# Summary of Quality Tests for Alternative Synthetic Data Models

## Description of Alternative Models

Along with the CTGAN model reported in the manuscript for Study 2, we considered several alternatives models that were generated using different approaches or parameters. Those models, and their specifications, are listed below. For each model, we conducted several of the same quality tests that we reported in the manuscript.

|  |  |  |
| --- | --- | --- |
| Model /Name Number | Batch Size | Epoch |
| 0 | 128 | 500 |
| 1 | 128 | 1000 |
| 2 | 128 | 2000 |
| 3 (ctgan-worst) | 128 | 3000 |
| 4 | 256 | 500 |
| 5 | 256 | 1000 |
| 6 | 256 | 2000 |
| 7 | 256 | 3000 |
| 8 (ctgan-best) | 500 | 500 |
| 9 | 500 | 1000 |
| 10 | 500 | 2000 |
| 11 | 500 | 3000 |

## Variable Correlation Tests

For each synthetic dataset, we compared differences in the correlation among the variables and the original data set. This is an important measure for gauging the substitutability of the synthetic data. Each pane in the figure below depicts a histogram representing the differences between the synthetic data and the original data. These results suggest that many of the alternative synthetic data generation measures produced correlations that differed substantially. The last two panels (depicting the best performing and worst performing ctgan models) help illustrate the range of differences that were observed.

A screenshot of a graph

Description automatically generated

## Comparison of Exploratory Factor Analyses

We also explored differences in the model’s substitutability regarding the factor loadings. As noted in the manuscript, we fit an exploratory factor analysis (EFA) using minimum residuals and varimax rotations that forced a two-factor solution onto the 11 leadership competencies for both self-ratings and superior’s ratings. We compared the differences in the factor loadings obtained in the original data with the alternative synthetic data models. We again observed markedly different levels of performance across the models, which are most pronounced when comparing the last two panes of the figures below (i.e., ctgan - best vs. ctgan – worst).

## Conclusion

These tests highlight the need to carefully consider a range of potential models when first generating synthetic data. Along with the model which we focused on in Study 2, we employed a range of hyperparameters (e.g., batch sizes that included 128, 256, and 500 and epochs including 500, 1,000, 2,000, and 3,000). The findings in this online supplement depict how modifying two hyperparameters can produce dramatically different levels of performance given the data one is seeking to synthetize. Thus, developing a suitable synthetic data generation model will likely represent an iterative, rather than a simply linear, process.

A screenshot of a graph

Description automatically generated