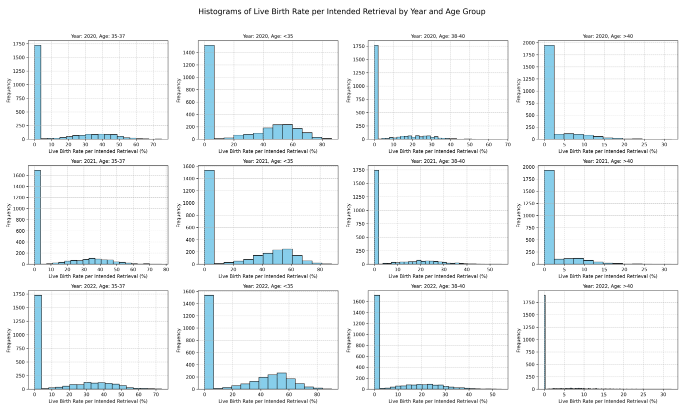

Figure 5. Histograms of live birth rate per intended retrieval for each year and age group, showing a high concentration of rates at or near zero, particularly for older age groups, which contributes to lower efficiency scores.

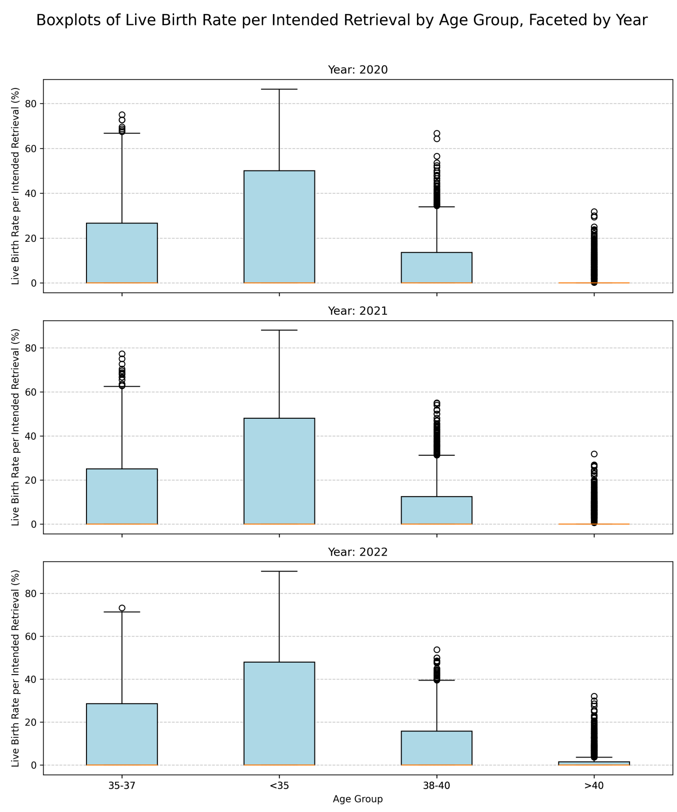

Figure 6. Boxplots of live birth rate per intended retrieval, stratified by patient age group and year, show a decreasing trend in live birth rate with increasing age.

The high incidence of zero live births and, consequently, zero live birth rates, particularly in older patient age groups, is a notable characteristic of the input data for DEA and is anticipated to significantly influence the efficiency scores. This confirms the motivation for age-group stratification outlined in the Introduction.

3.3. *Clinic Efficiency Scores*

The BCC-I DEA model, as described in the Methods section, yielded efficiency scores for 31,164 DMUs, stored in `data/dea_analysis_results/dea_scores/clinic_effic`
An efficiency score of 1.0 signifies that a DMU operates on the efficiency frontier (representing best practice within its stratum), while scores below 1.0 indicate relative inefficiency.

The aggregate statistics for all calculated efficiency scores are: mean = 0.2488, median = 0.1000, standard deviation = 0.2804, minimum = 0.0004, and maximum = 1.0000. The low mean and median scores suggest that, on average, DMUs operated considerably below the efficiency frontier established by their peers. No NaN scores were encountered, indicating successful computation for all DMUs.

3.3.1. *Stratum-Specific Efficiency Score Analysis*

A detailed breakdown of efficiency scores by year and age group is provided in Table 1 (derived from `data/dea_analysis_results/dea_scores/efficiency_score_su`

- **Mean and Median Efficiency:** As shown in Table 1, mean efficiency scores were consistently low, generally ranging from approximately 0.17 (for > 40 age group in 2022) to 0.34 (for < 35 age group in 2020). Median efficiency scores were notably lower than the means, particularly for older age groups (e.g., 0.0714 for > 40 in 2022), indicating a right-skewed distribution of efficiency scores within each stratum. This suggests that many DMUs exhibit low efficiency, while a smaller number achieve higher scores, thereby elevating the mean relative to the median.

- **Percentage of Efficient DMUs:** The proportion of DMUs operating on the efficiency frontier (score ≥ 0.9999) was small across all strata, typically ranging between 2% and 5%. For example, in 2022, this percentage varied from 2.28% for the < 35 age group to 4.29% for the > 40 age group.

3.4. *Temporal and Age-Group Trends in Efficiency*

3.4.1. *Temporal Trends (2020-2022)*

The analysis of efficiency score trends over the three-year period is summarized in `data/dea_analysis_results/dea_scores/temporal_efficiency`

Table 1. Summary of DEA Efficiency Scores by Year and Patient Age Group

Year	Age Group	Num DMUs	Mean Efficiency	Median Efficiency	Min Efficiency	Max Efficiency	Std Efficiency	Num Efficien
2020	<35	2752	0.3434	0.2500	0.0007	1.0	0.3161	
2020	35-37	2548	0.2577	0.0769	0.0013	1.0	0.2921	
2020	38-40	2524	0.2139	0.0807	0.0013	1.0	0.2585	
2020	>40	2464	0.1859	0.0833	0.0018	1.0	0.2458	
2021	<35	2720	0.3247	0.2000	0.0004	1.0	0.3040	
2021	35-37	2496	0.2504	0.0769	0.0004	1.0	0.2851	
2021	38-40	2491	0.2171	0.0833	0.0004	1.0	0.2537	
2021	>40	2441	0.1783	0.0769	0.0004	1.0	0.2386	
2022	<35	2849	0.3188	0.2500	0.0025	1.0	0.2934	
2022	35-37	2699	0.2612	0.1000	0.0004	1.0	0.2835	
2022	38-40	2640	0.2353	0.1000	0.0004	1.0	0.2671	
2022	>40	2540	0.1719	0.0714	0.0004	1.0	0.2339	

- **Mean Efficiency Trends:**
  - For the < 35 age group, mean efficiency exhibited a slight decline from 0.343 in 2020 to 0.325 in 2021, and further to 0.319 in 2022.
  - The $35-37$ age group showed relative stability in mean efficiency: 0.258 (2020), 0.250 (2021), and 0.261 (2022).
  - The $38-40$ age group experienced a modest increase in mean efficiency, from 0.214 in 2020 to 0.217 in 2021, and then to 0.235 in 2022.
  - For the > 40 age group, mean efficiency slightly decreased from 0.186 in 2020 to 0.178 in 2021, and then to 0.172 in 2022.

  These year-over-year changes in mean efficiency were generally minor, suggesting no substantial widespread shifts in overall clinic efficiency during the observed period.
- **Median Efficiency Trends:** Median efficiency scores also showed varied but generally minor changes. For the < 35 age group, median efficiency was 0.250 in 2020, 0.200 in 2021, and 0.250 in 2022. The $35-37$ and $38-40$ age groups saw slight increases in median efficiency by 2022, while the > 40 age group's median remained low and showed a slight decrease. Overall, median efficiencies remained low, indicating that the typical DMU's performance relative to the frontier did not change dramatically.

3.4.2. *Age-Group Trends in Efficiency*

A consistent pattern was observed regarding the relationship between patient age group and clinic efficiency across all three years:

- **Higher Efficiency in Younger Age Groups:** DMUs associated with patients in the < 35 age group consistently demonstrated the highest mean and median efficiency scores (Table 1).
- **Decreasing Efficiency with Age:** As patient age increased through the $35-37$, $38-40$, and > 40 categories, both mean and median efficiency scores generally decreased. The > 40 age group consistently exhibited the lowest average efficiency scores. This pattern suggests that achieving high relative efficiency is more challenging, or performance is more variable, in older patient populations, even when compared against peers treating the same age group. This is consistent with the known decline in fertility with age, as discussed in the Introduction.

The distribution of these efficiency scores is visualized in Figure 7 and Figure 8. The temporal trends of mean efficiency scores for each age group are shown in Figure 9.

3.5. *Impact of Zero Live Births on Efficiency (Sensitivity Analysis)*

A sensitivity analysis was performed to assess the impact of DMUs reporting zero live births (`Output_LiveBirths = 0`) on their efficiency scores. The findings are summarized in `data/dea_analysis_results/dea_scores/sensitivity_analysi` Figure 10 displays the distribution of efficiency scores for clinics with zero live births versus clinics with positive live births, stratified by year and patient age group.

- **Significant Efficiency Discrepancy:** A pronounced difference in efficiency scores was observed between DMUs with zero live births and those with at least one live birth.

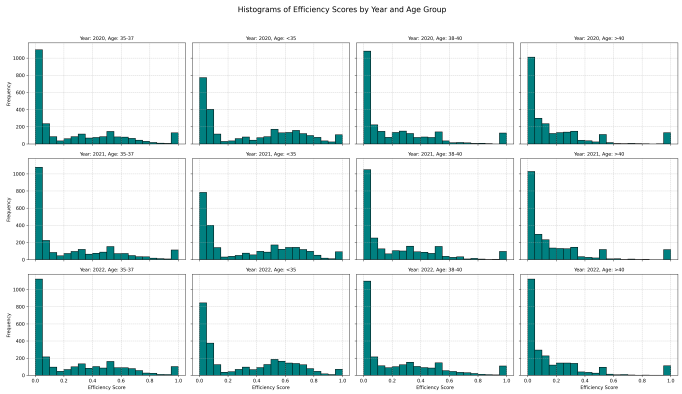

Figure 7. Histograms of clinic efficiency scores, stratified by year and patient age group, show a right-skewed distribution with many clinics having low efficiency and a few operating at the efficient frontier.

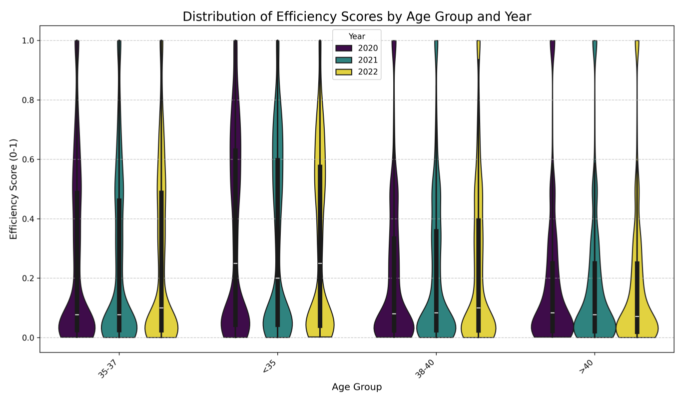

Figure 8. Efficiency score distributions for each age group, shown for each year. The distributions are skewed towards lower efficiency scores, particularly for older age groups.

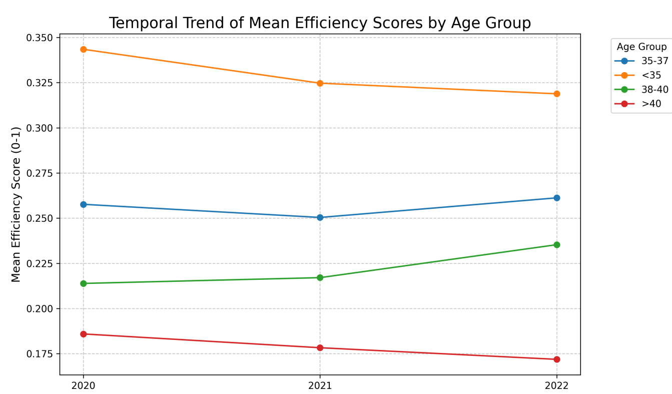

Figure 9. Temporal trend of mean efficiency scores for each patient age group from 2020 to 2022, showing a slight decline in efficiency for younger patients and relative stability for older patients.

- DMUs reporting **zero live births** consistently had very low mean and median efficiency scores across all strata. For example, in 2020 for the $35-37$ age group, DMUs with zero output had a mean efficiency of 0.139 and a median of 0.034, whereas DMUs with positive output had a mean efficiency of 0.503 and a median of 0.503. This pattern was consistent across all years and age groups.
- DMUs reporting **positive live births** had substantially higher mean and median efficiency scores. While these scores were still generally below 1.0 (indicating some relative inefficiency even among those achieving success), they were markedly greater than those of DMUs with no live births.

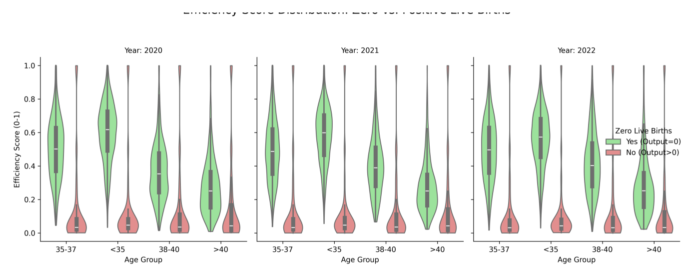

Figure 10. Distribution of efficiency scores for clinics with zero live births versus clinics with positive live births, stratified by year and patient age group. Clinics reporting zero live births exhibit substantially lower efficiency scores.

3.6. *Benchmarking and Implications*

The DEA results facilitate the identification of benchmark performance and highlights areas for potential improvement.

- **Benchmarking Performance:** DMUs achieving an efficiency score of 1.0 (or $\geq$0.9999) are considered to be on the efficiency frontier, representing best practice within their stratum. As indicated in Table 1, the percentage of such fully efficient DMUs was consistently low (generally 2-5%), implying that a vast majority of DMUs operate below the observed benchmark.
- **Performance Variation:** The wide distribution of efficiency scores and the low overall averages suggest significant heterogeneity in the performance of U.S. ART clinics concerning the conversion of intended retrievals into live births, even after accounting for patient age.

- **Implications of Zero-Output Cases:** The profound impact of zero live births on efficiency scores suggests that a critical factor influencing a clinic's relative efficiency is its ability to achieve at least some level of success (i.e., at least one live birth) within a given age group and year. Addressing the underlying causes of zero-output cycles could be a key area for improving overall efficiency metrics.

In summary, the DEA of U.S. ART clinics from 2020-2022 reveals generally low technical efficiency, with significant variation influenced by patient age and, critically, by whether any live births were achieved. Younger patient age groups are associated with higher efficiency scores. Temporal changes in efficiency were modest over the study period. A small fraction of clinics consistently operate on the efficiency frontier, indicating substantial opportunities for performance improvement across the sector. The sensitivity analysis highlights the importance of minimizing zero-output cycles as a key driver of overall efficiency. These results underscore the potential for clinics to learn from best practices and improve their resource utilization to enhance ART outcomes.

4. CONCLUSIONS

4.1. *Summary of Findings*

This study addressed the problem of variable efficiency among U.S. ART clinics in converting intended egg retrieval cycles into live births, an issue of increasing importance given rising infertility rates. We employed Data Envelopment Analysis (DEA) to assess the relative technical efficiency of clinics using clinic-level data from the 2020-2022 National ART Surveillance System (NASS) dataset. The input-oriented Banker, Charnes, Cooper (BCC) model, assuming variable returns to scale, was used to model the input-output relationship, with intended retrievals as the input and live births as the output, stratified by patient age groups.

4.2. *Key Results*

The DEA revealed generally low mean and median efficiency scores across all strata, indicating that, on average, clinics operated considerably below the efficiency frontier. Significant performance heterogeneity was observed, suggesting substantial variation in how effectively clinics convert resources into successful outcomes. A negative correlation was found between patient age and clinic efficiency, with younger age groups consistently demonstrating higher efficiency scores. Furthermore, a sensitivity analysis highlighted the substantial impact of zero-output cycles on efficiency scores, indicating that achieving at least one live birth is a critical factor influencing a clinic's relative efficiency. Temporal changes in efficiency were modest over the three-year study period.

4.3. *Learnings and Implications*

The results of this study underscore the potential for performance improvement and best practice dissemination within the U.S. ART sector. The low overall efficiency scores suggest that many clinics could benefit from optimizing their resource utilization and treatment protocols. The negative correlation between patient age and clinic efficiency highlights the need for targeted strategies to improve outcomes for older patients. The sensitivity analysis emphasizes the importance of minimizing zero-output cycles, suggesting that addressing the underlying causes of these cycles could be a key area for improving overall efficiency metrics. The identification of benchmark performers through DEA allows for the dissemination of best practices and the establishment of targets for improvement. Ultimately, this study provides valuable insights for clinics seeking to enhance their efficiency and improve ART outcomes for their patients.