



Considerations for TPP Evaluations with Multiple Follow-Up Assessments

Technical Appendix

This appendix includes the following sections:

[A. Model specification details and data structures across analytic approaches](#)

[B. Note on linear growth models](#)

A. Model specification and data structure

The brief's discussions about decisions made in estimating impacts across time points will guide the user in choosing the appropriate regression model specification. This specification, in turn, will determine how to organize the data to fit the model (either in wide format, wherein data are organized with one row per individual, or in long format, wherein data are organized with one row per assessment per individual. Data structures are discussed in detail later in this section).

Model specification

For the purposes of this discussion, assume a researcher plans to account for within-person correlation in any longitudinal models (the fourth decision) through cluster-robust standard errors; at the end of this section, we discuss how the model changes when the researcher includes random effects.

In describing the model specification, we will use the following notation:

i : indexes the individual

t : indexes time

y_i , y_{it} : the outcome for individual i , at time t

y_i^0 : the baseline measure of the outcome for ANCOVA models

In DID, the baseline measure is indicated by y_{i0}

X_i : baseline covariates for individual i

T_i : treatment indicator for individual i

$time_t$: the continuous measurement of time corresponding to index t

$post_t$: indicator that time t corresponds to a follow-up time point

Decisions 1–3 determine the model specification as follows:

- Decision 1 determines whether the model is a DID or ANCOVA model.
1. DID models use the observed measures at both baseline and follow-up (y_{it}) as the outcome. Interactions with $post_t$ distinguish baseline from follow-up.



2. ANCOVA models use follow-up measurements (y_{it}) as the outcome and include the baseline measure (y_i^0) as a predictor
- Depending on the choice of model type described in Decision 1 (DID or ANCOVA), Decisions 2 and 3 suggest three potential model specifications:
 1. When modeling follow-up outcomes separately (Decision 2), Decision 3 (treating time as continuous or discrete) does not apply. In this case, the ANCOVA models (under Decision 1) are not longitudinal; they include only one time point per individual as an outcome. Thus, time does not play a role in these models, and a single treatment effect (represented by θ) is estimated. If under Decision 1 the DID approach is selected, again Decision 3 does not apply, as the metric of time does not play a role in the estimation. The separate analysis of each outcome under a DID approach means each model will depend on two observations (a given follow-up period and the baseline); the treatment-by-post period interaction effect (also represented by θ) is the target estimand.
 2. For longitudinal models that treat time as a discrete variable under Decision 3, we estimate a different time effect (represented by δ_t) and treatment effect (represented by θ_t) for each follow-up time.
 3. For longitudinal models that treat time as continuous in a linear growth curve, we estimate a linear outcome trend in the control group (represented by ζ) and a separate linear treatment effect (represented by θ). Researchers and evaluators can estimate such a model and use θ as the estimate of program effectiveness as a simple method to use their longitudinal data. Although this analytic approach will yield two different trend lines representing the treatment and control groups, θ does not indicate *when* the difference in treatment and control group means occurred. Therefore, researchers should not interpret it as a test of program effectiveness. In fact, it can be misleading, suggesting there is a treatment effect when impacts are small (and potentially in the wrong direction), during the post-intervention assessment period of interest. See section B for a graphical illustration of this.

Table A.1 lists the resulting model specifications, with additional detail on how to structure the data for a given model type (more detail on this in the next section).

Table A.1. Model specification and data structure implied by Decisions 1, 2, and 3

Decision 1	Decision 2	Decision 3	Model specification	Data structure
DID	Separate	n.a.	$y_{it} = \alpha + \beta X_i + \gamma T_i + \delta post_t + \theta T_i \times post_t + \epsilon_{it}$	Long
DID	Longitudinal	Discrete	$y_{it} = \alpha + \beta X_i + \gamma T_i + \delta_t + \theta_t T_i \times post_t + \epsilon_{it}$	Long
DID	Longitudinal	Linear growth	$y_{it} = \alpha + \beta X_i + \gamma T_i + \delta post_t + \zeta post_t \times time_t + \xi post_t \times T_i + \theta T_i \times time_t \times post_t + \epsilon_{it}$	Long
ANCOVA	Separate	n.a.	$y_i = \alpha + \beta X_i + \xi y_i^0 + \theta T_i + \epsilon_i$	Wide
ANCOVA	Longitudinal	Discrete	$y_{it} = \alpha + \beta X_i + \xi y_i^0 + \delta_t + \theta_t T_i + \epsilon_{it}$	Long, but with baseline measure as a column, not a row



Decision 1	Decision 2	Decision 3	Model specification	Data structure
ANCOVA	Longitudinal	Linear growth	$y_{it} = \alpha + \beta X_i + \xi y_i^0 + \zeta time_t + \gamma T_i + \theta T_i \times time_t + \epsilon_{it}$	Long, but with baseline measure as a column, not a row

n.a. = not applicable.

Including random effects changes the model specification slightly by replacing the ϵ_{it} terms in any of the longitudinal models of Table A.1 with a new error specification, $u_i + e_{it}$. In this specification, the terms u_i are individual-level random effects, assumed to come from one distribution (usually normally distributed), and the terms e_{it} are residuals that come from a separate normal distribution. As a reminder, random effects apply only to longitudinal models that have multiple observations per individual.

Data structure

Different model specifications require different ways to organize the data. The primary determinants of the necessary data structure are whether the model is a DID model (which determines whether to show the baseline measure of the outcome as a separate observation in the data set or as an extra variable in each observation), and whether follow-up measures are modeled separately or in a single longitudinal analysis.

To illustrate, consider a research study that assesses the impact of an intervention aimed at reducing the number of episodes of risky sex. To do this, researchers ask participants to report the number of episodes of risky sex they have had in the prior 30 days at three separate three time points—once pre-intervention (baseline_score), once six months post-intervention (post1_score), and once 12 months post-intervention (post2_score). The researcher also records the baseline age of the participant (bl_age), and whether the participant received the intervention (treatment). Table A.2 gives an example of a data set that organizes these data with one row per individual, and one column for all recorded data. The table uses “.” to designate missing values.

Table A.2. Example data in wide format

Id	bl_age	treatment	baseline_score	post1_score	post2_score
1	15	1	2	2	2
2	17	1	3	2	.
3	16	1	3	3	2
4	18	0	4	3	4
5	19	0	3	3	.

A wide format signifies data sets containing longitudinal data organized with one row per individual. In this format, it is easy to see there are five study participants and three follow-up measurements, some of which are missing.

This wide data structure is most appropriate to fit cross-sectional models, which contain one observation per individual. The model we have discussed that is cross-sectional includes an ANCOVA model that



assesses each follow-up time separately.

A message about missing data. In the absence of data imputation, most statistical packages use listwise deletion to handle missing data on outcomes and covariates in a given observation. As a result, researchers conduct analyses on the subset of individuals with complete data on both baseline and follow-up outcome assessments, and on any covariates included in the models. Therefore, in the example study data in Table A.2, the analytic sample used in Follow-Up 1 would contain a larger set of individuals than the sample used in Follow-Up 2.

Longitudinal models typically require data to be in long format, with one row for every outcome assessment as an observation in the model, rather than having a separate observation for every individual. In the working example, each individual in the transformed long format data set would have three observations: one representing the baseline outcome score, one representing the six-month follow-up assessment score, and one representing the 12-month follow-up assessment score. In this case, a variable called “time” will index the different assessment points, and the value of the measurement at the corresponding visit will serve as the outcome score. The participant identifier, baseline covariates, and baseline measurement will all appear in separate columns; these columns will have the same value for all observations of a given individual. In addition, to simplify discrete time analyses under Decision 3, we present the continuous time variable as a baseline indicator (bl_i) and two follow-up indicators (post1_i and post2_i).

Note: In longitudinal analyses that use an ANCOVA approach in Decision 1, the baseline assessment will appear as a column of data, analogous to “bl_age”, rather than as a separate row for each individual; this helps differentiate the long data structure with baseline as a column, not a row, as presented in Table A.1.

Table A.3 organizes the same example data in long format.¹

Table A.3. Example data in long format

id	bl_age	treatment	visit	time	bl_i	post1_i	post2_i	outcome
1	15	1	0	0	1	0	0	2
1	15	1	1	6	0	1	0	2
1	15	1	2	12	0	0	1	2
2	17	1	0	0	1	0	0	3
2	17	1	1	6	0	1	0	2
2	17	1	2	12	0	0	1	.
3	16	1	0	0	1	0	0	3
3	16	1	1	6	0	1	0	3
3	16	1	2	12	0	0	1	2
4	18	0	0	0	1	0	0	4
4	18	0	1	6	0	1	0	3
4	18	0	2	12	0	0	1	4
5	19	0	0	0	1	0	0	3
5	19	0	1	6	0	1	0	3
5	19	0	2	12	0	0	1	.

¹ Standard statistical packages, such as STATA, SAS and R have procedures to quickly convert data files from wide to long (and vice versa).



Additional notes on model fitting

The exact syntax to fit regression models will depend on the software being used. However, we do make two additional notes related to model fitting and estimation:

- 1. Indices for time:** The specifications for longitudinal models that treat time as discrete include some parameters indexed by time, such as δ_t and $\theta_t T_i$. This means there is one value of each parameter for every time point. The easiest approach to fitting models with these terms will depend on the software used for model fitting. However, a solution that is relatively transportable across software packages is to create indicator or dummy variables for each time point, and directly include these predictors in the model. More specifically, we note that we can express δ_t and $\theta_t T_i$ as follows:

$$- \delta_t = \sum_s I_{\{t=s\}} \delta_s$$

$$- \theta_t T_i = \sum_s I_{\{t=s\}} \theta_s T_i$$

where $I_{\{t=s\}}$ has the value of 1 for observations that have $t = s$, and 0 otherwise. For example, in the situation described in Table A.3, if a researcher wants to fit a longitudinal DID model with discrete time, they can first create variables `bl_i`, `post1_i`, and `post2_i` that have the value 1 for observations at the corresponding time and 0 for all other times. They can then fit the model described in Table A.1, replacing δ_t with $\delta_1 \text{bl}_i + \delta_2 \text{post1}_i + \delta_3 \text{post2}_i$ and $\theta_t T_i$ with $\theta_1 \text{bl}_i \times T_i + \theta_2 \text{post1}_i \times T_i + \theta_3 \text{post2}_i \times T_i$. Note further that in this case the term $\delta_1 \text{bl}_i$ will drop out of the model, as it is not identifiable due to the intercept.

- 2. Estimating point-in-time effects:** As mentioned, growth curve models do not directly estimate the point-in-time impacts of the intervention. However, we can calculate this as a linear combination of the model coefficients. For example, in the ANCOVA model with linear growth, $\gamma + \theta \times \text{time}_t$ gives the treatment effect at any time t . Thus, researchers who want to estimate the treatment effect at six months post intervention would want to calculate $\gamma + 6 \times \theta$. This is an important reporting consideration for growth models: many audiences (including the TPPER) will expect to see point-in-time estimates of program effectiveness as a means to document the magnitude of the effect of a program at a well-defined time point.

B. Note on linear growth models

This section offers a graphical representation of why focusing on the time*treatment interaction (θ , the equation below) as the sole estimate of program effectiveness in a linear growth model has the potential to yield an incorrect interpretation.

As a reminder, we use the following notation when describing model specification:

i : indexes the individual

t : indexes time

y_i, y_{it} : the outcome for individual i , at time t

y_i^0 : the baseline measure of the outcome for ANCOVA models

In DID, the baseline measure is indicated by y_{i0}

X_i : baseline covariates for individual i

T_i : treatment indicator for individual i

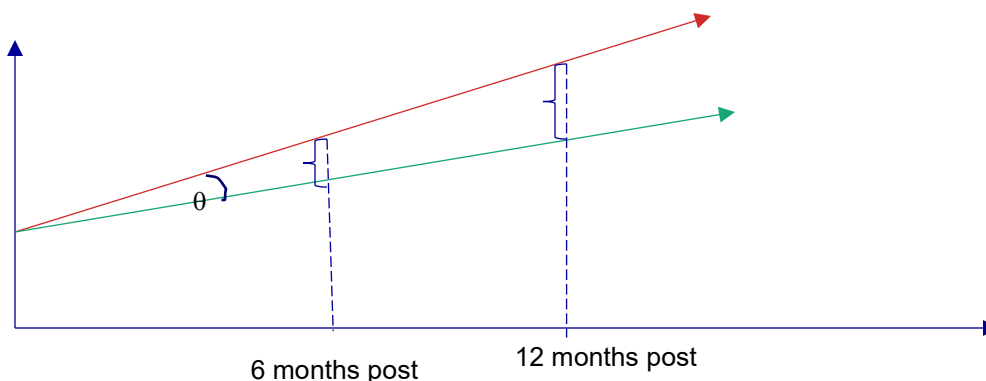
$time_t$: the continuous measurement of time corresponding to index t

$$(A.1) y_{it} = \alpha + \beta X_i + \gamma T_i + \zeta time_t + \theta T_i \times time_t + \epsilon_{it}$$

Although this interaction can show differences in trends (or slopes) between treatment and control groups, it can also mask exactly where this difference in outcomes (the true impact estimate) occurs.

Consider a hypothetical example study that examines the effect of an intervention on sexual initiation rates, with survey assessment periods at baseline and 6- and 12-months post program. Assume the evaluator conducting this study estimates a simple growth trend model for this study using Equation A.1. Under this assumption, in Exhibit A.4, the green line represents the linear trend for the treatment group and the red line represents the trend of the control group. Looking only at this subset of the graph, we could infer the program attenuates initiation at both follow-up points, and the magnitude of the impact increases over time. Therefore, θ seems to indicate this program is effective during the intervention period at slowing initiation rates over time.

Exhibit A.4. Illustrative graph

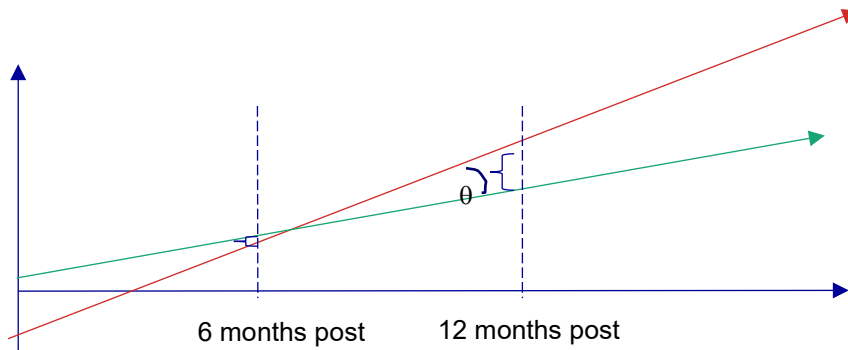


Consider an alternate presentation in Exhibit A.5, in which we have an identical value for θ as Exhibit A.4. The difference in the slopes between the treatment and control groups is exactly the same. In this situation, however, the interpretation of program effectiveness is far less clear. Now, it looks like the program actually has a small unfavorable impact at six months post, with higher initiation rates among the treatment group. Although the control group has overtaken the treatment group in terms of initiation at 12 months, the magnitude of the impact is smaller than what we saw in Exhibit A.4 at this same time. Therefore, although the parameter estimate is identical in these two exhibits, the interpretation of the

effect of the program is markedly different.

The main difference in Exhibit A.4 compared to A.5 is there were underlying baseline differences between the treatment and control groups that complicated interpreting the θ term as representing a program impact. This illustration aims to highlight why focusing solely on the time*treatment interaction term (θ) presents an incomplete and inaccurate story of the effect of the program.

Exhibit A.5. Illustrative graph



An important takeaway of this section and its illustrations is to highlight that when fitting a functional form for time, it is also important to supplement it with point-in-time impact estimates and to ensure the analysis demonstrates the treatment and control groups are equivalent at baseline.