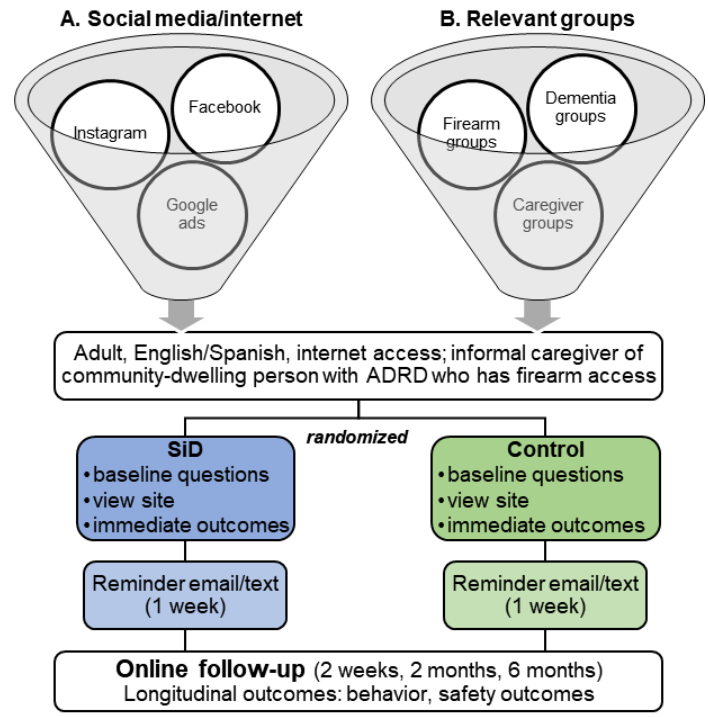
Project Title: "Safety in Dementia": An Online Caregiver Intervention
NCT Number: NCT05173922
Version Date: 5/6/24

STATISTICAL DESIGN AND POWER

Study Design: We will use an online randomized controlled trial of informal caregivers of community-dwelling individuals with ADRD and firearms at home (n=500), with longitudinal follow-up (**Figure 4.3.a**). We propose a parallel design: SiD vs web control (two arms). Randomization will occur at the level of the individual participant. To reduce selection bias, we will use block randomization stratified by recruitment groups (social media/internet and via relevant organizations). We will assess outcome measures immediately after the intervention (Aim 1a) and longitudinally (Aim 1b).

Overall Strategy: We will continue following recent published guidelines on statistical analyses in clinical trials.[1] Analyses will be performed according to the principle of intention-to-treat, including all randomized participants. Unless otherwise specified, hypothesis tests will be two-sided with alpha=.05, with 95% confidence intervals or p values reported. All statistical analyses will be performed using R[2] or SAS version 9.4 software (SAS Institute Inc., Cary, N.C.). Descriptive statistics will be computed for baseline caregiver characteristics (including key covariates), initially reporting on differences between: (a) two intervention arms and (b) complete follow-up vs. dropouts. Covariates will be screened in bivariate analyses and included in multivariate analysis if related to the outcome at p<.2, associated with dropout, or hypothesized *a priori* to be adjusted for. If there is evidence that normality assumptions were violated, we will use appropriate transformations, or the appropriate link function (e.g., logit link for dichotomized measures). We will employ general (or generalized) linear mixed models (GLMMs) to incorporate data structures that are longitudinal.[2,3] Analyses of SiD's longitudinal effects may be vulnerable to bias, as control arm participants may be exposed to intervention arm messages through exposure to other publicly available materials. We will assess exposure to other sources of information through structured questions. To mitigate the impact of this potential bias, we will adjust for contamination as applicable. Goodness of fit statistics and model fitting diagnostics will be used to assess for influential points, outliers, and heteroscedasticity, and to evaluate alternative model specifications. For Aim 2, data will come from the participants enrolled in the Aim 1 trial, including the longitudinal follow-up. As shown in **Figure 4.3.a**, trial enrollment will be stratified by two general methods: (A) social media and internet (Instagram, Facebook, and Google ads) and (B) relevant groups (dementia, firearm, and caregiver groups). The use of duplicated, identical sites for eligibility screening and for SiD review, each with a unique URL, will allow us to know a participant's recruitment group (A versus B), subgroup (e.g., Instagram versus Facebook), and language (English versus Spanish).

Figure 4.3.a. Study flow. Eligibility and SiD websites will have separate URLs (with identical content) to allow comparison (Aim 2).



Missing Data: Substantial efforts will be made to minimize missing data in data collection. We expect minimal missing in the primary outcome since this will be collected at enrollment. To minimize missing for longitudinal outcomes we will, based on our prior successes, use multi-modal outreach for follow-ups, including email, text, and telephone outreach to non-responders. Prior to beginning analyses, we will examine the data carefully to determine whether patterns of missing seem ignorable (MCAR or MAR) or non-ignorable (MNAR).[5–7] For longitudinal analyses we will use likelihood–based methods that utilize all available data through GLMMs, adjusting for covariates that are associated with missingness. If missingness seems non-ignorable we will consider, as data permits, pattern mixture models.[5,8] We will also examine in sensitivity analyses change scores as outcomes, adjusting for baseline.

Analysis:

Aim 1: To test the efficacy of SiD on firearm safety decision quality and behaviors, among a national sample of informal caregivers of community-dwelling people with ADRD and firearm access (n=500).

Hypothesis 1a: SiD is significantly associated with higher immediate decision quality, when compared to control. Our primary outcome measure will be the mean Preparation for Decision Making score at baseline, after viewing SiD or control. We will utilize multiple linear regression to test SiD's effect on decision quality. We will consider adjustment for baseline characteristics (e.g., age, gender, caregiver burden, severity of ADRD). Secondary analyses will be conducted similarly and will use as outcomes components of decision quality like decision self-efficacy and knowledge. Linear mixed models (LMM) will also be used to assess changes in self-efficacy and knowledge through time by treatment arm, allowing an interaction of treatment arm and time.

Hypothesis 1b: SiD is significantly associated with taking action to reduce firearm access, when compared to control. For access to firearms and actions to reduce access, we will use GLMM to account for correlations across measurements from the same individual, with adjustment for baseline characteristics as needed. We will test for an interaction between intervention and time to assess if intervention's effect varies over time. Secondary analyses will examine longitudinal change in caregiver burden and reported safety concerns.

Heterogeneity of treatment effect: Differential effect of treatment will be assessed by caregiver age group, whether the caregiver lives with the person with ADRD (versus not), and recruitment group (see Aim 2). GLMM models will be used including an interaction between the intervention variable (SiD versus control) and the factor of interest (e.g., age of caregiver, dichotomized into young and older adult). Exploratory heterogeneity of treatment effect will also be conducted by caregiver gender, urban versus rural residence, and language (English versus Spanish).

Aim 2: To compare varied methods in reaching informal caregivers.

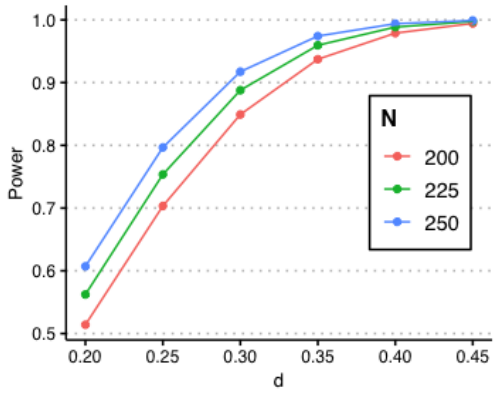
We will report descriptive statistics, using proportions (with 95% confidence intervals) and means (SD) or medians (with IQRs) for continuous variables. We will use t-test, Wilcoxon, and Chi-square test, as appropriate for pairwise statistical comparisons between and across groups and subgroups. We will use appropriate statistical methods to compare more than two groups at a time including ANOVA, Kruskal Wallis and Chi-square tests.

Hypothesis 2a: Caregivers reached via social media/internet will differ from those reached via relevant organizations. For our initial analyses, we will compare participant characteristics between group A (social media/internet) and group B (relevant groups), with subgroup analysis by language (English versus Spanish). We will then further divide analyses by the recruitment subgroup (e.g., Facebook versus Instagram).

Hypothesis 2b: Reach via social media/internet (versus relevant organizations) will be faster and lower cost. Using website analytics from the eligibility screening websites, we will calculate (by recruitment group and subgroup, including separately by language) the proportion of individuals who were eligible, the proportion of eligible individuals who enrolled, and the rate of recruitment. We will calculate the cost of recruitment per participant based on advertising costs, fees paid to organizations, and rough estimates of study staff time.

Sample size and Power: We have powered this study conservatively for the primary outcome to allow comparisons by recruitment groups and other participant characteristics. The large sample size will also protect against uncertainty in the estimate of the effect size given the small sample size (n=15) in our pilot study. In our pilot trial of 15 caregivers, the mean Preparation for Decision Making scores was 3.9 (SD=0.7) after SiD versus 3.6 (SD=0.2) in the control group; this corresponds to a standardized effect size of 0.42 (3.9 versus 3.6 with a pooled SD of 0.7; d=0.3/0.7=0.42)[4]. Based on t-test, with the proposed sample size of N=500 participants, we will be able to detect a d=0.42 effect size with very high power of >99% (alpha=0.05) between SiD and control for the primary outcome (preparation for decision making; **Figure 4.3.b**. Even if the effect is smaller (e.g., d=0.30) and the sample size is 80% (400 total) the power remains high at 84.9%. With a sample size of 500, we will have 125 in each of the final four analytic groups (stratified by recruitment group A versus B, SiD versus control). This sample will yield 80% power to detect any subgroup-specific standardized effect of d=0.36 or greater between SiD and control,

Figure 4.3.b. Power vs effect size (d), by N1. Alpha=0.05; N2=N1; 2-sided



including by recruitment group and other participant characteristics (using t-test and assuming sample sizes of 125 in each subgroup). Power calculations were computed using R version 4.0.2.[5]

References

1. Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, Williamson PR, Altman DG, Montgomery A, Lim P, Berlin J, Senn S, Day S, Barbachano Y, Loder E. Guidelines for the content of statistical analysis plans in clinical trials. JAMA. 2017 Dec;318(23):2337–43.
2. R: A language and environment for statistical computing. [Internet]. R Core Team. Available from: https://www.R-project.org/
3. Diggle P, Heagerty P, Liang KY, Zeger S. Analysis of Longitudinal Data. Oxford, United Kingdom: Oxford University Press; 2002.
4. Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G. Handbooks of Modern Statistical Methods: Longitudinal Data Analysis. Boca Raton, FL: Chapman & Hall/CRC; 2009.
5. Fairclough DL. Design and Analysis of Quality of Life Studies in Clinical Trials. New York, NY: Chapman and Hall/CRC; 2010.
6. Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. J R Stat Soc Ser C Appl Stat. 1994;43(1):49–73.
7. Hogan JW, Roy J, Korkontzelou C. Handling drop-out in longitudinal studies. Stat Med. 2004 May;23(9):1455–97.
8. Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. Psychol Methods. 1997;2(1):64–78.
9. Cohen J. Statistical power analysis for the behavioral sciences. Mahwah, NJ: Lawrence Erlbaum Associates; 1988.
10. PASS 2020: Power Analysis and Sample Size Software. [Internet]. NCSS, LLC; 2020. Available from: https://www.ncss.com/software/pass/