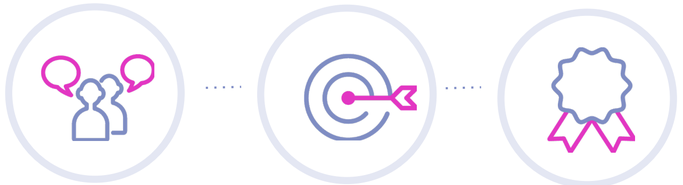
The Goal Attainment Scaling Method is Robust to Violations of Normality in Goal Scales: A Simulation Study.

Taylor Dunn[1], Justin Stanley[1], Kari Knox[1], Sanja Stanojevic[1,2], Susan Howlett[1,2], Kenneth Rockwood[1,2]

[1] Ardea Outcomes [2] Dalhousie University

Background

- **Goal attainment scaling (GAS)** is a patient-centric outcome measure that captures meaningful change through personally identified goals of treatment.
- GAS is generally a **three-step process:**



IDENTIFY GOALS

Clinician interviews subject/caregiver to identify goals of treatment (usually 3)

BUILD GAS SCALES

Set 5-point attainment scales for each identified goal

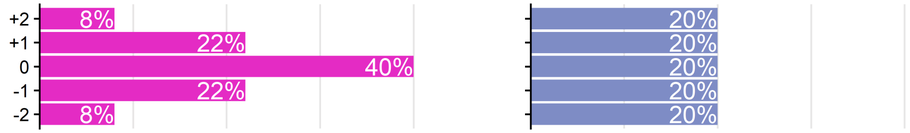
MEASURE ATTAINMENT

Rate during follow-up whether the goals have been attained

- A **key assumption** in the GAS method is that scores on the 5-point scales approximate a **normal distribution**.
- Using data simulation techniques, we investigated whether **GAS statistical properties varied** if the assumption **of normality was violated**.

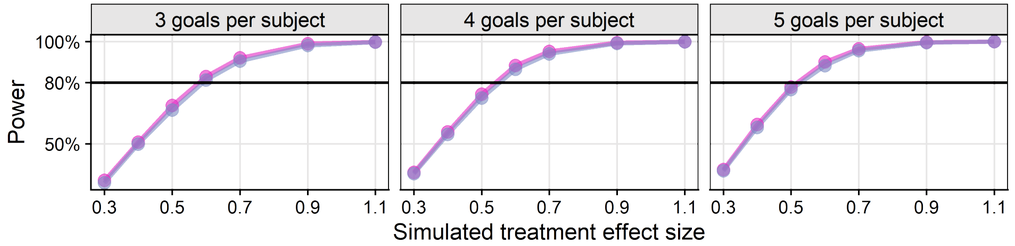
Results

Normally-distributed scores vs uniformly-distributed scores



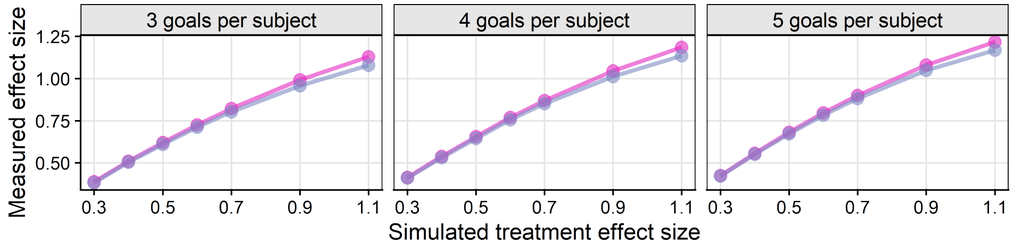
Statistical power does not differ significantly between normally-distributed and uniformly-distributed scores

Power vs simulated treatment effect for 60 subjects



Measured effect sizes slightly higher with normally-distributed compared to uniformly-distributed scores

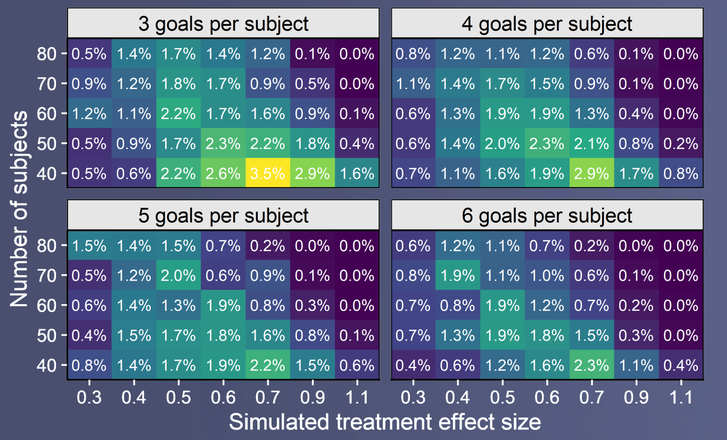
Simulated vs mean measured effect size for 60 subjects



ARDEA OUTCOMES

Results (cont.)

Difference in power (normal - uniform) for all simulation scenarios



- For the most part, there is less than a 2% difference in power to detect a treatment effect.

Methods

- We employed a latent variable model (Urach et al. 2019) to generate GAS data.
- The following parameters were varied: number of subjects, treatment effect size, and number of goals per subject.
- Latent goal scores were discretized into 5-point scales following **uniform** and **normal** distributions.
- 10,000 trials were simulated for each set of parameters.
- Two-sided *t*-tests on GAS T-scores were used to test the null hypothesis of no treatment effect.
- Power was the percentage of simulations detecting a significant effect at $\alpha = 0.05$. Standardized effect sizes were computed as Cohen's *d*.