Power Calculations for SEED 3 Supplemental COVID-19 Impact Assessments

The number of participants from SEED 3 with complete data collection prior to the end of March 2020 and an anticipated final study class of ASD, DD or POP is approximately 1,900 (ASD [n=500], DD [n = 800], POP [n = 600]). Based on the study completion rate observed in the initial follow-up study of SEED 1 children (SEED Teen) of 60%, the COVID-19 Impact Study can anticipate that >1,100 participants will complete the data collection protocol (ASD: n ~= 300; DD: n ~= 480; POP: n ~= 360).

Summary

The following analyses should illustrate that an anticipated sample size of N = 1,100 provides strong statistical power (1-β >= .95) to answer a number of important research questions such as a) whether the number of health and behavior problems indicated at follow-up are greater in the ASD group relative to the DD and POP groups, b) whether these children experienced any changes in the ability to access or receive health and mental health services following the implementation of COVID-19 mitigation strategies (e.g., stay-at-home orders), and finally c) whether changes in health and mental health services following the implementation of COVID-19 mitigation strategies resulted in greater disruption of household structure, finances, or employment among families of children with ASD relative to families of children with DD or from the general population (POP).

Conventional definitions of effect size:

Test Type	**effect size**	**Small**	**Medium**	**Large**
t-test for means	Cohen's *d*	.20	.50	.80
t-test for correlation	*r*	.10	.30	.50
F-test for regression	f^2	.02	.15	.35
F-test for ANOVA	Cohen's *f*	.10	.25	.40
Partial Eta for ANOVA	η_p^2	.02	.06	.14
Chi-square	*w*	.10	.30	.50
Odds Ratio	*OR*	1.7	3.5	6.7
Prevalence Ratio	*PR*	1.2	1.9	3.0

Broad Research Aims

1. **Aim 1**: Between Group Differences: ASD, DD, POP

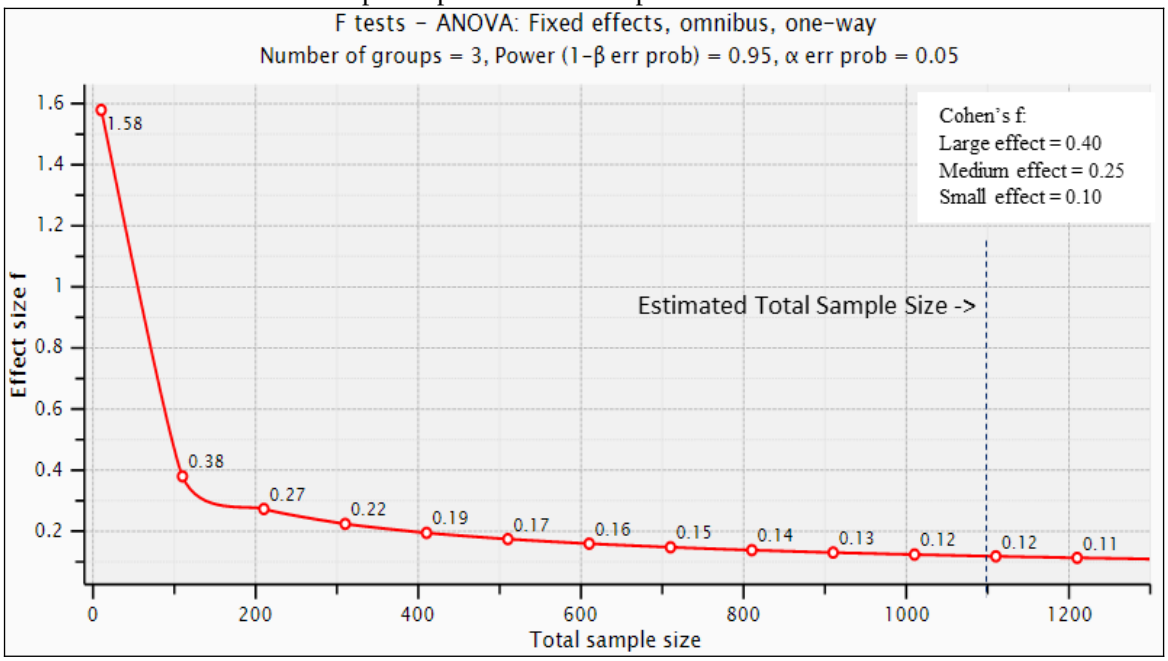
   Compare disruptions in out-of-school and in- school services; use of telemedicine/remote learning strategies; well-being and resilience activities between ASD, DD, & POP groups:
   - Computer and internet availability
   - Free or reduced lunch
   - Missed or delayed services, including general medical, other medial, mental health and preschool program services
   - Use and satisfaction with telehealth services
   - Activities, routines, and well-being
   - Individual Education Plan Progress

2. **Aim 2**: Within Group Comparisons - Pre to post-test changes, and changes over time periods for ASD, DD and POP groups
    a. Increases in sleep and problem behaviors, as measured through standardized assessments to include:
        - Internalizing behaviors t-score
        - Externalizing behaviors t-score
        - Sleep t-score
    b. Pre- and post-test differences in VABS scores among children with ASD
        - Adaptive Behavior Composite standard score
        - Communication standard score
        - Socialization standard score
        - Daily Living Skills standard score
    c. Retrospectively reported changes over time (January-February 2020 [pre-COVID], March-April 2020 [immediate post-COVID, Spring school term], June-July 2020 [Summer], and September-October 2020 [Fall school term] in:
        - Computer and internet availability
        - Free or reduced lunch
        - Missed or delayed services, including general medical, other medial, mental health and preschool program services
        - Use and satisfaction with telehealth services
        - Activities, routines, and well-being

3. **Aim 3**: Within-between group interaction – does the change from pre to post-test differ by group (ASD, DD, POP)?
   Assess whether the change or disruption of health and mental health services following the implementation of COVID-19 mitigation strategies resulted in:
    a. Greater increases in sleep and problem behaviors, as measured through standardized assessments to include:
        - Internalizing behaviors t-score
        - Externalizing behaviors t-score
        - Sleep t-score
    b. Differences across groups in retrospectively reported changes over time (January-February 2020 [pre-COVID], March-April 2020 [immediate post-COVID, Spring school term], June-July 2020 [Summer], and September-October 2020 [Fall school term]
        - Computer and internet availability
        - Free or reduced lunch
        - Missed or delayed services, including general medical, other medial, mental health and preschool program services
        - Use and satisfaction with telehealth services
        - Activities, routines, and well-being

Proposed Statistical Analyses for Aim 1 (between group differences)

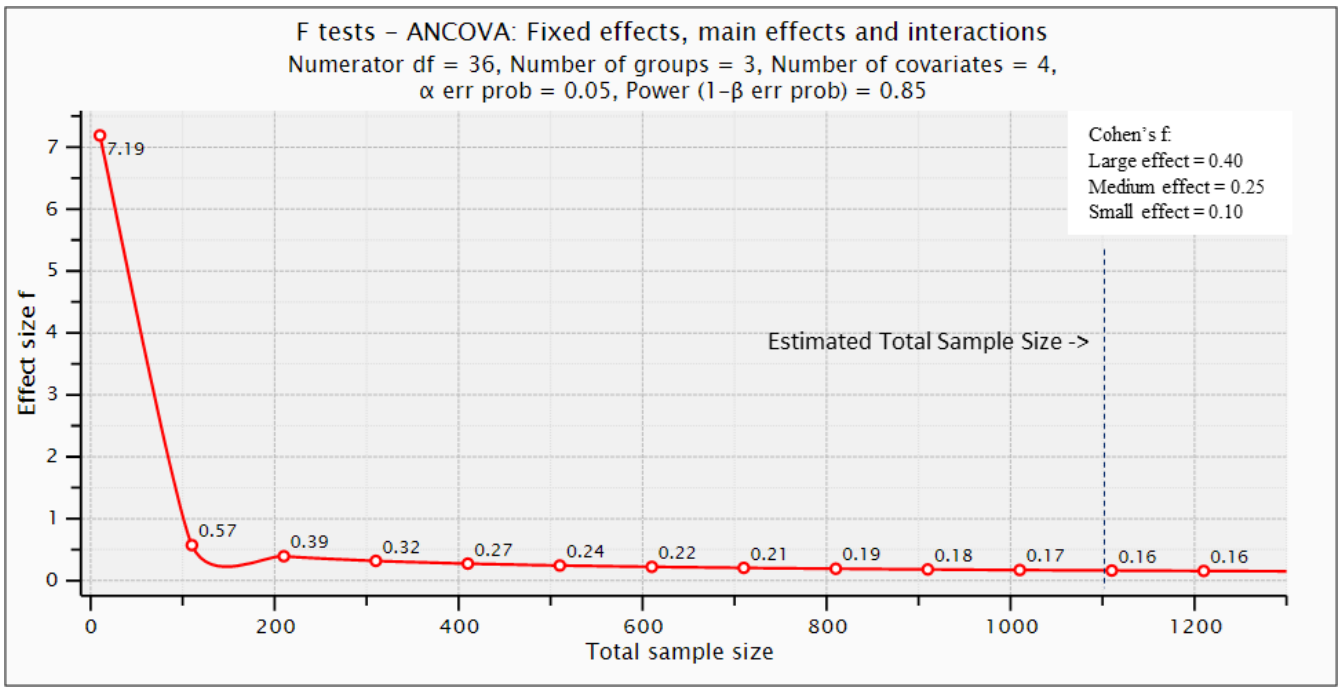
1. Proposed Tests for Continuous Outcomes (e.g., CBCL Internalizing/Externalizing behaviors t-score)
    a. Parametric Tests
        i. One-way ANOVA (3 group comparison)
            1. A sample size of 1,100 provides good statistical power (1-β = .95) to detect small but meaningful differences between three groups (Cohen's f = 0.10), and strong statistical power (1-β = .99) to detect small to medium effects (Cohen's f = 0.12)
                a. Note: The figure below illustrates the sample size needed to detect small but significant ($p < 0.05$) between-group differences among ASD, DD, & POP participants at follow-up



            2. Note: For two-group comparisons, a sample size of 600 (e.g., ASD: n=300, POP: n =300) provides good statistical power (1-β = .95) to detect between-group differences with a small to medium effect size (Cohen's f = 0.16)

ii. ANCOVA (3-group comparison adjusting for demographics)

1. A sample size of 1100 provides good statistical power (1-β = .85) to detect a between group difference with a small to medium effect size (Cohen's f = 0.16) after adjusting for covariates (e.g., education level, income, insurance, child's sex)



b. Non-parametric tests

i. Mann-Whitney U test

1. Assuming a non-normal distribution and a slightly higher proportion of POP relative to ASD participants, the sample size needed to achieve strong statistical power (1-β = .95) capable of detecting a between-group difference (e.g., ASD vs. POP) with a small to medium effect size (Cohen's d = .35), would be approximately 400 participants (ASD: n = 175; POP: n = 225).

2. Proposed Tests for Categorical Outcomes (e.g., computer/internet availability – yes/no)

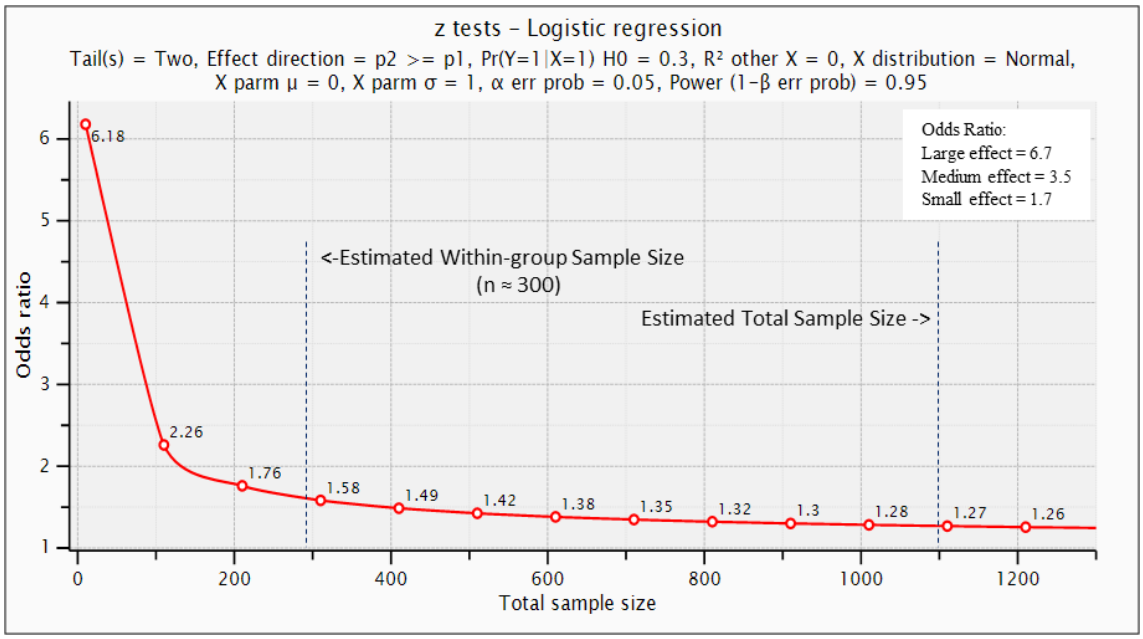
a. Binary Variables

i. Chi-square

1. A sample size of 1100 provides good statistical power (1-β = .85) to detect a small effect of w = 0.10.

ii. Logistic Regression (OR)

1. Assuming a population occurrence for a binary outcome is 30% (0.3) and an observed occurrence is 40% (0.4), a sample size of 1100 provides good statistical power (1-β = .95) to detect an odds ratio of OR = 1.27 and an alpha level of $\alpha \leq 0.05$



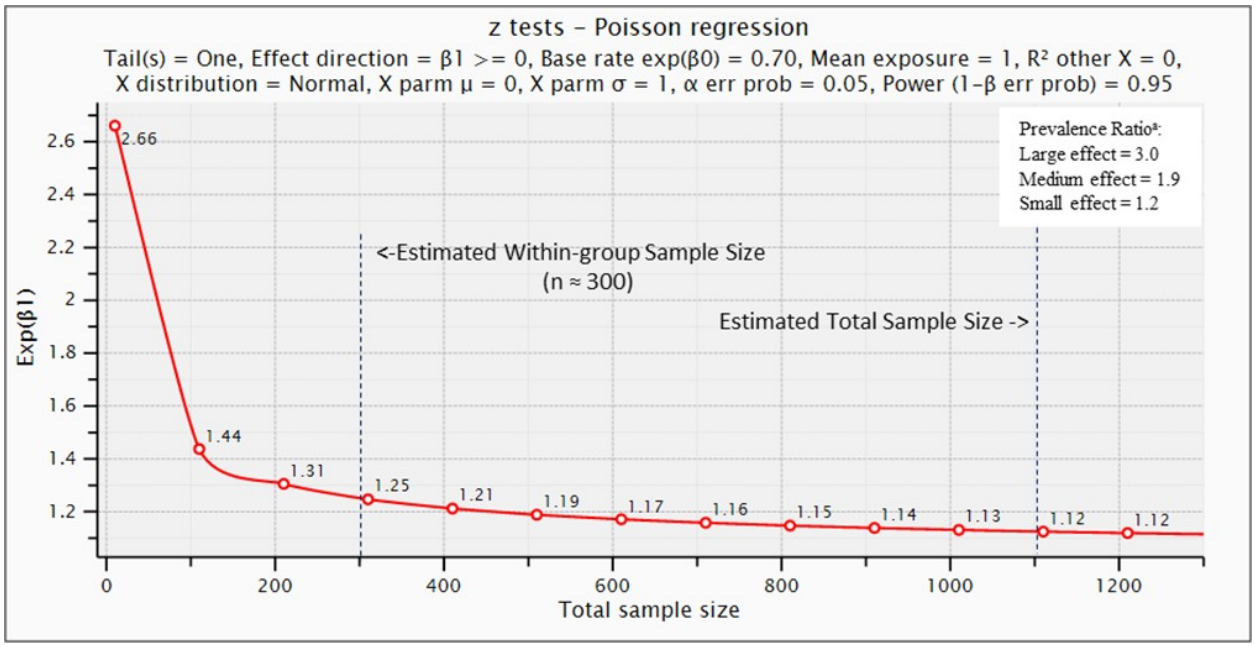
b. Count Variables

i. Poisson Regression (PR) (count variable – e.g., number of missed or delayed services)

1. Example: Determine the sample size needed to achieve good statistical power (1-β = .95) and minimum effect size with an alpha level of $\alpha \leq 0.05$ that would indicate a significant higher prevalence in the ASD relative to the POP group at follow-up.

a. In order to achieve good statistical power (1-β = .95) and detect a significant ($p < 0.05$, one-tailed) prevalence ratio of PR = 1.17, a total sample size of 600 would be needed.

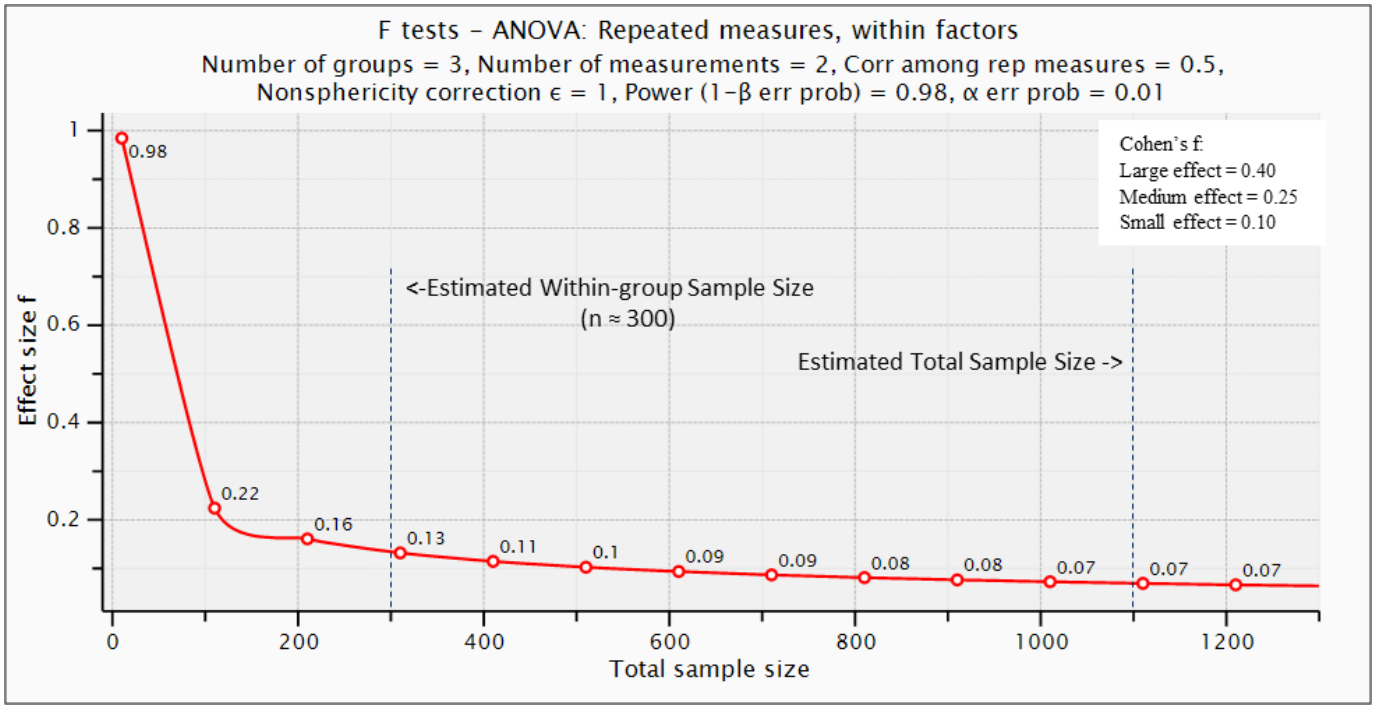
i. Note: this example represents a hypothetical situation where the researcher is interested in comparing the prevalence of a binary outcome in the ASD relative to the POP group, and assumes relatively equal number of ASD and POP participants (e.g., ASD: n = 300, POP: n = 300)



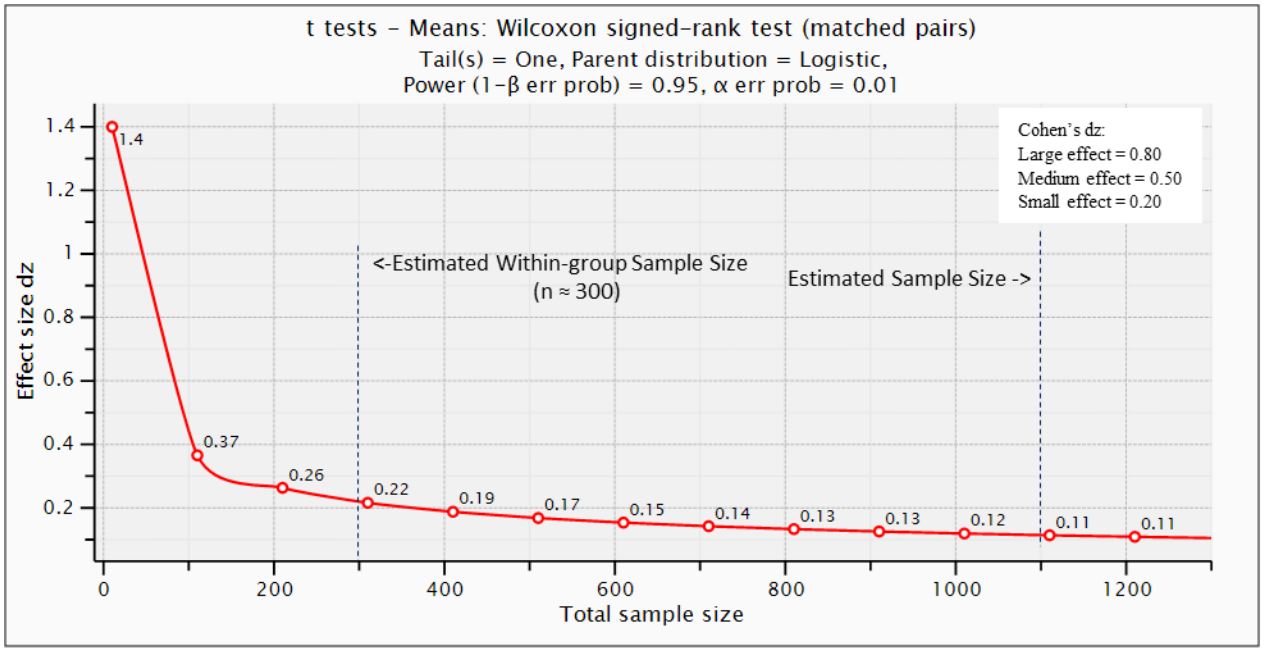
[a] Magnitude of effect size (i.e., small, medium large) for prevalence ratio assumes outcome is relatively rare event and that the sample size of each study group (e.g., ASD, DD, & POP) are equivalent

Proposed Statistical Analyses for Aim 2 (within-group comparisons)

1. Proposed Tests for Continuous Outcomes (e.g., changes in CBCL scores pre- vs. post-test)
    a. Parametric Tests
        i. Repeated measures ANOVA
            1. A sample size of 1100 provides strong statistical power (1-β = .98) to detect a small but meaningful pre-post change in a continuous outcome (Cohen's f = 0.07)
            2. Note: this example illustrates the sample size needed to detect a relatively small but significant ($p < 0.05$) effect of time (pre-post) across the entire sample, regardless of final study classification



b. Non-parametric test
    i. Wilcoxon signed-rank test
        1. Assuming a non-normal distribution, a sample size of 1100 provides strong statistical power (1-β = .95) to detect a small but meaningful pre-post change in a continuous outcome (dz = 0.11)



Note: Effect size 'dz' is equivalent to Cohen's d: small = 0.20, medium = 0.50, large = 0.80, the only difference being the continuity correction needed when using repeated measures

2. Proposed Tests for Changes in Categorical Outcomes (e.g., proportion of children needing free or reduced lunch pre- vs. post-test)
    a. Binary
        i. McNemar's tests
            1. Assuming equal sample sizes between groups (e.g., ASD: n = 300; POP: n =300), a total sample size of 600 would provide good statistical power (1-β = .94) to detect an odds ratio of OR ≥ 1.6 at an alpha level of α = 0.04.
        ii. Poisson Regression
            1. Assuming the prevalence of a specified outcome is 70% at pre-test, a sample size of 1100 provides good statistical power (1-β = .95) to detect an increase of 15% (PR = 1.15) with an alpha level of α ≤ 0.01 at post-test.

Proposed Analyses for Aim 3 (within-between interaction)

1. Proposed Tests for Continuous Outcomes
    i. Parametric Tests
        1. Repeated measures ANOVA within-between interaction
            a. A sample size of 1100 provides strong statistical power (1-β = .99) to detect a small time (pre-post) by group (ASD, DD, POP) interaction (Cohen's f = 0.08) with an alpha level of α = 0.01.
            b. Note: This example illustrates the sample size needed to detect a relatively small but significant time (pre-post) by group (ASD, DD, POP) interaction.