Pandemic-Related Analysis Challenges : The Duo PACT Study

Tor Neilands, PhD
Methods Core
UCSF Center for AIDS Prevention Studies

Adapting HIV and Other Interventions to Distance Models and Interruptions in Times of COVID-19: A 4-Part Implementation Science Series

Session 3: Data Analysis Considerations for Studies Spanning the Pre-COVID and During-COVID Eras

November 17, 2020

Duo PACT Study

- The Duo Pact Study (R01 NR010187; Mallory Johnson, PI; NCT02925949) is a 2-arm RCT to test the efficacy of a couple-level social support behavioral intervention vs. an individual-level Life Steps control to improve engagement in HIV care measured by HIV virologic control (1 = yes; 0 = no) derived from lab tests.
- Sample: 300 individuals from 150 couples with at least one partner being a sexual or gender minority and having suboptimal HIV care engagement. Equally allocated to control and intervention groups.
- Assessments: Baseline and 3, 6, and 9-month follow-ups.
- Intervention: Delivered between baseline and 3-month assessments.
- Planned Primary Analysis : Time-averaged comparison of odds of VL control by arm across 3-9 months. Alternating logistic regression (ALR) to address clustering due to repeated measures and dyads.

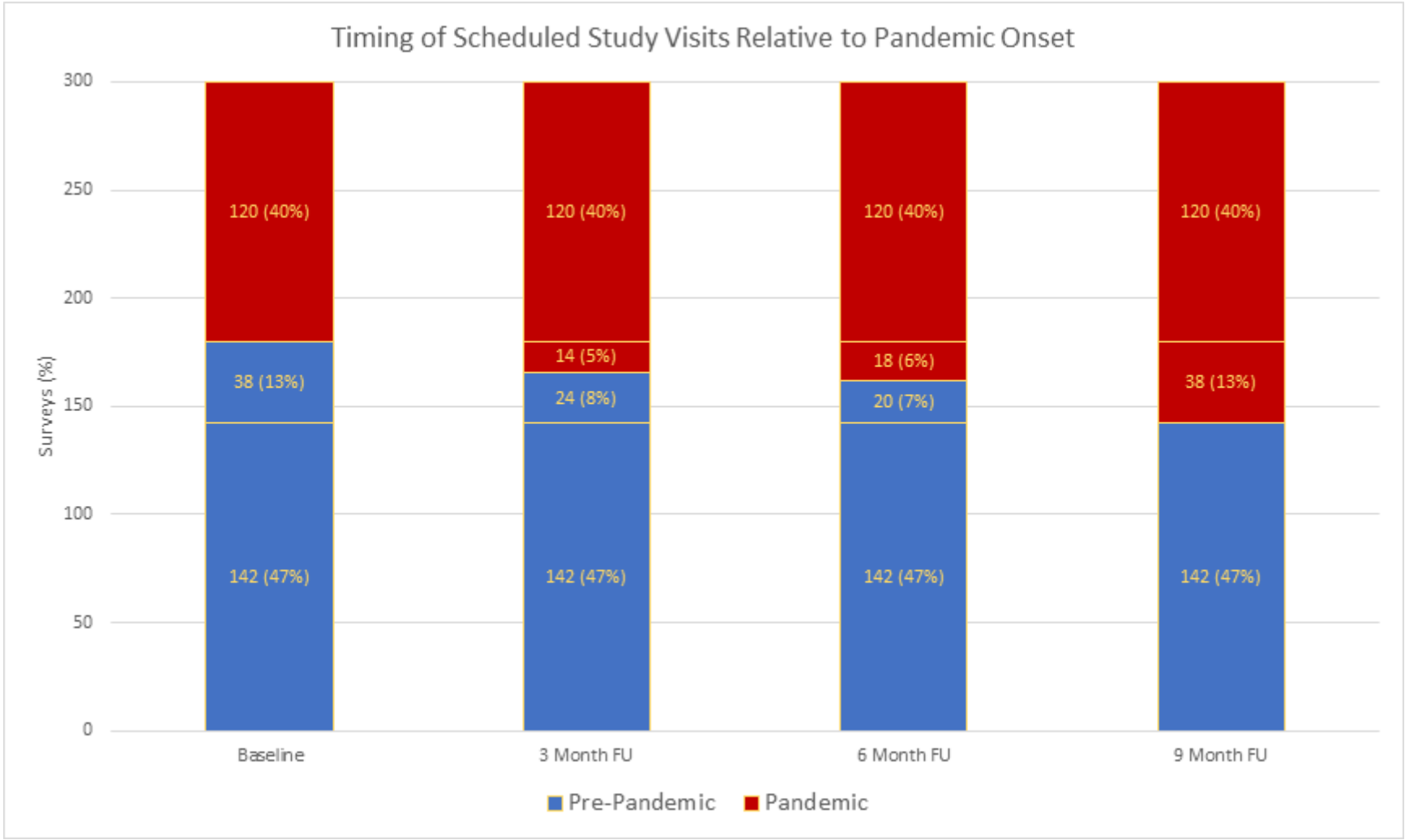
Pandemic-related Study Changes

- For Duo PACT, there were several changes to the study protocol resulting from the COVID-19 pandemic:
  1. Recruitment was halted on March 16th and resumed on July 1st. Upon resumption, recruitment was expanded from the San Francisco Bay Area to California-wide.
  2. The original study design featured enrollment, consent and baseline data collection in person with follow-up survey data to be collected online. All consent and baseline visits were changed from in-person to Zoom on July 1st.
  3. All control and intervention counseling sessions were changed from in-person to Zoom on March 16th. Anecdotally, attendance levels increased with Zoom, perhaps due to shelter-in-place boredom.
  4. The pandemic shut down lab data collection from mid-March until late May, at which point Quest had COVID safety procedures in place and lab data collection resumed.

Scheduled Visits

- For studies spanning the pre-pandemic and pandemic periods, it is useful to summarize scheduled visits broken down by pandemic status. As described in Steve's presentation, this summary will help us to:
  - Evaluate whether to make further protocol changes
  - Ascertain what is possible analytically
- Since Duo PACT data collection is still ongoing, we will use scheduled visits to differentiate pre-pandemic vs. during-pandemic time periods in the slides that follow.
- After data collection is concluded, we will substitute actual visits for scheduled visits to differentiate the pre-pandemic period from the pandemic period.

Scheduled Visits



Duo PACT Study Pandemic Response

- Assuming the scheduled visits projected in the previous slide are ultimately completed, Duo PACT will have collected sufficient amounts of data during the pre-pandemic and pandemic periods to explore pre-pandemic vs. pandemic era differences.
- Consequently, the Duo PACT study team plans to continue collecting data as originally planned, following the schedule shown on the previous slide.
- To explore pre-pandemic vs. pandemic differences, we will examine whether our planned time-averaged comparison differs across the pre-pandemic vs. pandemic time periods.

Duo Pact Analysis Plan

- The comparison will be conducted as a planned contrast based coefficients obtained from a model containing randomization group, time, and the group-by-time interaction.
- We will include a time-varying pandemic status indicator variable as a main effect and the following interaction terms:
  - Group-by-pandemic status
  - Time-by-pandemic status
  - Group-by-time-by-pandemic status
- We will follow a backward elimination strategy to remove non-significant interaction terms followed by the pandemic status indicator.
- If any interactions containing pandemic status are significant, we will report results differentiated by pandemic status.

Duo Pact Analysis Plan (2)

- If no interaction effects are present, we will evaluate the significance of the main effect for pandemic status. If it is significant, we will retain it; otherwise we will drop it. We will then report results from a reduced model containing the group, time, and group-by-time effects (adjusting for the pandemic status indicator if it is significant).
- We will consider sensitivity analyses where the definition of pandemic status is varied. On Slide 5, n=38 (13%) of participants experienced some data collection during both time periods. For these 38 participants, sensitivity analyses could evaluate whether our results change if we pool this subgroup with each of the two larger groups. We will consider the following two pooling scenarios:
  - Set the subgroup pandemic status to “pre-pandemic” (red -> blue)
  - Set the subgroup pre-pandemic status to “pandemic” (blue -> red)

What about Missing Data in Duo PACT?

- 109 participants completed all measurements pre-COVID (defined as prior to March 16th). There were 436 possible viral load (VL) measurements.
- Of those, 93 participants completed 407 (93%).
- During COVID, there are 47 participants with 82 possible VLs; of those, 37 had 68 VLs (83%).
- First adaptation strategy is to backfill missing VLs by requesting of medical information (ROI) from non-Duo PACT lab visits during March through May.
  - 18 ROIs sent; 12 returned; 3 with VL results; 3 pending
- How to address the remaining missing values?

Addressing Missing Data

- Cro et al (2020a) recommend performing the primary analysis assuming that incomplete data arise from a missing-at-random (MAR) missingness mechanism.
- Multiple imputation (MI) is a flexible and accessible method for addressing missing data under the MAR assumption in the analysis of clinical trial data.
- Including auxiliary variables in the imputation-generating phase of multiple imputation (MI) can increase the likelihood of meeting the MAR assumption.
- The original analysis plan for Duo PACT proposed multiple imputation (MI) assuming MAR missingness for the primary analysis, which is consistent with Cro et al's recommendation.

MI Variables in Duo PACT

- All variables and effects to be used in the analysis model(s), including group, time, pandemic status, and their interactions.
- Candidate auxiliary variables to include in the imputation model (but not the analysis model):
  - Demographic variables measured at baseline (e.g., race/ethnicity, age, gender identity, income, etc.)
  - Survey self-reports/patient-reported outcomes (PROs), including HIV medication adherence and attending HIV medical appointments
- Prefer not to use MI? Auxiliary variables can be included in likelihood-based (e.g., MLE) or Bayesian analyses.
  - Model specification can be more complicated
  - No need for imputing – analysis model addresses missing data automatically under the MAR assumption
  - SEM software can include auxiliary variables under MLE & Bayes

MAR-based MI Strategy for Duo PACT

- To address clustering by couple and person, first convert the data to "super wide" structure so there is one row per couple.
- Impute separately by randomization group to allow for all-possible interactions containing group (White et al., 2011).
- Include pre-pandemic vs. pandemic indicators (time-varying: one for each time point in the super wide data structure) and their interactions with the outcome variable at each time point.
- Include auxiliary variables such as demographics and patient self-reports as described on the previous slide. Follow recommended criteria for including auxiliary variables (Collins et al., 2001).
- The resulting imputations will account for all interactions involving randomization group, assessment time, and pandemic status.
- Reshape the imputed datasets back to "super long" format with one row per couple-person-time point for analysis.

Missing Data-based Sensitivity Analyses

- Cro et al (2020a) suggest optionally following the MAR-based primary analysis with sensitivity analyses which assume missing data arise from a not-missing-at-random (NMAR) mechanism.
- NMAR-based imputations may be most useful for situations where there are missing data unrelated to other variables with observed data (Allison, 2002). E.g., in a drug or device trial where missingness is due to the intervention only and may be unrelated to other variables.
- In social and behavioral studies like DUO Pact, it is often possible to help the analysis to meet the MAR assumption via inclusion of relevant auxiliary variables (Schafer and Graham, 2002).
- See Cro et al (2016; 2020b) and the appendix for resources to learn more about NMAR-based sensitivity analyses and available software options.

Summary: Pandemic Analytic Strategies

- Chart scheduled visits during the pre-pandemic and during-pandemic time windows to gauge whether any data collection protocol changes are needed and what will be possible analytically once all data are collected.
- Identify amounts and causes of missing data. Make all possible attempts to retrieve missing data (e.g., ROIs from medical visits; collect survey data by phone/online).
- Check for pre-pandemic vs. pandemic-era differences in key effects if the data support testing for those differences. Consider sensitivity analyses which vary the operationalization of pre-pandemic vs. pandemic status.
- Address missing data under the MAR assumption for the primary analysis. Include auxiliary variables to help meet the MAR assumption. If using MI, be sure to include all effects you will examine in your analysis models, including interaction terms.
- Optional: Follow the MAR-based primary analysis with NMAR-based missing data sensitivity analyses if they are relevant to your study.

References

- Allison P. *Missing Data.* Thousand Oaks, CA: Sage Publications; 2002.
- Collins LM, Schafer JL, Kam C-M. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods.* 2001;6(4):330-351.
- Cro S, Morris TP, Kahan BC, Cornelius VR, Carpenter JR. A four-step strategy for handling missing outcome data in randomised trials affected by a pandemic. *BMC Med Res Methodol.* 2020;20(1):208.
- Cro S, Morris TP, Kenward MG, Carpenter JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. *Stat Med.* 2020;39(21):2815-2842.
- Cro S, Morris TP, Kenward MG, Carpenter J. Reference-based sensitivity analysis via multiple imputation for longitudinal trials with protocol deviation. *Stata Journal.* 2016;16(2):443-463.
- Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychological Methods.* 2002;7(2):147-177.
- White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med.* 2011;30(4):377-399.

Acknowledgements

- Duo PACT Study Information:
  - Mallory Johnson, PhD
  - Lara Coffin
  - Sam Dilworth (including Figure shown on Slide 5)
- Extensive input on slides and multiple valuable discussions:
  - Steve Gregorich and Chuck McCulloch
- Reminder not to overlook data collection instrument programming and database management challenges imposed by a pandemic:
  - Lance Pollack
- Panel organizers:
  - Sheri Lippman & Wayne Steward
  - Rochelle Blanco

Appendix: NMAR Sensitivity Analyses

- Recommended MI-based NMAR sensitivity analyses are based on pattern-mixture models that relax the MAR assumption.
  - Delta-based MI: Use a constant δ to shift the value of an estimated effect (e.g., log odds of post-BL difference between Duo Pact and Life Steps). Vary δ across a range of plausible values to assess how the intervention effect changes.
  - Reference-based MI: Impute values based on a counterfactual progression of participants with missing values (e.g., impute missing values for Duo PACT intervention participants based on the Life Steps control group's trajectory).
- Delta-based MI is more general and can be used for any design whereas reference-based MI requires a reasonable reference comparator and is typically used for RCTs.
- Specifying values for δ is typically not easy. Reference-based MI avoids the need to specify specific values.

Appendix: NMAR Multiple Imputation Tools

- Stata reference-based imputation is available via the user-written command -mixmi- (Cro et al, 2016).
- Stata commands for delta-based MI are shown in Cro et al, 2020(b)
- SAS PROC MI in SAS/STAT 15.1 supports both delta and reference-based MI scenarios with the MNAR statement (e.g., see example 79.15 in the SAS documentation).
- Cro et al (2020b) provide a link to SAS macros for implementing additional reference-based MI scenarios. The published link has been supplanted by this page: https://www.lshtm.ac.uk/research/centres-projects-groups/missing-data#dia-working-group. It contains links to a number of potentially useful SAS macro and R programs for various designs and analyses (e.g., clustered data structures, survival models, negative binomial-distributed outcomes).