# Supplementary Appendix: Difference-in-Differences with a Continuous Treatment

Brantly Callaway* Andrew Goodman-Bacon† Pedro H.C. Sant'Anna‡

January 26, 2024

This supplementary appendix provides a number of additional results for "Difference-in-Differences with a Continuous Treatment". Appendix SA contains a full discussion of the setting with multiple time periods and variation in treatment timing and dose intensity and expands upon the results provided in Appendix D in the main text. This section also provides a number of results about interpreting TWFE regressions in the multiple-period setting. Appendix SB provides proofs for all the results in the main text and in the supplementary appendix concerning multiple periods and variation in treatment timing and dose intensity. Appendix SC provides results and proofs for a number of additional results that were discussed in the main text: results for settings with no untreated units; additional results for TWFE decompositions with a continuous treatment; and TWFE decompositions with a multi-valued discrete treatment. Appendix SD provides a discussion and a proof for Theorem C.1 in the main text, which provided a comparison between different versions of parallel trends assumptions as well as characterized their relationship to restrictions on treatment effect heterogeneity. Finally, Appendix SE provides results on relaxing the strong parallel trends assumption, which was briefly discussed in Section 5.1 in the main text.

## SA Additional Details for Multiple Periods and Variation in Treatment Timing and Dose

In Section 5.2 and Appendix D in the main text, we briefly discussed some results concerning difference-in-differences with a continuous treatment when there are multiple time periods and variation in treatment timing across units. This section provides a full treatment of that setting.

### SA.1 Target Parameters

Following the discussion in the main text, we mainly consider identifying the disaggregated parameters $ATT(g,t,d|g,d)$, $ATE(g,t,d)$, $ACRT(g,t,d|g,d)$, and $ACR(g,t,d)$ which are all defined in Appendix D. In the main text, we also introduced more aggregated parameters that are easier to

*University of Georgia. Email: brantly.callaway@uga.edu
†Federal Reserve Bank of Minneapolis and NBER. Email: andrew@goodman-bacon.com
‡Emory University. Email: pedro.santanna@emory.edu

report and estimate than the fully disaggregated parameters previously mentioned. In particular, we discussed the identification of $ATE^{dose}(d)$ and $ACR^{dose}(d)$, which averaged $ATE(g,t,d)$ and $ACR(g,t,d)$ across timing groups and time periods to deliver summary parameters that were solely a function of the dose. We also discussed how to average these into scalar summary parameters, $ATE^o$ and $ACR^o$. Finally, we discussed the identification of the event-study parameters $ATE^{es}(e)$ and $ACR^{es}(e)$, which average across dose and partially average across timing groups and event times to deliver average treatment effects and average causal responses as a function of event-time. All of the identification results were in the context of a version of strong parallel trends in Assumption 5-MP.

In this section, we start by introducing "on-the-treated" versions of the summary causal effect parameters discussed in the previous paragraph. To start with, consider the summary parameter

$$ATT^{dose}(d|d) = \mathbb{E}\Big[\overline{TE}(d)\Big|D = d, G \leq T\Big]$$

which (recalling that, among units that participate in the treatment in any time period, $\overline{TE}_i$ was defined as the average treatment effect across all post-treatment periods for unit $i$ in the main text) is the average treatment effect of dose $d$ across post-treatment periods among those that experienced dose $d$. We can likewise similarly define a causal response parameter

$$ACRT^{dose}(d|d) = \frac{\partial ATT^{dose}(l|d)}{\partial l}\bigg|_{l=d}$$

$ATT^{dose}(d|d)$ and $ACRT^{dose}(d|d)$ are analogs of $ATT(d|d)$ and $ACRT(d|d)$ considered in the main text in the case with two time periods. Following the same line of thinking as in the main text, we can also further aggregate these parameters into scalar summary parameters:

$$ATT^o = \mathbb{E}\Big[ATT^{dose}(D|D)\Big|G \leq T\Big] \quad \text{and} \quad ACRT^o = \mathbb{E}\Big[ACRT^{dose}(D|D)\Big|G \leq T\Big]$$

$ATT^o$ and $ACRT^o$ are fully aggregated parameters that can summarize "on-the-treated" level effects and causal responses, respectively; they are analogous to the homonymous parameters discussed in the two-period case considered in the main text. Following the same line of argument as in the main text, next we show that $ATT^{dose}(d|d)$ can be expressed in terms of underlying $ATT(g,t,d|g,d)$ parameters. To see this, notice that

$$\begin{aligned} ATT^{dose}(d|d) &= \mathbb{E}\Big[\overline{TE}(d)\Big|D = d, G \leq T\Big] \\ &= \sum_{g \in \bar{\mathcal{G}}} \frac{1}{T-g+1} \sum_{t=2}^{T} \mathbf{1}\{t \geq g\} \mathbb{E}\Big[Y_{i,t}(g,d) - Y_{i,t}(0)\Big|G = g, D = d\Big] \mathbb{P}(G = g|D = d, G \leq T) \\ &= \sum_{g \in \bar{\mathcal{G}}} \sum_{t=2}^{T} \omega^{dose}(g,t,d) ATT(g,t,d|g,d) \end{aligned}$$

where $\omega^{dose}(g,t,d) = \frac{\mathbf{1}\{t \geq g\}}{T-g+1}\mathbb{P}(G = g|D = d, G \leq T)$ and where the first equality comes from the definition of $ATT^{dose}(d|d)$, the second equality holds by the definition of $\overline{TE}_i(d)$ and by the law of iterated expectations, and the third equality holds by the definition of $\omega^{dose}(g,t,d)$. An important difference relative to $ATE^{dose}(d)$ in the main text is that, here, the weights also depend on the dose $d$. It is also straightforward to see that, given some value of $d$, $\omega^{dose}(g,t,d)$ is non-negative for all values of $(g,t)$ and that $\sum_{g \in \bar{\mathcal{G}}} \sum_{t=2}^{T} \omega^{dose}(g,t,d) = 1$.

Next, we relate $ACRT^{dose}(d|d)$ to underlying $ACRT(g,t,d|g,d)$. Notice that

$$\begin{aligned}
ACRT^{dose}(d|d) &= \frac{\partial}{\partial l} \mathbb{E}\Big[\overline{TE}(l)\Big| D=d, G \leq T\Big]\Bigg|_{l=d} \\
&= \frac{\partial}{\partial l} \mathbb{E}\left[\frac{1}{T-G+1}\sum_{t=2}^{T} \mathbf{1}\{t \geq G\}\Big(Y_{i,t}(G,l) - Y_{i,t}(0)\Big)\Big| D=d, G \leq T\right]\Bigg|_{l=d} \\
&= \frac{\partial}{\partial l}\left\{\sum_{g \in \bar{\mathcal{G}}} \frac{1}{T-g+1}\sum_{t=2}^{T} \mathbf{1}\{t \geq g\}\mathbb{E}\Big[Y_{i,t}(g,l) - Y_{i,t}(0)\Big| G=g, D=d\Big]\mathbb{P}(G=g|D=d, G \leq T)\right\}\Bigg|_{l=d} \\
&= \sum_{g \in \bar{\mathcal{G}}} \sum_{t=2}^{T} \omega^{dose}(g,t,d) ACRT(g,t,d|g,d)
\end{aligned}$$

where the first line comes from the definition of $ACRT^{dose}(d|d)$, the second line holds from the definition of $\overline{TE}(d)$, the third line holds from the law of iterated expectations, and the last line holds because the only term that depends on $l$ is the inside conditional expectation (and then by the definition of $ACRT(g,t,d|g,d)$). The previous expression shows that the aggregated parameter $ACRT^{dose}(d|d)$ can be expressed as a weighted average of underlying $ACRT(g,t,d|g,d)$ parameters.

Next, we consider "on-the-treated" event study parameters. Recall that, for units that are ever observed to participate in the treatment for $e$ periods, we defined $TE_i(d|e) = Y_{i,G_i+e}(G_i,d) - Y_{i,G_i+e}(0)$. Next, we define intermediate "on-the-treated" parameters that are a function of the dose and event-time $e$:

$$\begin{aligned}
\widetilde{ATT}^{dose,es}(d|d,e) &= \mathbb{E}\Big[TE(d|e)\Big| D=d, G+e \in [2,T], G \leq T\Big], \\
\widetilde{ACRT}^{dose,es}(d|d,e) &= \frac{\partial \widetilde{ATT}^{dose,es}(l|d,e)}{\partial l}\Bigg|_{l=d}
\end{aligned}$$

where $\widetilde{ATT}^{dose,es}(d|d,e)$ is the average treatment effect of dose $d$ among those in dose group $d$ (conditioning on dose group $d$ is the difference relative to $\widetilde{ATE}^{dose,es}(d|e)$ discussed in the main text) for those that have been exposed to the treatment for $e$ periods. Similarly, $\widetilde{ACRT}^{dose,es}(d|d,e)$ is the average causal response to a marginal increase in the dose among those in dose group $d$ that have been exposed to the treatment for $e$ periods.

As in the main text, if a researcher wants to report an event study, one option is to pick a particular value of $d$ and report $\widetilde{ATT}^{dose,es}(d|d,e)$ and/or $\widetilde{ACRT}^{dose,es}(d|d,e)$ for that value of the dose while varying event time. Another option is to average these parameters across all doses, which is the route we follow now. In particular, we can consider the parameters

$$\begin{aligned}
ATT^{es}(e) &= \mathbb{E}\Big[\widetilde{ATT}^{dose,es}(D|D,e)\Big| G+e \in [2,T], G \leq T\Big] \\
ACRT^{es}(e) &= \mathbb{E}\left[\widetilde{ACRT}^{dose,es}(D|D,e)\Big| G+e \in [2,T], G \leq T\right]
\end{aligned}$$

Although we suspect that these are the natural event study target parameters in most applications, because they are simple averages $\widetilde{ATT}^{dose,es}(d|d,e)$ and $\widetilde{ACRT}^{dose,es}(d|d,e)$ over the dose, next we show that $\widetilde{ATT}^{dose,es}(d|d,e)$ and $\widetilde{ACRT}^{dose,es}(d|d,e)$ can be related to the corresponding underlying, disaggregated parameters $ATT(g,t,d|g,d)$ and $ACRT(g,t,d|g,d)$. Toward this end, let $\pi_g(e,d) = \mathbb{P}(G=g|D=d, G+e \in [2,T], G \leq T)$, and notice that

$$\begin{aligned}
\widetilde{ATT}^{dose,es}(d|d,e) &= \mathbb{E}\Big[TE(d|e)|D=d, G+e \in [2,T], G \leq T\Big] \\
&= \sum_{g \in \bar{\mathcal{G}}} \mathbf{1}\{g+e \in [2,T]\}\mathbb{E}[Y_{g+e}(g,d) - Y_{g+e}(0)|G=g, D=d]\pi_g(e,d) \\
&= \sum_{g \in \bar{\mathcal{G}}} \left\{ \mathbf{1}\{g+e \in [2,T]\}\mathbb{E}[Y_{g+e}(g,d) - Y_{g+e}(0)|G=g, D=d]\pi_g(e,d) \sum_{t=2}^{T} \mathbf{1}\{g+e=t\} \right\} \\
&= \sum_{g \in \bar{\mathcal{G}}} \sum_{t=2}^{T} \mathbf{1}\{g+e \in [2,T]\}\mathbf{1}\{g+e=t\}\mathbb{E}[Y_t(g,d) - Y_t(0)|G=g, D=d]\pi_g(e,d) \\
&= \sum_{g \in \bar{\mathcal{G}}} \sum_{t=2}^{T} \omega^{dose,es}(g,t,d|e) ATT(g,t,d|g,d)
\end{aligned}$$

where $\omega^{dose,es}(g,t,d|e) = \mathbf{1}\{g+e \in [2,T]\}\mathbf{1}\{g+e=t\}\pi_g(e,d)$ and where the first equality holds by the definition of $\widetilde{ATT}^{dose,es}(d|d,e)$, the second equality holds by the law of iterated expectations, the third equality holds because $\sum_{t=2}^{T} \mathbf{1}\{g+e=t\} = 1$ among groups that are observed to participate in the treatment for $e$ periods, the fourth equality holds by combining the summations, and the last equality holds by the definitions of $\omega^{dose,es}$ and $ATT(g,t,d|g,d)$. Notice that the weights, $\omega^{dose,es}(g,t,d|e)$ are similar to the event study weights $w^{dose,es}(g,t|e)$ discussed in the main text except for that the probability term here depends on $d$ while it did not in the main text—this difference arises because the parameters here are "on-the-treated" while the ones in the main text were not. Next, consider

$$\begin{aligned}
\widetilde{ACRT}^{dose,es}(d|d,e) &= \frac{\partial \mathbb{E}\Big[TE(l|e)|D=d, G+e \in [2,T], G \leq T\Big]}{\partial l}\Bigg|_{l=d} \\
&= \frac{\partial}{\partial l} \left\{ \sum_{g \in \bar{\mathcal{G}}} \sum_{t=2}^{T} \mathbf{1}\{g+e \in [2,T]\}\mathbf{1}\{g+e=t\}\mathbb{E}[Y_t(g,l) - Y_t(0)|G=g, D=d]\pi_g(e,d) \right\}\Bigg|_{l=d} \\
&= \sum_{g \in \bar{\mathcal{G}}} \sum_{t=2}^{T} \omega^{dose,es}(g,t,d|e) ACRT(g,t,d|g,d)
\end{aligned}$$

where the first equality holds by the definition of $\widetilde{ACRT}^{dose,es}(d|d,e)$, the second equality holds using the same sort of argument as for $\widetilde{ATT}^{dose,es}(d|d,e)$ above, and the last equality holds because the only term that depends on $l$ is the inside conditional expectation which is equal to $ATT(g,t,l|g,d)$ and then the result holds by the definitions of $\omega^{dose,es}(g,t,d|e)$ and $ACRT(g,t,d|g,d)$.

This discussion highlights that if $ATT(g,t,d|g,d)$ and/or $ACRT(g,t,d|g,d)$ are identified, then we can recover the more aggregated summary parameters that have been discussed in this section.

## SA.2 Identification

In this section, we consider the identification of the disaggregated parameters $ATT(g,t,d|g,d)$, $ATE(g,t,d)$, $ACRT(g,t,d|g,d)$, and $ACR(g,t,d)$. Given the discussion above and in Appendix D in the main text, if these parameters are identified, then it implies that we can recover more aggregated parameters (e.g., $ATT^{es}(e)$ or $ACR(d)$, among others) that were discussed in those sections.

Following the discussion in Callaway and Sant'Anna (2021) and Marcus and Sant'Anna (2021), we consider several alternative versions of the parallel trends and strong parallel trends assumptions that we considered in the main text. We note that, in a setting with two periods, all of the versions of parallel trends and strong parallel trends that we consider here are equivalent to each other though they differ in cases with multiple periods and variation in treatment timing.[1]

**Assumption 4-MP-Extended** (Parallel Trends with Multiple Periods and Variation in Treatment Timing)**.**

*(a)* *For all* $g \in \mathcal{G}$, $t = 2, \ldots, T$, $d \in \mathcal{D}$, $\mathbb{E}[\Delta Y_t(0)|G = g, D = d] = \mathbb{E}[\Delta Y_t(0)|D = 0]$.

*(b)* *For all* $g \in \mathcal{G}$, $t = g, \ldots, T$, $d \in \mathcal{D}$, $\mathbb{E}[\Delta Y_t(0)|G = g, D = d] = \mathbb{E}[\Delta Y_t(0)|D = 0]$.

*(c)* *For all* $g \in \mathcal{G}$, $t = g, \ldots, T$, $d \in \mathcal{D}$, $\mathbb{E}[\Delta Y_t(0)|G = g, D = d] = \mathbb{E}[\Delta Y_t(0)|G = k]$ *for all groups* $k \in \mathcal{G}$ *such that* $t < k$ *(i.e., pre-treatment periods for group* $k$*).*

Assumption 4-MP-Extended (a), (b), and (c) all provide ways to extend the idea of parallel trends from a two-period setting to a setting with multiple periods and variation in treatment timing. They differ on the basis of (i) the comparison group that they rationalize and (ii) whether parallel trends is assumed to hold in pre-treatment periods. Assumption 4-MP-Extended(a) is the strongest assumption about paths of untreated potential outcomes. It says that paths of untreated potential outcomes are the same for all groups and for all doses across all time periods. Assumption 4-MP-Extended(b) says that the path of outcomes for group $g$ in post-treatment time periods is the same as the path of untreated potential outcomes among never-treated units. Parallel pre-trends need not hold under part (b). Assumption 4-MP-Extended(c) says that the path of outcomes for group $g$ in post-treatment time periods is the same as the path of outcomes among all groups that are not treated yet in that period—this includes both the untreated group as well as groups that will eventually be treated but that are not treated yet. Based on the results in earlier sections, note that each parallel trends assumption in Assumption 4-MP-Extended is directed toward identifying $ATT(g,t,d|g,d)$ rather than $ATE(g,t,d)$.

Next, we provide an analogous set of assumptions that target identifying $ATE(g,t,d)$.

**Assumption 5-MP-Extended** (Strong Parallel Trends with Multiple Periods and Variation in Treatment Timing)**.**

[1] It is also straightforward to develop identification results under multi-period versions of aggregate parallel trends (Assumption 4-Agg) or alternative strong parallel trends (Assumption 5-Alt) that we discussed in the main text. These results hold using the same sort of extension ideas discussed here for parallel trends and strong parallel trends, so we do not provide formal results for these cases for the sake of brevity.

*(a) For all* $g \in \mathcal{G}$, $t = 2, \ldots, T$, *and* $d \in \mathcal{D}$, $\mathbb{E}[Y_t(g,d) - Y_{t-1}(g,d)|G = g, D = d] = \mathbb{E}[Y_t(g,d) - Y_{t-1}(g,d)|G = g]$ *and* $\mathbb{E}[\Delta Y_t(0)|G = g, D = d] = \mathbb{E}[\Delta Y_t(0)|D = 0]$

*(b) For all* $g \in \mathcal{G}$, $t = g, \ldots, T$, $d \in \mathcal{D}$, $\mathbb{E}[Y_t(g,d) - Y_{t-1}(g,d)|G = g, D = d] = \mathbb{E}[Y_t(g,d) - Y_{t-1}(g,d)|G = g]$ *and* $\mathbb{E}[\Delta Y_t(0)|G = g, D = d] = \mathbb{E}[\Delta Y_t(0)|D = 0]$

*(c) For all* $g \in \mathcal{G}$, $t = g, \ldots, T$, $d \in \mathcal{D}$, $\mathbb{E}[Y_t(g,d) - Y_{t-1}(g,d)|G = g, D = d] = \mathbb{E}[Y_t(g,d) - Y_{t-1}(g,d)|G = g]$ *and* $\mathbb{E}[\Delta Y_t(0)|G = g, D = d] = \mathbb{E}[\Delta Y_t(0)|G = k]$ *for all groups* $k \in \mathcal{G}$ *such that* $t < k$ *(i.e., pre-treatment periods for group* $k$*).*

Parts (a), (b), and (c) of the assumption correspond to the same parts in Assumption 4-MP-Extended and differ based on which group is used as the comparison group in terms of untreated potential outcomes. Part (a) additionally corresponds to Assumption 5-MP in the main text. Finally, the reason that there are two parts to these assumptions rather than just one as in Assumption 4-MP-Extended is that, in the setup of this section, conditional on being in group $g$ with $t \geq g$, by construction, there are no untreated units in the group; thus, the second part of the assumption handles untreated potential outcomes slightly differently than treated potential outcomes—essentially these multi-period versions of strong parallel trends allow us to compare paths of outcomes across doses within a particular time period and for a particular timing group while some untreated comparison group can be used to construct the trend in untreated potential outcomes. Before providing our main identification result with multiple periods and variation in treatment timing and dose, recall that (as defined in the main text) $W_{i,t} = D_i \mathbf{1}\{t \geq G_i\}$ which is equal to 0 for units that are untreated in period $t$ and equal to $D_i$ for units that have been treated by period $t$.

**Theorem S1.** *Under Assumptions 1-MP, 2-MP(a), and 3-MP, and for all* $g \in \mathcal{G}$, $t = 2, \ldots, T$ *such that* $t \geq g$, *and for all* $d \in \mathcal{D}$,

*(1a) If, in addition, either Assumption 4-MP-Extended(a) or (c) holds, then*

$$ATT(g,t,d|g,d) = \mathbb{E}[Y_t - Y_{g-1}|G = g, D = d] - \mathbb{E}[Y_t - Y_{g-1}|W_t = 0]$$

*(1b) If, in addition, Assumption 4-MP-Extended(b) holds, then*

$$ATT(g,t,d|g,d) = \mathbb{E}[Y_t - Y_{g-1}|G = g, D = d] - \mathbb{E}[Y_t - Y_{g-1}|D = 0]$$

*(2a) If, in addition, either Assumption 5-MP-Extended(a) or (c) holds, then*

$$ATE(g,t,d) = \mathbb{E}[Y_t - Y_{g-1}|G = g, D = d] - \mathbb{E}[Y_t - Y_{g-1}|W_t = 0]$$

*(2b) If, in addition, Assumption 5-MP-Extended(b) holds, then*

$$ATE(g,t,d) = \mathbb{E}[Y_t - Y_{g-1}|G = g, D = d] - \mathbb{E}[Y_t - Y_{g-1}|D = 0]$$

The proof of Theorem S1 is provided in Appendix SB. Part (1a) of Theorem S1 says that $ATT(g,t,d|g,d)$—the average effect of participating in the treatment in time period $t$ among units who became treated in period $g$ and experienced dose $d$—is identified under a parallel trends assumption and that it is equal to the average path of outcomes experienced by units in group $g$ under dose

$d$ adjusted by the average path of outcomes experienced among units that are not-yet-treated by period $t$. The results in the other parts are similar as well. For part (1b), the weaker parallel trends assumption in Assumption 4-MP-Extended(b) implies that the never-treated group should be used as the comparison group (this is a smaller comparison group relative to the not-yet-treated group). Parts (2a) and (2b) show that under Assumption 5-MP-Extended the same estimands identify $ATE(g,t,d)$.

Finally, for this section, we show that the same sort of selection bias terms as we emphasized in the main text can show up when making comparisons across doses (and, hence, show up in causal response parameters) in a setting with multiple periods and variation in treatment timing and dose under parallel trends assumptions. And, also like in the main text, strong parallel trends can be used to eliminate these selection bias terms. For simplicity, we provide these results under the strongest versions of Assumptions 4-MP-Extended and 5-MP-Extended, but analogous results hold in the other cases as well.

**Theorem S2.** *Under Assumptions 1-MP, 2-MP, and 3-MP, and for all $g \in \mathcal{G}$, $t = 2, \dots, T$ such that $t \geq g$, and for all $d \in \mathcal{D}_+^c$,*

*(1) If, in addition, Assumption 4-MP-Extended(a) holds, then*

$$\begin{aligned}\frac{\partial}{\partial d}\mathbb{E}[Y_t - Y_{g-1}|G=g, D=d] &= \frac{\partial}{\partial d}ATT(g,t,d|g,d) \\ &= ACRT(g,t,d|g,d) + \underbrace{\left.\frac{\partial ATT(g,t,d|g,l)}{\partial l}\right|_{l=d}}_{\text{selection bias}}.\end{aligned}$$

*(2) If, in addition, Assumption 5-MP-Extended(a) holds, then*

$$\frac{\partial}{\partial d}\mathbb{E}[Y_t - Y_{g-1}|G=g, D=d] = \frac{\partial}{\partial d}ATE(g,t,d) = ACR(g,t,d).$$

The proof of Theorem S2 is provided in Appendix SB. Theorem S2 provides an analogous result for the case with multiple periods and variation in treatment timing and dose to Theorems 3.2 and 3.3 in the main text. The theorem has implications for the aggregated parameters discussed above. If one maintains some version of strong parallel trends, then it rationalizes targeting causal response parameters such as $ACR^{dose}(d)$ or $ACR^{es}(e)$. However, parallel trends alone does not recover aggregated causal response such as $ACRT^{dose}(d|d)$ or $ACRT^{es}(e)$ due to comparisons of paths of outcomes across doses including, under parallel trends, both causal responses and selection bias. On the other hand, parallel trends alone does recover summary level-effect parameters such as $ATT^{dose}(d|d)$, $ATT^o$, or $ATT^{es}(e)$.[2]

**Remark S1.** *The parallel trends assumptions in Assumption 4-MP-Extended are not the only possible ones. Interestingly, with a continuous treatment, there are some possible (and reasonable) comparison groups that are available that are not available with a binary treatment. For example, one could assume that*

[2]That said, it is worth emphasizing again that parallel trends alone does not rationalize causally interpreting differences in $ATT^{dose}(d|d)$ across $d$—this is analogous to the discussion in the main text and is also closely related to the discussion about $ACRT^{dose}(d|d)$ in this paragraph.

*For all $g \in \mathcal{G}$, $t = g, \ldots, T$, $d \in \mathcal{D}$, $\mathbb{E}[\Delta Y_t(0)|G = g, D = d] = \mathbb{E}[\Delta Y_t(0)|G = k, D = d]$ for all groups $k \in \mathcal{G}$ such that $t < k$ (i.e., pre-treatment periods for group $k$).*

*This sort of assumption amounts to using as a comparison group the set of units that are not yet treated but will eventually experience the same dose. It is straightforward to adapt the approach described in Theorem S1 to this sort of case and propose related estimators that can deliver consistent estimates of $ATT(g, t, d|g, d)$ under this assumption.*

**Remark S2.** *If a researcher is interested in targeting a particular $ATT(g, t, d|g, d)$ or $ATE(g, t, d)$, it is generally possible to weaken Assumption 4-MP-Extended or 5-MP-Extended. For example, one could make parallel trends directly about long differences, $(Y_t - Y_{g-1})$, rather than all short differences (this sort of assumption is generally weaker), or, in part (c) of each assumption, use more aggregated comparison groups instead of imposing parallel trends for all possible comparison groups (which is also weaker), or alternatively only make parallel trends assumptions for the particular dose being considered.*

**Remark S3.** *We do not provide formal estimation results for the setting with multiple periods and variation in treatment timing though we note that, if one bases estimation on the sample analog of the results in Theorem S1, then the results in the main text for the case with two periods apply directly to the disaggregated parameters $ATT(g, t, d|g, d)$, $ATE(g, t, d)$, and $ACR(g, t, d)$. For the aggregated parameters discussed above, at a high-level, one can then proceed to combine estimation results for the disaggregated parameters with the estimation results for the related aggregation schemes proposed in Callaway and Sant'Anna (2021).*

**Remark S4.** *Notice that the expressions for $ATT(g, t, d|g, d)$ are the same under Assumption 4-MP-Extended(a) and (c) while Assumption 4-MP-Extended(a) is stronger than Assumption 4-MP-Extended(b). In estimation, it may be possible to propose more efficient estimators under Assumption 4-MP-Extended(a) that exploit parallel trends holding across all periods and groups. This is akin to similar issues that arise in a setting with binary treatment (see Callaway (2023) for a discussion in the context of a binary treatment). The same sort of comment applies to estimating $ATE(g, t, d)$ under the different versions of Assumption 5-MP-Extended.*

## SA.3 TWFE estimators with multiple time periods and variation in treatment timing

In applications with multiple periods and variation in treatment timing and dose, empirical researchers typically estimate the TWFE regression

$$Y_{i,t} = \theta_t + \eta_i + \beta^{twfe} W_{i,t} + v_{i,t}. \tag{S1}$$

Equation (S1) is exactly the same as the TWFE regression in the baseline case with two periods in Equation (1.1) in the main text only with the notation slightly adjusted to match this section. In the main text, we related $\beta^{twfe}$ to several different types of causal effect parameters (see Theorem 3.4 in the main text). In this section, we provide related results for the setting with multiple time periods and variation in treatment timing with a particular emphasis on the comparisons underlying $\beta^{twfe}$ and in causal interpretations (especially causal response interpretations) of $\beta^{twfe}$ in the presence of treatment effect heterogeneity. The results in this section generalize the results in several recent papers on TWFE estimates, including Goodman-Bacon (2021) and de Chaisemartin and D'Haultfœuille (2020) to our DiD setup with variation in treatment intensity. In this section, we modify our previous notation slightly by setting $G_i = T+1$ for units that do not participate in the treatment in any period (rather than $G_i = \infty$), which simplifies the exposition in several places in this section.

To start with, write population versions of TWFE adjusted variables by

$$\ddot{W}_{i,t} = (W_{i,t} - \bar{W}_i) - \left(\mathbb{E}[W_t] - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[W_t]\right), \quad \text{where} \quad \bar{W}_i = \frac{1}{T}\sum_{t=1}^{T} W_{i,t}.$$

The population version of the TWFE estimator is

$$\beta^{twfe} = \frac{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[Y_{i,t}\ddot{W}_{i,t}]}{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\ddot{W}_{i,t}^2]}. \tag{S2}$$

As in the main text, we present both a "mechanical" decomposition of the TWFE estimator and a "causal" decomposition of the estimand that relates assumptions to interpretation. In order to define these decompositions, we introduce a bit of new notation. First, define the fraction of periods that units in group $g$ spend treated as

$$\bar{G}_g = \frac{T-(g-1)}{T}.$$

For the untreated group $g = T+1$ so that $\bar{G}_{T+1} = 0$.

Next, we define time periods over which averages are taken. For averaging variables across time periods, we use the following notation, for $t_1 \leq t_2$,

$$\bar{Y}_i^{(t_1,t_2)} = \frac{1}{t_2 - t_1 + 1}\sum_{t=t_1}^{t_2} Y_{i,t}.$$

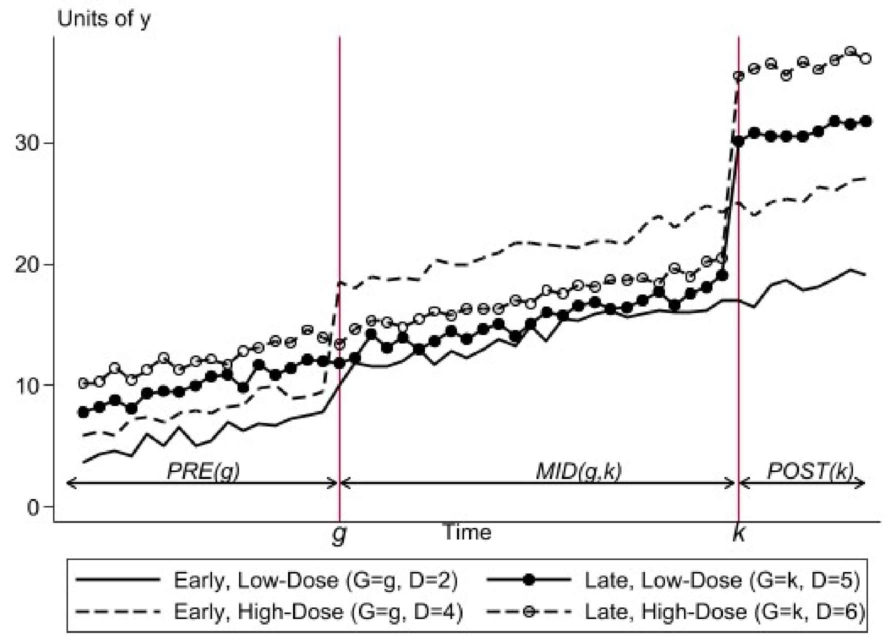
It is also convenient to define some particular averages across time periods. For two time periods $g$ and $k$, with $k > g$, (below, $g$ and $k$ will often index groups defined by treatment timing), we define

$$\bar{Y}_i^{PRE(g)} = \bar{Y}_i^{(1,g-1)}, \quad \bar{Y}_i^{MID(g,k)} = \bar{Y}_i^{(g,k-1)}, \quad \bar{Y}_i^{POST(k)} = \bar{Y}_i^{(k,T)}.$$

$\bar{Y}_i^{PRE(g)}$ is the average outcome for unit $i$ in periods 1 to $g-1$, $\bar{Y}_i^{MID(g,k)}$ is the average outcome for unit $i$ in periods $g$ to $k-1$, and $\bar{Y}_i^{POST(k)}$ is the average outcome for unit $i$ in periods $k$ to $T$. Below, when $g$ and $k$ index groups, $\bar{Y}_i^{PRE(g)}$ is the average outcome for unit $i$ in periods before units in either group are treated, $\bar{Y}_i^{MID(g,k)}$ is the average outcome for unit $i$ in periods after group $g$ has become treated but before group $k$ has been treated, and $\bar{Y}_i^{POST(k)}$ is the average outcome for unit $i$ after both groups have become treated.

To fix ideas about how the staggered-timing/continuous treatment case works, consider a setup with two timing groups, $g$ and $k$, with $k > g$. Some units in the "early-treated" group have $d = 2$, and others have $d = 4$. Some units in the late-treated group have $d = 5$, and others have $d = 6$. Thus, the four groups are early-treated/high-dose, early-treated/low-dose, late-treated/high-dose, and late-treated/low-dose. Figure S1 plots constructed outcomes for these groups with a treatment effect that is a one-time shift equal to $d^{1.5}$.

Figure S1: A Simple Set-Up with Staggered Timing and Variation in the Dose




*Notes:* The figure plots simulated data for four groups: early-treated/high-dose, early-treated/low-dose, late-treated/high-dose, and late-treated/low-dose.

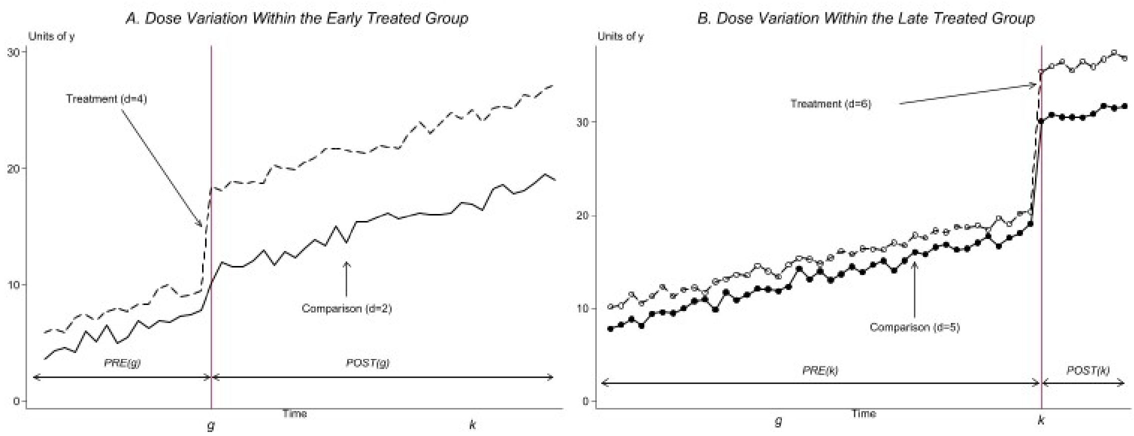
Following Goodman-Bacon (2021), we motivate the decomposition of the TWFE estimand by considering the four types of simple DiD estimands that can be formed using only one source of variation. The first comparison is a within timing-group comparison of paths of outcomes among units that experienced different amounts of the treatment.

$$\delta^{WITHIN}(g) = \frac{\text{Cov}(\bar{Y}^{POST(g)} - \bar{Y}^{PRE(g)}, D|G = g)}{\text{Var}(D|G = g)}. \tag{S3}$$

This term is essentially the same as the expression for the TWFE estimand in the baseline two-period case. It equals the OLS (population) coefficient from regressing the change in average outcomes before and after $g$ for units treated at time $g$ on their dose, $d$. Figure S2 uses the four-group example to show how $\delta^{WITHIN}(g)$ and $\delta^{WITHIN}(k)$ use higher-dose units as the "treatment group" and lower-dose units as the "comparison group".

The second comparison is based on treatment timing. It compares paths of outcomes between a particular timing group $g$ and a "later-treated" group $k$ (i.e., $k > g$) in the periods after group $g$ is

Figure S2: Within-Timing-Group Comparisons Across Doses




*Notes:* The figure shows the within-timing group comparison between higher- and lower-dose units defined by $\delta^{WITHIN}(g)$ and $\delta^{WITHIN}(k)$.

treated but before group $k$ becomes treated relative to their common pre-treatment periods.[3]

$$\delta^{MID,PRE}(g,k) = \frac{\mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G=g\right] - \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G=k\right]}{\mathbb{E}[D|G=g]}. \quad \text{(S4)}$$

Panel A of Figure S3 plots the outcomes used in this comparison with timing-group averages in black and the specific dose groups from Figure S1 in light gray. Under a parallel trends assumption, we show below that this term corresponds to a reasonable treatment effect parameter because the path of outcomes for group $k$ (which is still in its pre-treatment period here) is what the path of outcomes would have been for group $g$ if it had not been treated. Also note that this term encompasses comparisons of group $g$ to the "never-treated" group.

The third comparison is between paths of outcomes for the "later-treated" group $k$ in its post-treatment period relative to a pre-treatment period adjusted by the same path of outcomes for the "early-treated" group $g$.

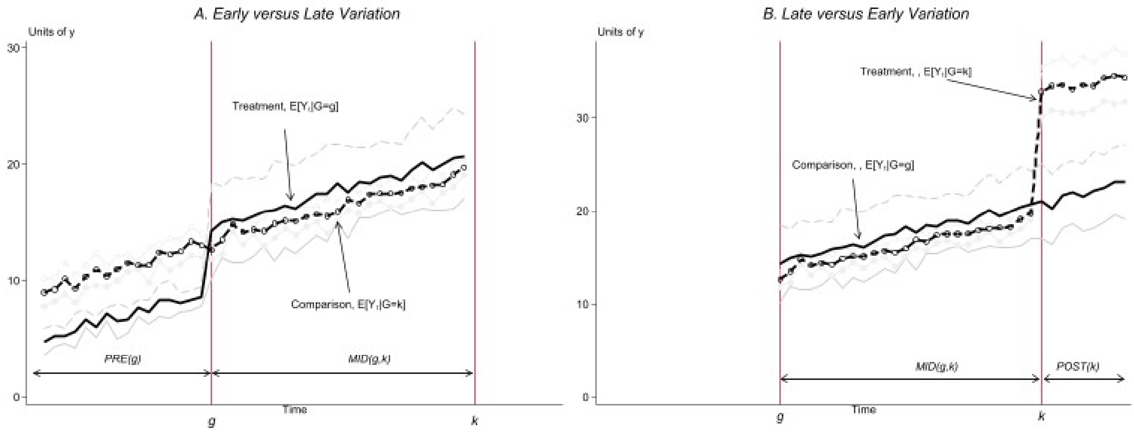
$$\delta^{POST,MID}(g,k) = \frac{\mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}\right)|G=k\right] - \mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}\right)|G=g\right]}{\mathbb{E}[D|G=k]}. \quad \text{(S5)}$$

These terms use the already-treated group $g$ as the comparison group for group $k$. Panel B of Figure S3 plots the outcomes used in this term. Mechanically, the TWFE regression exploits this comparison because group $g$'s treatment status/amount is not changing over these time periods. However, these are post-treatment periods for group $g$, and parallel trends assumptions do not place restrictions on paths of post-treatment outcomes, which are subtracted in Equation (S5). Therefore, it is undesirable that this term shows up in the expression for $\beta^{twfe}$.[4]

[3] Each of the following expressions also includes a term in the denominator. Below, this term is useful for interpreting differences across groups as partial effects of more treatment, but, for now, we largely ignore the expressions in the denominator.

[4] This sort of comparison also shows up in the case with a binary, staggered treatment. See, e.g., de Chaisemartin and D'Haultfœuille (2020), Goodman-Bacon (2021), and Borusyak, Jaravel, and Spiess (2023).

Figure S3: Between-Timing-Group Comparisons




*Notes:* The figure shows the between-timing-group comparisons that average the outcomes in groups $g$ and $k$ across dose levels and compare the early group to the later group (panel C) or the later group to the early group (panel D).

The final comparison that shows up in the TWFE estimator is between paths of outcomes between "early" and "late" treated groups in their common post-treatment periods relative to their common pre-treatment periods. In other words, this comparison comes from the "endpoints" where the two timing groups are either both untreated or both treated with possibly different average doses.

$$\delta^{POST,PRE}(g,k) = \frac{\mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)}\right)|G=g\right] - \mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)}\right)|G=k\right]}{\mathbb{E}[D|G=g] - \mathbb{E}[D|G=k]}. \tag{S6}$$

Figure S4 shows the outcomes that determine the comparisons that show up in this term. The reason that this term shows up in $\beta^{twfe}$ is that differences in the paths of outcomes between groups that have different distributions of the treatment are informative about $\beta^{twfe}$. For example, if more dose tends to increase outcomes and group $g$'s dose is higher on average than group $k$'s, then outcomes may increase more among group $g$ than group $k$ resulting in $\delta^{POST,PRE}(g,k)$ not being equal to 0.[5]

Next, we show how $\beta^{twfe}$ weights these simple DiD terms together and discuss its theoretical interpretation under parallel trends assumptions. To characterize the weights, first, define $p_g = \mathbb{P}(G = g)$ and

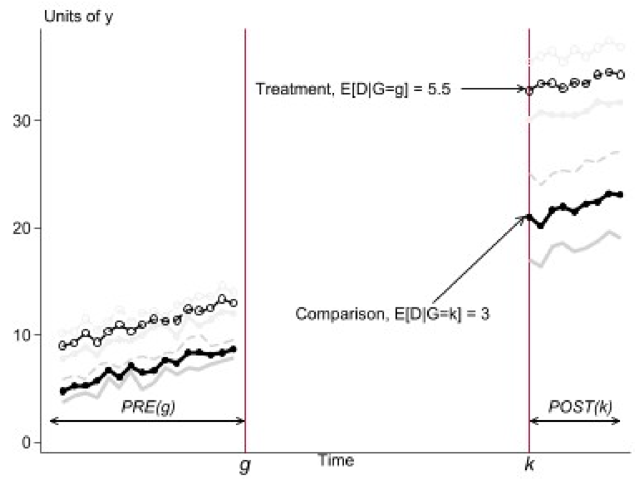
$$p_{g|\{g,k\}} = \mathbb{P}(G = g|G \in \{g,k\}),$$

which is the probability of being in group $g$ conditional on being in either group $g$ or $k$. We also define the following weights, which measure the variance of the treatment variable used to estimate each of the simple DiD terms in equations Equations (S3) to (S6).

$$w^{g,within}(g) = \text{Var}(D|G=g)(1-\bar{G}_g)\bar{G}_g p_g \bigg/ \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\ddot{W}_{i,t}^2],$$

[5]To be more precise, this term involves comparisons between groups $g$ and $k$ for the group with a higher dose on average to the group with a smaller dose on average. When $\mathbb{E}[D|G=g] > \mathbb{E}[D|G=k]$, this corresponds to the expression in Equation (S6). When $\mathbb{E}[D|G=g] < \mathbb{E}[D|G=k]$, one can multiply both the numerator and denominator by $-1$ so that we effectively make a positive-weight comparison for the group that experienced more dose relative to the group that experienced less dose.

Figure S4: Long Comparisons Between Timing Groups




*Notes:* The figure shows the comparisons between timing groups in the $POST(k)$ window when both are treated with potentially different average doses and the $PRE(g)$ window when neither group is treated.

$$w^{g,post}(g,k) = \mathbb{E}[D|G=g]^2(1-\bar{G}_g)(\bar{G}_g - \bar{G}_k)(p_g+p_k)^2 p_{g|\{g,k\}}(1-p_{g|\{g,k\}}) \Big/ \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\ddot{W}_{i,t}^2],$$

$$w^{k,post}(g,k) = \mathbb{E}[D|G=k]^2\bar{G}_k(\bar{G}_g - \bar{G}_k)(p_g+p_k)^2 p_{g|\{g,k\}}(1-p_{g|\{g,k\}}) \Big/ \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\ddot{W}_{i,t}^2],$$

$$w^{long}(g,k) = (\mathbb{E}[D|G=g]-\mathbb{E}[D|G=k])^2\bar{G}_k(1-\bar{G}_g)(p_g+p_k)^2 p_{g|\{g,k\}}(1-p_{g|\{g,k\}}) \Big/ \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\ddot{W}_{i,t}^2].$$

These weights are similar to the ones in Goodman-Bacon (2021) in the sense that they combine the size of the sample and the variance of treatment used to calculate each simple DiD term. In $w^{g,within}(g)$, for example, $\text{Var}(D|G=g)$ measures how much the dose varies across units with $G=g$, $(1-\bar{G}_g)\bar{G}_g$ measures the variance that comes from timing which falls when $g$ is closer to 0 or $T$, and $p_g$ measures the share of units with $G=g$ (i.e., subsample size). Since they only compare outcomes between timing-groups, $w^{g,post}(g,k)$ and $w^{k,post}(g,k)$ do not contain a within-timing-group variance of $D$, but they do include $\mathbb{E}[D|G=k]^2$ which reflects the fact that timing groups with higher average doses get more weight. The rest of the timing weights have the same interpretation as in Goodman-Bacon (2021). Finally, $w^{long}(g,k)$ includes the square of the difference in mean doses between groups $g$ and $k$—$(\mathbb{E}[D|G=g]-\mathbb{E}[D|G=k])^2$—which shows that the "endpoint" comparisons only influence $\beta^{twfe}$ to the extent that timing groups have different average doses. Two timing groups with the same average dose do not contribute a $\delta^{POST,PRE}(g,k)$ term because there is no differential change in their doses between the $PRE(g)$ window (when both groups are untreated) and the $POST(k)$ window (when both groups have $\mathbb{E}[D|G=g]=E[D|G=k]$).

Our next result combines the simple DiD terms and their variance weights to provide a mechanical decomposition of $\beta^{twfe}$ in DiD setups with variation in treatment timing and variation in treatment intensity.

**Proposition S1.** *Under Assumptions 1-MP, 2-MP(a), and 3-MP, $\beta^{twfe}$ in Equation (S1) can be written as*

$$\beta^{twfe} = \sum_{g \in \mathcal{G}} w^{g,within}(g)\delta^{WITHIN}(g)$$
$$+ \sum_{g \in \mathcal{G}} \sum_{k \in \mathcal{G}, k>g} \Big\{ w^{g,post}(g,k)\delta^{MID,PRE}(g,k) + w^{k,post}(g,k)\delta^{POST,MID}(g,k) + w^{long}(g,k)\delta^{POST,PRE}(g,k) \Big\}.$$

*In addition, (i) $w^{g,within}(g) \geq 0$, $w^{g,post}(g,k) \geq 0$, $w^{k,post}(g,k)$, and $w^{long}(g,k) \geq 0$ for all $g \in \mathcal{G}$ and $k \in \mathcal{G}$ with $k > g$, and (ii) $\sum_{g \in \mathcal{G}} w^{g,within}(g) + \sum_{g \in \mathcal{G}} \sum_{k \in \mathcal{G}, k>g} \big\{ w^{g,post(g,k)}(g,k) + w^{k,post}(g,k) + w^{long}(g,k) \big\} = 1$.*

Proposition S1 generalizes the decomposition theorem for binary staggered timing designs in Goodman-Bacon (2021) to our setup with variation in treatment intensity.[6] Notice that it does not require Assumption 2-MP(b) and is, therefore, compatible with a binary, multi-valued, continuous, or mixed treatment. It says that $\beta^{twfe}$ can be written as a weighted average of the four comparisons in Equations (S3) to (S6). These weights are all positive and sum to one.

Proposition S1 provides a new, explicit description of what kinds of comparisons TWFE uses to compute $\beta^{twfe}$, but it does not on its own provide guidance on how to interpret TWFE estimates. Our results for the two-period case in the main text, for example, show that simple estimators like $\delta^{WITHIN}(g)$ equal averages of $ACRT$ parameters plus selection bias. Similarly, the terms that compare outcomes across timing groups necessarily average over the dose-specific treatment effects of units within that timing group. We analyze the theoretical interpretation of each of these simple DiD estimands under different assumptions and then discuss what this implies about the (arguably implicit) identifying assumptions and estimand for TWFE.

To begin we define additional weights that apply to the underlying causal parameters in the DiD terms in Equations (S3) through (S6):

$$w_1^{within}(g,l) = \frac{\Big(\mathbb{E}[D|G=g, D \geq l] - \mathbb{E}[D|G=g]\Big)}{\text{Var}(D|G=g)} \mathbb{P}(D \geq l|G=g),$$

$$w_1(g,l) = \frac{\mathbb{P}(D \geq l|G=g)}{\mathbb{E}[D|G=g]}, \qquad w_0(g) = \frac{d_L}{\mathbb{E}[D|G=g]},$$

$$w_1^{across}(g,k,l) = \frac{(\mathbb{P}(D \geq l|G=g) - \mathbb{P}(D \geq l|G=k))}{(\mathbb{E}[D|G=g] - \mathbb{E}[D|G=k])},$$
$$\tilde{w}_1^{across}(g,k,l) = \frac{\mathbb{P}(D \geq l|G=k)}{(\mathbb{E}[D|G=g] - \mathbb{E}[D|G=k])}, \qquad \tilde{w}_0^{across}(g,k) = \frac{d_L}{(\mathbb{E}[D|G=g] - \mathbb{E}[D|G=k])}.$$

In addition, define the following differences in paths of outcomes over time

$$\pi^{POST(\tilde{k}),PRE(\tilde{g})}(g) = \mathbb{E}\left[\left(\bar{Y}^{POST(\tilde{k})} - \bar{Y}^{PRE(\tilde{g})}\right)|G=g\right] - \mathbb{E}\left[\left(\bar{Y}^{POST(\tilde{k})} - \bar{Y}^{PRE(\tilde{g})}\right)|D=0\right],$$
$$\pi^{MID(\tilde{g},\tilde{k}),PRE(\tilde{g})}(g) = \mathbb{E}\left[\left(\bar{Y}^{MID(\tilde{g},\tilde{k})} - \bar{Y}^{PRE(\tilde{g})}\right)|G=g\right] - \mathbb{E}\left[\left(\bar{Y}^{MID(\tilde{g},\tilde{k})} - \bar{Y}^{PRE(\tilde{g})}\right)|D=0\right],$$

[6] In particular, in the special case of a staggered, binary treatment, $w^{g,within}(g)\delta^{WITHIN}(g) = 0$ (since there is no within-group variation in the dose in this case), and $w^{long}(g,k)\delta^{POST,PRE}(g,k) = 0$ (because the distribution of the dose is the same across all groups). Then, Proposition S1 collapses to Theorem 1 in Goodman-Bacon (2021) because the terms $w^{g,post(g,k)}\delta^{MID,PRE}(g,k)$ and $w^{k,post}(g,k)\delta^{POST,MID}(g,k)$ correspond exactly to between-timing-group comparisons.

$$\pi^{POST(\tilde{k}),MID(\tilde{g},\tilde{k})}(g) = \mathbb{E}\left[\left(\bar{Y}^{POST(\tilde{k})} - \bar{Y}^{MID(\tilde{g},\tilde{k})}\right)|G=g\right] - \mathbb{E}\left[\left(\bar{Y}^{POST(\tilde{k})} - \bar{Y}^{MID(\tilde{g},\tilde{k})}\right)|D=0\right],$$

and, similarly,

$$\pi_D^{POST(\tilde{k}),PRE(\tilde{g})}(g,d) = \mathbb{E}\left[\left(\bar{Y}^{POST(\tilde{k})} - \bar{Y}^{PRE(\tilde{g})}\right)|G=g, D=d\right] - \mathbb{E}\left[\left(\bar{Y}^{POST(\tilde{k})} - \bar{Y}^{PRE(\tilde{g})}\right)|D=0\right],$$

$$\pi_D^{MID(\tilde{g},\tilde{k}),PRE(\tilde{g})}(g,d) = \mathbb{E}\left[\left(\bar{Y}^{MID(\tilde{g},\tilde{k})} - \bar{Y}^{PRE(\tilde{g})}\right)|G=g, D=d\right] - \mathbb{E}\left[\left(\bar{Y}^{MID(\tilde{g},\tilde{k})} - \bar{Y}^{PRE(\tilde{g})}\right)|D=0\right],$$

$$\pi_D^{POST(\tilde{k}),MID(\tilde{g},\tilde{k})}(g,d) = \mathbb{E}\left[\left(\bar{Y}^{POST(\tilde{k})} - \bar{Y}^{MID(\tilde{g},\tilde{k})}\right)|G=g, D=d\right] - \mathbb{E}\left[\left(\bar{Y}^{POST(\tilde{k})} - \bar{Y}^{MID(\tilde{g},\tilde{k})}\right)|D=0\right],$$

which are the same paths of outcomes but conditional on having dose $d$.

The following result is our main result on interpreting TWFE estimates with continuous treatment.

**Theorem S3.** *Under Assumptions 1-MP, 2-MP, and 3-MP,*

*(1) The four comparisons in Equations (S3) to (S6) can be written as*

$$\delta^{WITHIN}(g) = \int_{d_L}^{d_U} w_1^{within}(g,l)\frac{\partial \pi_D^{POST(g),PRE(g)}(g,l)}{\partial l}\,dl,$$

$$\delta^{MID,PRE}(g,k) = \int_{d_L}^{d_U} w_1(g,l)\frac{\partial \pi_D^{MID(g,k),PRE(g)}(g,l)}{\partial l}\,dl + w_0(g)\frac{\pi_D^{MID(g,k),PRE(g)}(g,d_L)}{d_L} - w_0(g)\frac{\pi^{MID(g,k),PRE(g)}(k)}{d_L},$$

$$\delta^{POST,MID}(g,k) = \int_{d_L}^{d_U} w_1(k,l)\frac{\partial \pi_D^{POST(k),MID(g,k)}(k,l)}{\partial l}\,dl + w_0(k)\frac{\pi^{POST(k),MID(g,k)}(k,d_L)}{d_L} - w_0(k)\left(\frac{\pi^{POST(k),PRE(g)}(g) - \pi^{MID(g,k),PRE(g)}(g)}{d_L}\right),$$

$$\begin{aligned}\delta^{POST,PRE}(g,k) = &\int_{d_L}^{d_U} w_1^{across}(g,k,l)\frac{\partial \pi_D^{POST(k),PRE(g)}(g,l)}{\partial l}\,dl \\ &- \left\{\int_{d_L}^{d_U} \tilde{w}_1^{across}(g,k,l)\left(\frac{\partial \pi_D^{POST(k),PRE(g)}(k,l)}{\partial l} - \frac{\partial \pi_D^{POST(k),PRE(g)}(g,l)}{\partial l}\right)dl \right.\\ &\left.+ \tilde{w}_0^{across}(g,k)\left(\frac{\pi_D^{POST(k),PRE(g)}(k,d_L) - \pi_D^{POST(k),PRE(g)}(g,d_L)}{d_L}\right)\right\}.\end{aligned}$$

*(2) If, in addition, Assumption 5-MP-Extended(a) holds, then*

$$\delta^{WITHIN}(g) = \int_{d_L}^{d_U} w_1^{within}(g,l)\overline{ACR}^{POST(g)}(g,l)dl,$$

$$\delta^{MID,PRE}(g,k) = \int_{d_L}^{d_U} w_1(g,l)\overline{ACR}^{MID(g,k)}(g,l)\,dl + w_0(g)\frac{\overline{ATE}^{MID(g,k)}(g,d_L)}{d_L},$$

$$\delta^{POST,MID}(g,k) = \int_{d_L}^{d_U} w_1(k,l)\overline{ACR}^{POST(k)}(k,l)\,dl + w_0(k)\frac{\overline{ATE}^{POST(k)}(k,d_L)}{d_L} - w_0(k)\left(\frac{\pi^{POST(k),PRE(g)}(g) - \pi^{MID(g,k),PRE(g)}(g)}{d_L}\right),$$

$$\delta^{POST,PRE}(g,k) = \int_{d_L}^{d_U} w_1^{across}(g,k,l)\overline{ACR}^{POST(k)}(g,l)\,dl$$
$$- \left\{ \int_{d_L}^{d_U} \tilde{w}_1^{across}(g,k,l) \left( \frac{\partial \pi_D^{POST(k),PRE(g)}(k,l)}{\partial l} - \frac{\partial \pi_D^{POST(k),PRE(g)}(g,l)}{\partial l} \right) dl \right.$$
$$\left. + \tilde{w}_0^{across}(g,k) \left( \frac{\pi_D^{POST(k),PRE(g)}(k,d_L) - \pi_D^{POST(k),PRE(g)}(g,d_L)}{d_L} \right) \right\}.$$

*In addition, (i)* $w_1^{within}(g,d) \geq 0$*,* $w_1(g,d) \geq 0$*, and* $w_0(g) \geq 0$*, for all* $g \in \mathcal{G}$ *and* $d \in \mathcal{D}_+^c$ *and (ii)* $\int_{d_L}^{d_U} w_1^{within}(g,l)\,dl = 1$*,* $\int_{d_L}^{d_U} w_1(g,l)\,dl + w_0(g) = 1$*, and* $\int_{d_L}^{d_U} w_1^{across}(g,k,l)\,dl = 1$*.*

Part (1) of Theorem S3 links the four sets of comparisons in the TWFE estimator in Proposition S1 to derivatives of conditional expectations (this is broadly similar to Equation (B.7) in the proof of Theorem 3.4 in the main text) along with some additional (nuisance) paths of outcomes.

Part (2) of Theorem S3 imposes the multi-period version of strong parallel trends in Assumption 5-MP-Extended(a).[7] Under Assumption 5-MP-Extended(a), $\delta^{WITHIN}(g)$ and $\delta^{MID,PRE}(g,k)$ both deliver weighted averages of $ACR$-type parameters. However, $\delta^{POST,MID}(g,k)$ and $\delta^{POST,PRE}(g,k)$ still involve non-negligible nuisance terms. Under Assumption 5-MP-Extended(a), the additional term in $\delta^{POST,MID}(g,k)$ involves the difference between treatment effects for group $g$ in group $k$'s post-treatment periods relative to treatment effects for group $g$ in the periods after group $g$ is treated but before group $k$ is treated—that is, treatment effect dynamics. Parallel trends assumptions do not imply that this term is equal to 0. And, in the special case where the treatment is binary, this term corresponds to the "problematic" term related to treatment effect dynamics in Goodman-Bacon (2021).

The additional nuisance term in $\delta^{POST,PRE}(g,k)$ involves differences in partial effects of more treatment across groups in their common post-treatment periods. Parallel trends does not restrict these partial effects to be equal to each other. This term does not show up in the case with a binary treatment because, by construction, the distribution of the dose is the same across groups. It is helpful to further consider where this expression comes from. For simplicity, temporarily suppose that the partial effect of more dose is positive and constant across groups, time, and dose. In this case, if group $g$ has more dose on average than group $k$, then its outcomes should increase more from group $g$ and $k$'s common pre-treatment period to their common post-treatment period. This is the comparison that shows up in $\delta^{POST,PRE}(g,k)$. However, when partial effects are not the same across groups and times (which is not implied by any parallel trends assumption), then, for example, it could be the case that the partial effect of dose is positive for all groups and time periods but greater for group $k$ relative to group $g$. If these differences are large enough, it could lead to the cross-group, long-difference comparisons in $\delta^{POST,PRE}(g,k)$ having the opposite sign.

Next, we discuss what sort of extra conditions can (i) guarantee that $\beta^{twfe}$ is a (positively) weighted average of underlying causal responses or (ii) for $\beta^{twfe} = ACR^o$. To do so, one must further restrict different types of treatment effect heterogeneity.

[7] In Theorem S3-Extended below, we provide an analogous result under the parallel trends assumption in Assumption 4-MP-Extended(a).

**Assumption S1** (Assumptions Limiting Treatment Effect Heterogeneity)**.**

*(a) [No Treatment Effect Dynamics] For all $g \in \mathcal{G} \setminus (T+1)$ and $t \geq g$ (i.e, post-treatment periods for group $g$), $ACR(g,t,d)$ and $ATE(g,t,d_L)$ do not vary with $t$.*

*(b) [Homogeneous Causal Responses across Groups] For all $g \in \mathcal{G} \setminus (T+1)$ with $t \geq g$ and $k \in \mathcal{G} \setminus (T+1)$ with $t \geq k$, $ACR(g,t,d) = ACR(k,t,d)$ and $ATE(g,t,d_L) = ATE(k,t,d_L)$.*

*(c) [Homogeneous Causal Responses across Dose] For all $g \in \mathcal{G} \setminus (T+1)$ with $t \geq g$, $ACR(g,t,d)$ does not vary across $d$, and, in addition, $ATE(g,t,d_L)/d_L = ACR(g,t,d)$.*

Assumption S1 introduces three additional conditions limiting treatment effect heterogeneity. Assumption S1(a) imposes that, within a timing-group, the causal response to the treatment does not vary across time which rules out treatment effect dynamics. Assumption S1(b) imposes that, for a fixed time period, causal responses to the treatment are constant across timing-groups. Assumption S1(c) imposes that, within timing-group and time period, the causal response to more dose is constant across different values of the dose.

**Proposition S2.** *Under Assumptions 1-MP, 2-MP, 3-MP, and 5-MP-Extended(a),*

*(a) If, in addition, Assumption S1(a) holds, then*

$$\delta^{POST,MID}(g,k) = \int_{d_L}^{d_U} w_1(k,l)\overline{ACR}^{POST(k)}(k,l)\,dl + w_0(k)\frac{\overline{ATE}^{POST(k)}(k,d_L)}{d_L}.$$

*(b) If, in addition, Assumption S1(b) holds, then*

$$\delta^{POST,PRE}(g,k) = \int_{d_L}^{d_U} w_1^{across}(g,k,l)\overline{ACR}^{POST(k)}(g,l)\,dl.$$

*(c) If, in addition, Assumption S1(a), (b) and (c) hold, then*

$$\beta^{twfe} = ACR^o.$$

Proposition S2 provides additional conditions under which the nuisance terms in $\delta^{POST,MID}(g,k)$ and $\delta^{POST,PRE}(g,k)$ are equal to 0. For $\delta^{POST,MID}(g,k)$, these nuisance terms will be equal to 0 if there are no treatment effect dynamics; that is, the causal response to more dose does not vary across time. Ruling out these sorts of treatment effect dynamics is analogous to the kinds of conditions that are required to rule out negative weights TWFE estimates with a binary treatment. For $\delta^{POST,PRE}(g,k)$, the nuisance terms will be equal to 0 if there are homogeneous causal responses across groups—that the causal response to more dose is the same across groups conditional on having the same amount of dose and being in the same time period. Neither of these assumptions is implied by any of the parallel trends assumptions that we have considered, and they are both potentially very strong. Therefore, under both Assumption S1(a) and (b), $\beta^{twfe}$ is equal to a weighted average of average causal response parameters, but these weights continue to be driven by the TWFE estimation strategy and, like in the baseline two-period case, can continue to deliver poor estimates of the overall average causal response to the treatment. If all of the conditions in Assumption S1(a), (b), and (c) hold, then it implies that $ACR(g,t,d)$ does not vary by timing group, time period, or the amount of dose, and part (c) of Proposition S2 says that $\beta^{twfe}$ is equal to the overall average causal response under these additional, strong conditions.

The results in part (2) of Theorem S3 and in Proposition S2 relied on the multi-period version of strong parallel trends in Assumption 5-MP-Extended(a). To conclude this section, we provide a version of Theorem S3 under the parallel trends assumption in Assumption 4-MP-Extended(a) that only involves paths of untreated potential outcomes rather than strong parallel trends.

**Theorem S3-Extended.** *Under Assumptions 1-MP, 2-MP, 3-MP, and 4-MP-Extended(a),*

$$\delta^{WITHIN}(g) = \int_{d_L}^{d_U} w_1^{within}(g,l) \left( \overline{ACRT}^{POST(g)}(g,l|g,l) + \frac{\partial \overline{ATT}^{POST(g)}(g,l|g,h)}{\partial h}\Big|_{h=l} \right) dl$$

$$\delta^{MID,PRE}(g,k) = \int_{d_L}^{d_U} w_1(g,l) \left( \overline{ACRT}^{MID(g,k)}(g,l|g,l) + \frac{\partial \overline{ATT}^{MID(g,k)}(g,l|g,h)}{\partial h}\Big|_{h=l} \right) dl + w_0(g) \frac{\overline{ATT}^{MID(g,k)}(g,d_L|g,d_L)}{d_L}$$

$$\delta^{POST,MID}(g,k) = \int_{d_L}^{d_U} w_1(k,l) \left( \overline{ACRT}^{POST(k)}(k,l|k,l) + \frac{\partial \overline{ATT}^{POST(k)}(k,l|k,h)}{\partial h}\Big|_{h=l} \right) dl + w_0(k) \frac{\overline{ATT}^{POST(k)}(k,d_L|k,d_L)}{d_L} - w_0(k) \left( \frac{\pi^{POST(k),PRE(g)}(g) - \pi^{MID(g,k),PRE(g)}(g)}{d_L} \right)$$

$$\begin{aligned} \delta^{POST,PRE}(g,k) = & \int_{d_L}^{d_U} w_1^{across}(g,k,l) \left( \overline{ACRT}^{POST(k)}(g,l|g,l) + \frac{\partial \overline{ATT}(g,l|g,h)}{\partial h}\Big|_{h=l} \right) dl \\ & - \Bigg\{ \int_{d_L}^{d_U} \tilde{w}_1^{across}(g,k,l) \left( \frac{\partial \pi_D^{POST(k),PRE(g)}(k,l)}{\partial l} - \frac{\partial \pi_D^{POST(k),PRE(g)}(g,l)}{\partial l} \right) dl \\ & + \tilde{w}_0^{across}(g,k) \left( \frac{\pi_D^{POST(k),PRE(g)}(k,d_L) - \pi_D^{POST(k),PRE(g)}(g,d_L)}{d_L} \right) \Bigg\} \end{aligned}$$

*where the weights are the same as in Theorem S3 and satisfy the same properties.*

This result is similar to part (2) of Appendix SA.3 except that $\overline{ACR}^{\cdot}(\cdot,d)$ should be replaced by $\overline{ACRT}^{\cdot}(\cdot,d|\cdot,d) + \frac{\partial \overline{ATT}^{\cdot}(\cdot,d|\cdot,l)}{\partial l}\Big|_{l=d}$ where the second term is a selection bias term, and $\overline{ATE}^{\cdot}(\cdot,d_L)$ should be replaced by $\overline{ATT}^{\cdot}(\cdot,d_L|\cdot,d_L)$. The main additional takeaway from Theorem S3-Extended is that, under a standard version of parallel trends, all four comparisons in Equations (S3) to (S6) include selection bias terms.

## SA.4 Discussion

The results in this section suggest four important weaknesses of TWFE estimands in a difference-in-differences framework with multiple time periods, and variation in treatment timing and intensity. First, all of the results in this section have used the strongest versions of the parallel trends assumptions discussed above (Assumptions 4-MP-Extended(a) or 5-MP-Extended(a)) which involve parallel trends holding across all periods (including pre-treatment periods). If there are violations of parallel trends in pre-treatment periods, these violations will contribute to $\beta^{twfe}$.

Second, like the TWFE estimands considered in the main text in the case with two time periods, TWFE estimands have weights that are driven by the estimation method. These weights may have undesirable properties in settings where there is treatment effect heterogeneity.

Third, in addition to reasonable treatment effect parameters, TWFE estimands also include undesirable components due to treatment effect dynamics and heterogeneous causal responses across groups and time periods. That these show up in the TWFE estimand is potentially problematic and can possibly lead to very poor performance of the TWFE estimator. Ruling out these problems requires substantially stronger conditions in addition to any kind of parallel trends assumption.

Finally, even when these extra conditions hold (i.e., the best case scenario for TWFE), if a researcher invokes a standard parallel trends assumption, the TWFE estimand delivers weighted averages of derivatives of $ATT$-type parameters which are themselves hard to interpret because, like in the two-period case, they include both actual causal responses and selection bias terms.

Of these four weaknesses, the first three can be completely avoided by using the estimands presented in Theorem S1. These estimands rely only on parallel trends assumptions; in particular, they are available without imposing any conditions on treatment effect dynamics or how causal responses vary across groups. The fourth weakness, though, is a more fundamental challenge of difference-in-differences approaches with variation in treatment intensity as comparing treatment effect parameters across different values of the dose appears to fundamentally require imposing stronger assumptions that rule out some forms of selection into different amounts of the treatment. Although undesirable, we are not aware of any other practical solution to this empirically relevant DiD problem. Thus, we urge practitioners to (i) transparently discuss their assumptions, potentially exploiting context-specific knowledge to justify the plausibility of a stronger parallel trends assumption in the given application or (ii) to focus on other parameters that do not involve comparisons across doses.

## SB Proofs of Results from Section D and Appendix SA

This section contains the proofs of results from Appendix D in the main text and Appendix SA, which encompasses our results on DiD with a continuous treatment and with multiple periods and variation in treatment timing and dose intensity.

### SB.1 Proof of Results from Appendix SA.2

This section proves Theorem S1 and Theorem S2; note that Theorem D.1, in the main text, corresponds to part (2a) of Theorem S1 (under Assumption 5-MP-Extended(a)).

#### Proof of Theorem S1

*Proof.* For part (1a), we show the result under Assumption 4-MP-Extended(c), which is strictly weaker than Assumption 4-MP-Extended(a). First, notice that,

$$\begin{aligned} ATT(g,t,d|g,d) &= \mathbb{E}[Y_t(g,d) - Y_t(0)|G=g, D=d] \\ &= \mathbb{E}[Y_t(g,d) - Y_{g-1}(0)|G=g, D=d] - \mathbb{E}[Y_t(0) - Y_{g-1}(0)|G=g, D=d] \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}[Y_t(g,d) - Y_{g-1}(0)|G=g, D=d] - \sum_{s=g}^{t} \mathbb{E}[Y_s(0) - Y_{s-1}(0)|G=g, D=d] \quad \text{(S7)} \\
&= \mathbb{E}[Y_t(g,d) - Y_{g-1}(0)|G=g, D=d] - \sum_{s=g}^{t} \mathbb{E}[Y_s(0) - Y_{s-1}(0)|W_t=0] \\
&= \mathbb{E}[Y_t(g,d) - Y_{g-1}(0)|G=g, D=d] - \mathbb{E}[Y_t(0) - Y_{g-1}(0)|W_t=0] \\
&= \mathbb{E}[Y_t - Y_{g-1}|G=g, D=d] - \mathbb{E}[Y_t - Y_{g-1}|W_t=0]
\end{aligned}
$$

where the first equality is the definition of $ATT(g,t,d|g,d)$, the second equality holds by adding and subtracting $\mathbb{E}[Y_{g-1}(0)|G=g, D=d]$, the third equality holds by adding and subtracting $\mathbb{E}[Y_s(0)|G = g, D = d]$ for $s = g, \ldots, (t-1)$, the fourth equality holds under Assumption 4-MP-Extended(c), the fifth equality holds by canceling all the terms involving $\mathbb{E}[Y_s(0)|W_t = 0]$ for $s = g, \ldots, (t-1)$ (i.e., from the reverse of the argument for the third equality), and the last equality holds from writing the potential outcomes in terms of their observed counterparts.

For part (1b), in Equation (S7), $\sum_{s=g}^{t} \mathbb{E}[Y_s(0) - Y_{s-1}(0)|G=g, D=d] = \sum_{s=g}^{t} \mathbb{E}[Y_s(0) - Y_{s-1}(0)|D = 0]$ under Assumption 4-MP-Extended(b). Then, the result holds by otherwise following the same arguments as in part (1a).

For part (2a), we show the result under Assumption 5-MP-Extended(c) which is strictly weaker than Assumption 5-MP-Extended(a). Notice that,

$$
\begin{aligned}
ATE(g,t,d) &= \mathbb{E}[Y_t(g,d) - Y_t(0)|G=g] \\
&= \mathbb{E}[Y_t(g,d) - Y_{g-1}(g,d)|G=g] - \mathbb{E}[Y_t(0) - Y_{g-1}(0)|G=g] \\
&= \sum_{s=g}^{t} \mathbb{E}[Y_s(g,d) - Y_{s-1}(g,d)|G=g] - \sum_{s=g}^{t} \mathbb{E}[Y_s(0) - Y_{s-1}(0)|G=g] \quad \text{(S8)} \\
&= \sum_{s=g}^{t} \mathbb{E}[Y_s(g,d) - Y_{s-1}(g,d)|G=g, D=d] - \sum_{s=g}^{t} \mathbb{E}[Y_s(0) - Y_{s-1}(0)|W_t=0] \\
&= \mathbb{E}[Y_t(g,d) - Y_{g-1}(g,d)|G=g, D=d] - \mathbb{E}[Y_t(0) - Y_{g-1}(0)|W_t=0] \\
&= \mathbb{E}[Y_t - Y_{g-1}|G=g, D=d] - \mathbb{E}[Y_t - Y_{g-1}|W_t=0]
\end{aligned}
$$

where the first equality holds by the definition of $ATE(g,t,d)$, the second equality adds and subtracts $\mathbb{E}[Y_{g-1}(g,d)|G = g]$ (this equation also uses the no anticipation condition in Assumption 3-MP which implies that $\mathbb{E}[Y_{g-1}(g,d)|G = g] = \mathbb{E}[Y_{g-1}(0)|G = g]$), the third equality holds by writing both "long differences" as summations over "short differences", the fourth equality holds by Assumption 5-MP-Extended(c), the fifth equality holds by canceling all of the intermediate terms in the summations over short differences, and the last equality holds by writing potential outcomes in terms of their corresponding observed outcomes and is the result.

Finally, for part (2b), in Equation (S8), $\sum_{s=g}^{t} \mathbb{E}[Y_s(0) - Y_{s-1}(0)|G = g] = \sum_{s=g}^{t} \mathbb{E}[Y_s(0) - Y_{s-1}(0)|D = 0]$ under Assumption 5-MP-Extended(b). The result then follows using the same subsequent arguments as in part (2a). □

**Proof of Theorem S2**

*Proof.* To start with, notice that

$$\frac{\partial}{\partial d}\mathbb{E}[Y_t - Y_{g-1}|G=g, D=d] = \frac{\partial}{\partial d}\Big\{\mathbb{E}[Y_t - Y_{g-1}|G=g, D=d] - \mathbb{E}[Y_t - Y_{g-1}|W_t = 0]\Big\} \tag{S9}$$

which holds because the second term does not depend on $d$. Thus, under Assumption 4-MP-Extended(a), we have that

$$\begin{aligned}\frac{\partial}{\partial d}\mathbb{E}[Y_t - Y_{g-1}|G=g, D=d] &= \frac{\partial}{\partial d}ATT(g,t,d|g,d) \\ &= ACRT(g,t,d|g,d) + \frac{\partial ATT(g,t,d|g,l)}{\partial l}\bigg|_{l=d}\end{aligned}$$

where the first equality holds by Equation (S9) and Theorem S1(1a), and the second equality holds by the linearity of differentiation and the definition of $ACRT(g,t,d|g,d)$.

Under Assumption 5-MP-Extended(2a), we have that

$$\begin{aligned}\frac{\partial}{\partial d}\mathbb{E}[Y_t - Y_{g-1}|G=g, D=d] &= \frac{\partial}{\partial d}ATE(g,t,d) \\ &= ACR(g,t,d)\end{aligned}$$

where the first equality holds by Equation (S9) and Theorem S1(c), and the second equality holds by the definition of $ACR(g,t,d)$. This completes the proof. ☐

## SB.2 Proofs of Results from Appendix SA.3

This section contains the proofs for interpreting TWFE regressions in the case with a continuous treatment, multiple periods, and variation in treatment timing as in Appendix SA.3.

Before proving the main results in this section, we introduce some additional notation. Let

$$v(g,t) = \mathbf{1}\{t \geq g\} - \bar{G}_g \tag{S10}$$

where the term $\mathbf{1}\{t \geq g\}$ is equal to one in post-treatment time periods for units in group $g$ and recalling that we defined $\bar{G}_g = \frac{T-g+1}{T}$ which is the fraction of periods that units in group $g$ are exposed to the treatment (and notice that this latter term does not depend on the particular time period $t$). Further, notice that $v(g,t)$ is positive in post-treatment time periods and negative in pre-treatment time periods for units in a particular group. Finally, also note that, for the "never-treated" group, $g = T+1$, so that both terms in the expression for $v$ are equal to 0.

Furthermore, recall that, for $1 \leq t_1 \leq t_2 \leq T$, we defined

$$\bar{Y}_i^{(t_1,t_2)} = \frac{1}{t_2 - t_1 + 1}\sum_{t=t_1}^{t_2} Y_{i,t}$$

where below (and following the notation used throughout the paper), we sometimes leave the subscript $i$ implicit.

We next state and prove some additional results that are helpful for proving the main results. The first lemma re-writes (overall) expected dose experienced in period $t$ adjusted by the overall expected dose (across periods and units) in a form that is useful in proving later results.

**Lemma S1.** *Under Assumptions 1-MP, 2-MP(a), and 3-MP,*

$$\mathbb{E}[W_t] - \frac{1}{T}\sum_{s=1}^{T}\mathbb{E}[W_s] = \sum_{g\in\mathcal{G}}\int_{\mathcal{D}} dv(g,t)\, dF_{D|G}(d|g)p_g$$

*Proof.* First, notice that

$$\begin{aligned}\mathbb{E}[W_t] &= \sum_{g\in\mathcal{G}}\int_{\mathcal{D}} \mathbb{E}[W_t|G=g,D=d]\, dF_{D|G}(d|g)p_g \\ &= \sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\mathbf{1}\{t\geq g\}\, dF_{D|G}(d|g)p_g \qquad \text{(S11)}\end{aligned}$$

where the first equality holds by the law of iterated expectations and the second equality holds because, after conditioning on group and dose, $W_t$ is fully determined. Thus,

$$\begin{aligned}\mathbb{E}[W_t] - \frac{1}{T}\sum_{s=1}^{T}\mathbb{E}[W_s] &= \sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\mathbf{1}\{t\geq g\}\, dF_{D|G}(d|g)p_g - \frac{1}{T}\sum_{s=1}^{T}\sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\mathbf{1}\{s\geq g\}\, dF_{D|G}(d|g)p_g \\ &= \frac{1}{T}\sum_{s=1}^{T}\sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\left(\mathbf{1}\{t\geq g\} - \mathbf{1}\{s\geq g\}\right)\, dF_{D|G}(d|g)p_g \\ &= \sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\left\{\frac{1}{T}\sum_{s=1}^{T}\mathbf{1}\{t\geq g\} - \mathbf{1}\{s\geq g\}\right\} dF_{D|G}(d|g)p_g \\ &= \sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\left\{\mathbf{1}\{t\geq g\} - \frac{T-g+1}{T}\right\} dF_{D|G}(d|g)p_g \\ &= \sum_{g\in\mathcal{G}}\int_{\mathcal{D}} dv(g,t)\, dF_{D|G}(d|g)p_g\end{aligned}$$

where the first equality applies Equation (S11) to both terms, the second equality combines terms by averaging the first term across time periods, the third equality re-orders the summations/integrals, the fourth equality holds because $\mathbf{1}\{t\geq g\}$ does not depend on $s$ and by counting the fraction of periods where $s\geq g$, and the last equality holds by the definition of $v(g,t)$. □

The next lemma provides an intermediate result for the expression for the numerator of $\beta^{twfe}$ in Equation (S2).

**Lemma S2.** *Under Assumptions 1-MP, 2-MP(a), and 3-MP,*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[Y_{i,t}\ddot{W}_{i,t}] = \frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\left(\mathbb{E}[Y_t|G=g,D=d] - \mathbb{E}[Y_t]\right)v(g,t)\, dF_{D|G}(d|g)p_g\right\}$$

*Proof.* Starting with the term on the left-hand side, we have that

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[Y_{i,t}\ddot{W}_{i,t}] \\ &= \frac{1}{T}\sum_{t=1}^{T}\left\{\mathbb{E}[Y_{i,t}W_{i,t}] - \mathbb{E}[Y_{i,t}\bar{W}_i] - \mathbb{E}[Y_t]\left(\mathbb{E}[W_t] - \frac{1}{T}\sum_{s=1}^{T}\mathbb{E}[W_s]\right)\right\}\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{T}\sum_{t=1}^{T}\left\{\mathbb{E}[Y_t D \mathbf{1}\{t \geq G\}] - \mathbb{E}\left[Y_t D \frac{T-G+1}{T}\right] - \mathbb{E}[Y_t]\left(\mathbb{E}[W_t] - \frac{1}{T}\sum_{s=1}^{T}\mathbb{E}[W_s]\right)\right\} \\
&= \frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{g\in\mathcal{G}}\int_{\mathcal{D}}\left(\mathbb{E}[Y_t d \mathbf{1}\{t \geq g\}|G=g, D=d] - \mathbb{E}\left[Y_t \frac{T-g+1}{T} d \Big| G=g, D=d\right]\right) dF_{D|G}(d|g)p_g \right.\\
&\qquad\qquad \left. - \mathbb{E}[Y_t]\left(\mathbb{E}[W_t] - \frac{1}{T}\sum_{s=1}^{T}\mathbb{E}[W_s]\right)\right\} \\
&= \frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\Big(\mathbb{E}[Y_t|G=g, D=d]v(g,t)\Big)\, dF_{D|G}(d|g)p_g - \mathbb{E}[Y_t]\left(\mathbb{E}[W_t] - \frac{1}{T}\sum_{s=1}^{T}\mathbb{E}[W_s]\right)\right\} \\
&= \frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\Big(\mathbb{E}[Y_t|G=g, D=d]v(g,t)\Big)\, dF_{D|G}(d|g)p_g - \mathbb{E}[Y_t]\left(\sum_{g\in\mathcal{G}}\int_{\mathcal{D}} dv(g,t)\, dF_{D|G}(d|g)p_g\right)\right\} \\
&= \frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\left(\mathbb{E}[Y_t|G=g, D=d] - \mathbb{E}[Y_t]\right) v(g,t)\, dF_{D|G}(d|g)p_g\right\}
\end{aligned}
$$

where the first equality holds by the definition of $\ddot{W}_{i,t}$, the second equality holds by plugging in for $W_{i,t}$ and $\bar{W}_i$, the third equality holds by the law of iterated expectations, the fourth equality holds by the definition of $v(g,t)$, the fifth equality holds by Lemma S1, and the sixth equality just combines terms. □

Next, based on the result in Lemma S2, we can write the numerator in the expression for $\beta^{twfe}$ as

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[Y_{i,t}\ddot{W}_{i,t}] \\
&\quad= \frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\left(\mathbb{E}[Y_t|G=g, D=d] - \mathbb{E}[Y_t]\right) v(g,t)\, dF_{D|G}(d|g)p_g\right\} \\
&\quad= \frac{1}{T}\sum_{t=1}^{T}\sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\Big(\mathbb{E}[Y_t|G=g, D=d] - \mathbb{E}[Y_t|G=g]\Big)v(g,t)\, dF_{D|G}(d|g)p_g && \text{(S12)} \\
&\qquad\quad + \frac{1}{T}\sum_{t=1}^{T}\sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\Big(\mathbb{E}[Y_t|G=g] - \mathbb{E}[Y_t]\Big)v(g,t)\, dF_{D|G}(d|g)p_g && \text{(S13)}
\end{aligned}
$$

where the first equality holds from Lemma S2 and the second equality holds by adding and subtracting $\mathbb{E}[Y_t|G=g]$.

The expression in Equation (S12) involves comparisons between units in the same group but that have different doses. The expression in Equation (S13) involves comparisons across different groups. We consider each of these terms in more detail below.

**Lemma S3.** *Under Assumptions 1-MP, 2-MP(a), and 3-MP,*

$$
\frac{1}{T}\sum_{t=1}^{T}\sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\Big(\mathbb{E}[Y_t|G=g, D=d] - \mathbb{E}[Y_t|G=g]\Big)v(g,t)\, dF_{D|G}(d|g)p_g
$$

$$= \sum_{g \in \mathcal{G}} \left\{ (1-\bar{G}_g)\bar{G}_g \text{Cov}\left(\bar{Y}^{POST(g)} - \bar{Y}^{PRE(g)}, D|G=g\right) \right\} p_g$$

*Proof.* Notice that

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\Big(\mathbb{E}[Y_t|G=g,D=d] - \mathbb{E}[Y_t|G=g]\Big)v(g,t)\,dF_{D|G}(d|g)p_g \\
&\qquad = \frac{1}{T}\sum_{t=1}^{T}\left\{ \sum_{g\in\mathcal{G}} \mathbb{E}[Y_t(D-\mathbb{E}[D|G=g])|G=g]v(g,t)p_g \right\} \\
&\qquad = \sum_{g\in\mathcal{G}}\left\{ \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[Y_t(D-\mathbb{E}[D|G=g])|G=g]v(g,t) \right\}p_g \\
&\qquad = \sum_{g\in\mathcal{G}}\left\{ -\frac{1}{T}\frac{(T-g+1)}{T}\sum_{t=1}^{g-1} \mathbb{E}[Y_t(D-\mathbb{E}[D|G=g])|G=g] \right. \\
&\qquad\qquad\qquad \left. + \frac{1}{T}\frac{(g-1)}{T}\sum_{t=g}^{T} \mathbb{E}[Y_t(D-\mathbb{E}[D|G=g])|G=g] \right\}p_g \\
&\qquad = \sum_{g\in\mathcal{G}}\left\{ \frac{g-1}{T}\frac{(T-g+1)}{T}\left( \frac{1}{T-g+1}\sum_{t=g}^{T} \mathbb{E}[Y_t(D-\mathbb{E}[D|G=g])|G=g] \right.\right. \\
&\qquad\qquad\qquad \left.\left. - \frac{1}{g-1}\sum_{t=1}^{g-1} \mathbb{E}[Y_t(D-\mathbb{E}[D|G=g])|G=g] \right) \right\}p_g \\
&\qquad = \sum_{g\in\mathcal{G}}\left\{ \frac{g-1}{T}\frac{(T-g+1)}{T}\left( \mathbb{E}\big[\big(\bar{Y}^{POST(g)} - \bar{Y}^{PRE(g)}\big)(D-\mathbb{E}[D|G=g])|G=g\big] \right) \right\}p_g \\
&\qquad = \sum_{g\in\mathcal{G}}\left\{ (1-\bar{G}_g)\bar{G}_g\left( \mathbb{E}\big[\big(\bar{Y}^{POST(g)} - \bar{Y}^{PRE(g)}\big)(D-\mathbb{E}[D|G=g])|G=g\big] \right) \right\}p_g \\
&\qquad = \sum_{g\in\mathcal{G}}\left\{ (1-\bar{G}_g)\bar{G}_g \text{Cov}\left(\bar{Y}^{POST(g)} - \bar{Y}^{PRE(g)}, D|G=g\right) \right\}p_g
\end{aligned}$$

where the first equality holds by the law of iterated expectations (and combining terms involving $d$ and $Y_t$), the second equality changes the order of the summations, the third equality holds by splitting the summation involving $t$ in time period $g$ and plugs in for $v(g,t)$ (which is constant within group $g$ and across time periods from $1,\ldots,g-1$ and from $g,\ldots,T$), the fourth equality multiplies and divides by terms so that the inside expressions can be written as averages, the fifth equality holds by changing the order of the expectation and averaging over time periods, the sixth equality holds by the definition of $\bar{G}_g$, and the last equality holds by the definition of covariance. □

Lemma S3 shows that part of the TWFE estimator comes from a weighted average of post- vs. pre-treatment outcomes within group but who experienced different doses. In particular, notice that for units in group $g$, $\bar{Y}_i^{POST(g)}$ is their average post-treatment outcome while $\bar{Y}_i^{PRE(g)}$ is their average pre-treatment outcome.

Next, we consider the expression from Equation (S13) above which arises from differences in outcomes across groups. We handle this term over several following results.

**Lemma S4.** *Under Assumptions 1-MP, 2-MP(a), and 3-MP,*

$$\frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\Big(\mathbb{E}[Y_t|G=g]-\mathbb{E}[Y_t]\Big)v(g,t)\,dF_{D|G}(d|g)p_g\right\}$$
$$=\frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{g\in\mathcal{G}}\sum_{k\in\mathcal{G},k>g}\Big(\mathbb{E}[D|G=g]v(g,t)-\mathbb{E}[D|G=k]v(k,t)\Big)\Big(\mathbb{E}[Y_t|G=g]-\mathbb{E}[Y_t|G=k]\Big)p_kp_g\right\}$$

*Proof.* Notice that

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\Big(\mathbb{E}[Y_t|G=g]-\mathbb{E}[Y_t]\Big)v(g,t)\,dF_{D|G}(d|g)p_g\right\}\\
&\quad=\frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{g\in\mathcal{G}}\mathbb{E}[D|G=g]\Big(\mathbb{E}[Y_t|G=g]-\mathbb{E}[Y_t]\Big)v(g,t)p_g\right\}\\
&\quad=\frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{g\in\mathcal{G}}\mathbb{E}[D|G=g]\Big(\mathbb{E}[Y_t|G=g]-\sum_{k\in\mathcal{G}}\mathbb{E}[Y_t|G=k]p_k\Big)v(g,t)p_g\right\}\\
&\quad=\frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{g\in\mathcal{G}}\sum_{k\in\mathcal{G}}\mathbb{E}[D|G=g]v(g,t)\Big(\mathbb{E}[Y_t|G=g]-\mathbb{E}[Y_t|G=k]\Big)p_kp_g\right\}\\
&\quad=\frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{g\in\mathcal{G}}\sum_{k\in\mathcal{G},k>g}\Big(\mathbb{E}[D|G=g]v(g,t)-\mathbb{E}[D|G=k]v(k,t)\Big)\Big(\mathbb{E}[Y_t|G=g]-\mathbb{E}[Y_t|G=k]\Big)p_kp_g\right\}
\end{aligned}$$

where the first equality holds by integrating over $\mathcal{D}$, the second equality holds by the law of iterated expectations, the third equality holds by combining terms, and the last equality holds because all combinations of $g$ and $k$ occur twice. □

Lemma S4 is helpful because it shows that the cross-group part of the TWFE estimator can be written as comparisons for each group relative to later-treated groups.

Next, we provide an important intermediate result. Before stating this result, we define the following weights

$$\begin{aligned}
\tilde{w}^{g,within}(g)&=\mathrm{Var}(D|G=g)(1-\bar{G}_g)\bar{G}_gp_g\\
\tilde{w}^{g,post}(g,k)&=\mathbb{E}[D|G=g]^2(1-\bar{G}_g)(\bar{G}_g-\bar{G}_k)p_kp_g\\
\tilde{w}^{k,post}(g,k)&=\mathbb{E}[D|G=k]^2\bar{G}_k(\bar{G}_g-\bar{G}_k)p_kp_g\\
\tilde{w}^{long}(g,k)&=(\mathbb{E}[D|G=g]-\mathbb{E}[D|G=k])^2\bar{G}_k(1-\bar{G}_g)p_kp_g
\end{aligned}$$

which correspond to $w^{g,post}$, $w^{k,post}$, and $w^{long}(g,k)$ above except they do not divide by $T^{-1}\sum_{t=1}^{T}\mathbb{E}[\ddot{W}_{i,t}^2]$. In addition, notice that

$$\begin{aligned}
&\mathbb{E}[D|G=g]v(g,t)-\mathbb{E}[D|G=k]v(k,t)\\
&\quad=\begin{cases}-\mathbb{E}[D|G=g]\bar{G}_g+\mathbb{E}[D|G=k]\bar{G}_k & \text{for } t<g<k\\ \mathbb{E}[D|G=g](1-\bar{G}_g)+\mathbb{E}[D|G=k]\bar{G}_k & \text{for } g\le t<k\\ \mathbb{E}[D|G=g](1-\bar{G}_g)-\mathbb{E}[D|G=k](1-\bar{G}_k) & \text{for } g<k\le t\end{cases}
\end{aligned}\tag{S14}$$

which holds by the definition of $v$ and is useful for the proof of the following lemma.

**Lemma S5.** *Under Assumptions 1-MP, 2-MP(a), and 3-MP,*

$$\frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\Big(\mathbb{E}[Y_t|G=g]-\mathbb{E}[Y_t]\Big)v(g,t)\,dF_{D|G}(d|g)p_g\right\}$$
$$=\sum_{g\in\mathcal{G}}\sum_{k\in\mathcal{G},k>g}\Bigg\{\tilde{w}^{g,post}(g,k)\bigg(\mathbb{E}\left[\big(\bar{Y}^{MID(g,k)}-\bar{Y}^{PRE(g)}\big)|G=g\right]-\mathbb{E}\left[\big(\bar{Y}^{MID(g,k)}-\bar{Y}^{PRE(g)}\big)|G=k\right]\bigg)$$
$$+\tilde{w}^{k,post}(g,k)\bigg(\mathbb{E}\left[\big(\bar{Y}^{POST(k)}-\bar{Y}^{MID(g,k)}\big)|G=k\right]-\mathbb{E}\left[\big(\bar{Y}^{POST(k)}-\bar{Y}^{MID(g,k)}\big)|G=g\right]\bigg)$$
$$+\tilde{w}^{long}(g,k)\bigg(\mathbb{E}\left[\big(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\big)|G=g\right]-\mathbb{E}\left[\big(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\big)|G=k\right]\bigg)\Bigg\}$$

*Proof.* The result holds as follows

$$\frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{g\in\mathcal{G}}\int_{\mathcal{D}} d\Big(\mathbb{E}[Y_t|G=g]-\mathbb{E}[Y_t]\Big)v(g,t)\,dF_{D|G}(d|g)p_g\right\}$$
$$=\sum_{g\in\mathcal{G}}\sum_{k\in\mathcal{G},k>g}\left\{\frac{1}{T}\sum_{t=1}^{T}\Big(\mathbb{E}[D|G=g]v(g,t)-\mathbb{E}[D|G=k]v(k,t)\Big)\Big(\mathbb{E}[Y_t|G=g]-\mathbb{E}[Y_t|G=k]\Big)\right\}p_kp_g$$
$$=\sum_{g\in\mathcal{G}}\sum_{k\in\mathcal{G},k>g}\Bigg\{\frac{1}{T}\left(-\mathbb{E}[D|G=g]\bar{G}_g+\mathbb{E}[D|G=k]\bar{G}_k\right)\sum_{t=1}^{g-1}\Big(\mathbb{E}[Y_t|G=g]-\mathbb{E}[Y_t|G=k]\Big)$$
$$+\frac{1}{T}\left(\mathbb{E}[D|G=g](1-\bar{G}_g)+\mathbb{E}[D|G=k]\bar{G}_k\right)\sum_{t=g}^{k-1}\Big(\mathbb{E}[Y_t|G=g]-\mathbb{E}[Y_t|G=k]\Big)$$
$$+\frac{1}{T}\left(\mathbb{E}[D|G=g](1-\bar{G}_g)-\mathbb{E}[D|G=k](1-\bar{G}_k)\right)\sum_{t=k}^{T}\Big(\mathbb{E}[Y_t|G=g]-\mathbb{E}[Y_t|G=k]\Big)\Bigg\}p_kp_g$$
$$=\sum_{g\in\mathcal{G}}\sum_{k\in\mathcal{G},k>g}\Bigg\{(1-\bar{G}_g)\left(-\mathbb{E}[D|G=g]\bar{G}_g+\mathbb{E}[D|G=k]\bar{G}_k\right)\Big(\mathbb{E}[\bar{Y}^{PRE(g)}|G=g]-\mathbb{E}[\bar{Y}^{PRE(g)}|G=k]\Big)$$
$$+(\bar{G}_g-\bar{G}_k)\left(\mathbb{E}[D|G=g](1-\bar{G}_g)+\mathbb{E}[D|G=k]\bar{G}_k\right)\Big(\mathbb{E}[\bar{Y}^{MID(g,k)}|G=g]-\mathbb{E}[\bar{Y}^{MID(g,k)}|G=k]\Big)$$
$$+\bar{G}_k\left(\mathbb{E}[D|G=g](1-\bar{G}_g)-\mathbb{E}[D|G=k](1-\bar{G}_k)\right)\Big(\mathbb{E}[\bar{Y}^{POST(k)}|G=g]-\mathbb{E}[\bar{Y}^{POST(k)}|G=k]\Big)\Bigg\}p_kp_g$$
$$=\sum_{g\in\mathcal{G}}\sum_{k\in\mathcal{G},k>g}\Bigg\{(1-\bar{G}_g)\left(-\mathbb{E}[D|G=g](\bar{G}_g-\bar{G}_k)+(\mathbb{E}[D|G=k]-\mathbb{E}[D|G=g])\bar{G}_k\right)\Big(\mathbb{E}[\bar{Y}^{PRE(g)}|G=g]-\mathbb{E}[\bar{Y}^{PRE(g)}|G=k]\Big)$$
$$+(\bar{G}_g-\bar{G}_k)\left(\mathbb{E}[D|G=g](1-\bar{G}_g)+\mathbb{E}[D|G=k]\bar{G}_k\right)\Big(\mathbb{E}[\bar{Y}^{MID(g,k)}|G=g]-\mathbb{E}[\bar{Y}^{MID(g,k)}|G=k]\Big)$$
$$+\bar{G}_k\left((\mathbb{E}[D|G=g]-\mathbb{E}[D|G=k])(1-\bar{G}_g)-\mathbb{E}[D|G=k](\bar{G}_g-\bar{G}_k)\right)\Big(\mathbb{E}[\bar{Y}^{POST(k)}|G=g]-\mathbb{E}[\bar{Y}^{POST(k)}|G=k]\Big)\Bigg\}p_kp_g$$
$$=\sum_{g\in\mathcal{G}}\sum_{k\in\mathcal{G},k>g}\Bigg\{\mathbb{E}[D|G=g](1-\bar{G}_g)(\bar{G}_g-\bar{G}_k)\bigg(\mathbb{E}\left[\big(\bar{Y}^{MID(g,k)}-\bar{Y}^{PRE(g)}\big)|G=g\right]-\mathbb{E}\left[\big(\bar{Y}^{MID(g,k)}-\bar{Y}^{PRE(g)}\big)|G=k\right]\bigg)$$
$$+\mathbb{E}[D|G=k]\bar{G}_k(\bar{G}_g-\bar{G}_k)\bigg(\mathbb{E}\left[\big(\bar{Y}^{POST(k)}-\bar{Y}^{MID(g,k)}\big)|G=k\right]-\mathbb{E}\left[\big(\bar{Y}^{POST(k)}-\bar{Y}^{MID(g,k)}\big)|G=g\right]\bigg)$$
$$+(\mathbb{E}[D|G=g]-\mathbb{E}[D|G=k])\bar{G}_k(1-\bar{G}_g)\bigg(\mathbb{E}\left[\big(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\big)|G=g\right]-\mathbb{E}\left[\big(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\big)|G=k\right]\bigg)\Bigg\}p_kp_g$$
$$=\sum_{g\in\mathcal{G}}\sum_{k\in\mathcal{G},k>g}\Bigg\{\tilde{w}^{g,post}(g,k)\bigg(\mathbb{E}\left[\big(\bar{Y}^{MID(g,k)}-\bar{Y}^{PRE(g)}\big)|G=g\right]-\mathbb{E}\left[\big(\bar{Y}^{MID(g,k)}-\bar{Y}^{PRE(g)}\big)|G=k\right]\bigg)$$
$$+\tilde{w}^{k,post}(g,k)\bigg(\mathbb{E}\left[\big(\bar{Y}^{POST(k)}-\bar{Y}^{MID(g,k)}\big)|G=k\right]-\mathbb{E}\left[\big(\bar{Y}^{POST(k)}-\bar{Y}^{MID(g,k)}\big)|G=g\right]\bigg)$$
$$+\tilde{w}^{long}(g,k)\bigg(\mathbb{E}\left[\big(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\big)|G=g\right]-\mathbb{E}\left[\big(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\big)|G=k\right]\bigg)\Bigg\}$$

where the first equality uses the result in Lemma S4, the second equality changes the order of the summations (splitting them at $g$ and $k$ where the values of $v(g,t)$ and $v(k,t)$ change) and uses Equation (S14), the third equality holds by averaging over time periods (which involves multiplying and dividing by $g-1$ in the first line, multiplying and dividing by $k-g$ in the second line, and multiplying and dividing by $T-k+1$ in the last line), the fourth equality rearranges the expressions for the weights, the fifth equality holds by rearranging terms with common weights, and the last equality holds by the definitions of $\tilde{w}^{g,post}$, $\tilde{w}^{k,post}$, and $\tilde{w}^{long}$ and by noticing that

$$p_k p_g = (p_g + p_k)^2 p_{g|\{g,k\}}(1 - p_{g|\{g,k\}})$$

which holds by multiplying and dividing both $p_k$ and $p_g$ by $(p_g + p_k)$ and by the definition of $p_{g|\{g,k\}}$.

□

The result in Lemma S5 is very closely related to the result on interpreting TWFE regressions with a binary treatment and multiple time periods and variation in treatment timing in Goodman-Bacon (2021).[8] In particular, it says that, even with a continuous or multi-valued discrete treatment, the TWFE regression estimator involves comparisons between (i) the path of outcomes for units that become treated relative to the path of outcomes for units that are not treated yet, (ii) the path of outcomes for units that become treated relative to the path of outcomes for units that have already been treated, and (iii) comparisons of the paths of outcomes across groups from their common pre-treatment periods to their common post-treatment periods. Intuitively, the first set of comparisons is very much in the spirit of DiD, but, as we show below, the second and third sets of comparisons are not (except under additional specialized conditions).

**Lemma S6.** *Under Assumptions 1-MP, 2-MP(a), and 3-MP,*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\ddot{W}_{i,t}^2] = \sum_{g\in\mathcal{G}}\tilde{w}^{g,within}(g) + \sum_{g\in\mathcal{G}}\sum_{k\in\mathcal{G},k>g}\Big\{\tilde{w}^{g,post}(g,k) + \tilde{w}^{k,post}(g,k) + \tilde{w}^{long}(g,k)\Big\}$$

*Proof.* To start with, notice that $\mathbb{E}[\ddot{W}_{i,t}^2] = \mathbb{E}[W_{i,t}\ddot{W}_{i,t}]$. Then, we can apply the arguments of Lemmas S2 to S5 but with $W_{i,t}$ replacing $Y_{i,t}$. This implies that

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\ddot{W}_{i,t}^2] \\
&\quad= \sum_{g\in\mathcal{G}}\tilde{w}^{g,within}(g)\frac{\mathrm{Cov}(\bar{W}^{POST(g)} - \bar{W}^{PRE(g)}, D|G=g)}{\mathrm{Var}(D|G=g)} \\
&\quad+ \sum_{g\in\mathcal{G}}\sum_{k\in\mathcal{G},k>g}\Bigg\{\tilde{w}^{g,post}(g,k)\frac{\mathbb{E}\left[(\bar{W}^{MID(g,k)} - \bar{W}^{PRE(g)})|G=g\right] - \mathbb{E}\left[(\bar{W}^{MID(g,k)} - \bar{W}^{PRE(g)})|G=k\right]}{\mathbb{E}[D|G=g]} \\
&\qquad+ \tilde{w}^{k,post}(g,k)\frac{\mathbb{E}\left[(\bar{W}^{POST(k)} - \bar{W}^{MID(g,k)})|G=k\right] - \mathbb{E}\left[(\bar{W}^{POST(k)} - \bar{W}^{MID(g,k)})|G=g\right]}{\mathbb{E}[D|G=k]}
\end{aligned}$$

[8] One difference worth noting is that the weights are slightly different due to the terms involving $\mathbb{E}[D|G=g]$ and $\mathbb{E}[D|G=k]$. With a binary treatment, these expectations are equal to each other by construction, but with a continuous treatment these terms are no longer generally equal to each other. This also implies that the third term does not show up in the case with a binary treatment.

$$+ \tilde{w}^{long}(g,k)\frac{\mathbb{E}\left[(\bar{W}^{POST(k)} - \bar{W}^{PRE(g)})|G=g\right] - \mathbb{E}\left[(\bar{W}^{POST(k)} - \bar{W}^{PRE(g)})|G=k\right]}{\mathbb{E}[D|G=g] - \mathbb{E}[D|G=k]}\Bigg\}$$

$$= \sum_{g\in\mathcal{G}} \tilde{w}^{g,within}(g) + \sum_{g\in\mathcal{G}}\sum_{k\in\mathcal{G},k>g}\left\{\tilde{w}^{g,post}(g,k) + \tilde{w}^{k,post}(g,k) + \tilde{w}^{long}(g,k)\right\}$$

where the last equality holds by noting that $\bar{W} = D$ in post-treatment periods and $\bar{W} = 0$ in pre-treatment periods, and then by canceling terms. □

**Proof of Proposition S1**

*Proof.* Proposition S1 immediately holds by combining the results in Lemma S2, from Equations (S12) and (S13), and by Lemmas S3 to S5 (which all concern the numerator in the expression for $\beta^{twfe}$ in Equation (S2)), and then dividing by $(1/T)\sum_{t=1}^{T}\mathbb{E}[\ddot{W}_{i,t}^2]$ (which corresponds to the denominator in the expression for $\beta^{twfe}$ in Equation (S2)) using the result in Lemma S6. That the weights are all positive holds immediately by their definitions. That they sum to one holds by the definitions of the weights and is an immediate implication of Lemma S6. □

Next, we move to proving Theorem S3. To do this we provide expressions for each of the comparisons that show up in Proposition S1 in terms of derivatives of paths of outcomes. These results invoke Assumption 2-MP(b) and, therefore, use that the treatment is actually continuous, but they do not invoke any parallel trends assumptions. That said, it would be straightforward to adapt these results to the case with a discrete multi-valued treatment along the lines of the baseline two-period case considered in the main text.

It is also useful to note that

$$\frac{\partial \pi_D^{POST(\tilde{k}),PRE(\tilde{g})}(g,d)}{\partial d} = \frac{\partial \mathbb{E}\left[\left(\bar{Y}^{POST(\tilde{k})} - \bar{Y}^{PRE(\tilde{g})}\right)|G=g, D=d\right]}{\partial d},$$

$$\frac{\partial \pi_D^{MID(\tilde{g},\tilde{k}),PRE(\tilde{g})}(g,d)}{\partial d} = \frac{\mathbb{E}\left[\left(\bar{Y}^{MID(\tilde{g},\tilde{k})} - \bar{Y}^{PRE(\tilde{g})}\right)|G=g, D=d\right]}{\partial d},$$

$$\frac{\partial \pi_D^{POST(\tilde{k}),MID(\tilde{g},\tilde{k})}(g,d)}{\partial d} = \frac{\partial \mathbb{E}\left[\left(\bar{Y}^{POST(\tilde{k})} - \bar{Y}^{MID(\tilde{g},\tilde{k})}\right)|G=g, D=d\right]}{\partial d},$$

which holds because the second parts of each $\pi_D$ term do not vary with the dose.

Next, we consider a result for the numerator (which is the main term) of $\delta^{WITHIN}(g)$ in Equation (S3).

**Lemma S7.** *Under Assumptions 1-MP, 2-MP, and 3-MP,*

$$\mathrm{Cov}\left(\bar{Y}^{POST(g)} - \bar{Y}^{PRE(g)}, D|G=g\right)$$
$$= \int_{d_L}^{d_U}\Big(\mathbb{E}[D|G=g, D\geq l] - \mathbb{E}[D|G=g]\Big)\mathbb{P}(D\geq l|G=g)\frac{\partial \mathbb{E}[\bar{Y}^{POST(g)} - \bar{Y}^{PRE(g)}|G=g, D=l]}{\partial l}\, dl$$

*Proof.* First, notice that

$$\mathrm{Cov}\left(\bar{Y}^{POST(g)} - \bar{Y}^{PRE(g)}, D|G=g\right) = \mathbb{E}\left[\left(\bar{Y}^{POST(g)} - \bar{Y}^{PRE(g)}\right)(D - \mathbb{E}[D|G=g])|G=g\right]$$

Then, the proof follows essentially the same arguments as in Theorem 3.4(a) in the main text with $\bar{Y}^{POST(g)} - \bar{Y}^{PRE(g)}$ replacing $\Delta Y$ and the other arguments relating to the distribution of the dose holding conditional on being in group $g$. The second term, involving $d_L$, in Theorem 3.4(a) does not show up here as, by construction, there are no untreated units in group $g$. $\square$

Lemma S7 says that part of $\delta^{WITHIN}(g)$ in the TWFE regression estimator comes from a weighted average of $\frac{\partial \mathbb{E}[\bar{Y}^{POST(g)} - \bar{Y}^{PRE(g)}|G=g,D=d]}{\partial d}$.

Next, we consider the numerator (which is the main term) in the expression for $\delta^{MID,PRE}(g,k)$ in Equation (S4). This term is quite similar to the baseline two-period case considered in Theorem 3.4(a) because units in group $k$ have not been treated yet.

**Lemma S8.** *Under Assumptions 1-MP, 2-MP, and 3-MP, and for $k > g$,*

$$\begin{aligned}
&\mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G=g\right] - \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G=k\right] \\
&\quad = \int_{d_L}^{d_U} \mathbb{P}(D \geq l|G=g) \frac{\partial \mathbb{E}[\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}|G=g, D=l]}{\partial l}\, dl \\
&\qquad + d_L \frac{\mathbb{E}[\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}|G=g, D=d_L] - \mathbb{E}[\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}|D=0]}{d_L} \\
&\qquad - d_L \frac{\mathbb{E}[\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}|G=k] - \mathbb{E}[\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}|D=0]}{d_L}
\end{aligned}$$

*Proof.* To start with, notice that

$$\begin{aligned}
&\mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G=g\right] - \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G=k\right] \\
&\quad = \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G=g\right] - \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|D=0\right] \\
&\qquad - \left(\mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G=k\right] - \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|D=0\right]\right) \\
&\quad = \int_{d_L}^{d_U} \mathbb{P}(D \geq l|G=g) \frac{\partial \mathbb{E}[\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}|G=g, D=l]}{\partial l}\, dl \\
&\qquad + d_L \frac{\mathbb{E}[\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}|G=g, D=d_L] - \mathbb{E}[\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}|D=0]}{d_L} \\
&\qquad - d_L \frac{\mathbb{E}[\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}|G=k] - \mathbb{E}[\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}|D=0]}{d_L}
\end{aligned}$$

where the first equality holds by adding and subtracting $\mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|D=0\right]$. For the second equality, notice that

$$\begin{aligned}
&\mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G=g\right] - \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|D=0\right] \\
&\quad = \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G=g\right] - \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G=g, D=d_L\right] \\
&\qquad + \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G=g, D=d_L\right] - \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|D=0\right]
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G=g\right] - \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G=g, D=d_L\right] \\
&= \int_{d_L}^{d_U} \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G=g, D=d\right] - \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G=g, D=d_L\right]\, dF_{D|G}(d|g)
\end{aligned}$$

$$= \int_{d_L}^{d_U} \int_{d_L}^{d_U} \mathbf{1}\{l \leq d\} \frac{\partial \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G = g, D = l\right]}{\partial l} \, dl \, dF_{D|G}(d|g)$$
$$= \int_{d_L}^{d_U} \mathbb{P}(D \geq l | G = g) \frac{\partial \mathbb{E}[\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}|G = g, D = l]}{\partial l} \, dl$$

where the first equality holds by the law of iterated expectations, the second equality holds by the fundamental theorem of calculus, and the last equality holds by changing the order of integration and simplifying.

Combining the above expressions implies the result. □

Next, we consider the numerator (which is the main term) of $\delta^{POST,MID}(g,k)$ in Equation (S5) which comes from comparing paths of outcomes for newly treated groups relative to already-treated groups.

**Lemma S9.** *Under Assumptions 1-MP, 2-MP, and 3-MP, and for $k > g$,*

$$\begin{aligned}
&\mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}\right)|G = k\right] - \mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}\right)|G = g\right] \\
&\quad = \int_{d_L}^{d_U} \mathbb{P}(D \geq l | G = k) \frac{\partial \mathbb{E}[\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}|G = k, D = l]}{\partial l} \, dl \\
&\qquad + d_L \frac{\mathbb{E}[\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}|G = k, D = d_L] - \mathbb{E}[\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}|D = 0]}{d_L} \\
&\quad - \Big\{ \mathbb{E}[\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)}|G = g] - \mathbb{E}[\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)}|D = 0] \\
&\qquad - \Big( \mathbb{E}[\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}|G = g] - \mathbb{E}[\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}|D = 0] \Big) \Big\}
\end{aligned}$$

*Proof.* Notice that

$$\begin{aligned}
&\mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}\right)|G = k\right] - \mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}\right)|G = g\right] \\
&\quad = \left( \mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}\right)|G = k\right] - \mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}\right)|D = 0\right] \right) \\
&\qquad - \left( \mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}\right)|G = g\right] - \mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}\right)|D = 0\right] \right) \\
&\quad = \left( \mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}\right)|G = k\right] - \mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}\right)|D = 0\right] \right) \qquad \text{(S15)} \\
&\qquad - \Big\{ \left( \mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)}\right)|G = g\right] - \mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)}\right)|D = 0\right] \right) \\
&\qquad\quad - \left( \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|G = g\right] - \mathbb{E}\left[\left(\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}\right)|D = 0\right] \right) \Big\} \\
&\quad = \int_{d_L}^{d_U} \mathbb{P}(D \geq l | G = k) \frac{\partial \mathbb{E}[\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}|G = k, D = l]}{\partial l} \, dl \qquad \text{(S16)} \\
&\qquad + d_L \frac{\mathbb{E}[\bar{Y}^{POST(k)} - \bar{Y}^{MID}(g,k)|G = g, D = d_L] - \mathbb{E}\left[\left(\bar{Y}^{POST(k)} - \bar{Y}^{MID(g,k)}\right)|D = 0\right]}{d_L}
\end{aligned}$$

$$
-\left\{\left(\mathbb{E}\left[\left(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\right)|G=g\right]-\mathbb{E}\left[\left(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\right)|D=0\right]\right)\right.
$$

$$
\left.-\left(\mathbb{E}\left[\left(\bar{Y}^{MID(g,k)}-\bar{Y}^{PRE(g)}\right)|G=g\right]-\mathbb{E}\left[\left(\bar{Y}^{MID(g,k)}-\bar{Y}^{PRE(g)}\right)|D=0\right]\right)\right\}
$$

where the first equality holds by adding and subtracting $\mathbb{E}\left[\left(\bar{Y}^{POST(k)}-\bar{Y}^{MID(g,k)}\right)|D=0\right]$, the second equality holds by adding and subtracting both $\mathbb{E}\left[\bar{Y}^{PRE(g)}|G=g\right]$ and $\mathbb{E}\left[\bar{Y}^{PRE(g)}|D=0\right]$, and the last equality holds by applying the same sort of arguments as in the proof of Lemma S8. □

The expression in Lemma S9 appears complicated and is worth explaining in some more detail. Consider Equation (S15) in the proof of Lemma S9. There are three parts to this expression. The first part compares the path of outcomes in post-treatment periods relative to some pre-treatment periods for units in group $k$ to the path of outcomes for units that never participate in the treatment. This sort of comparison is very much in the spirit of DiD and will correspond to a reasonable treatment effect parameter under appropriate parallel trends assumptions. Similarly, under suitable parallel trends assumptions, the terms in the second and third lines will correspond to treatment effects for group $g$ between periods $k$ and $T$ (the second line) and treatment effects for group $g$ between periods $g$ and $k-1$ (the third line). Therefore, the difference between these terms can be thought of as some form of treatment effect dynamics. That means, in general, for this overall term to correspond to a treatment effect parameter for group $k$, there needs to be no treatment effect dynamics for group $g$—and, to be clear, treatment effect dynamics are not ruled out by any of the parallel trends assumptions that we have considered above.

Finally, we consider the numerator (which is the main term) of $\delta^{POST,PRE}(g,k)$ in Equation (S6).

**Lemma S10.** *Under Assumptions 1-MP, 2-MP, and 3-MP, and for $k > g$,*

$$
\begin{aligned}
&\mathbb{E}\left[\left(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\right)|G=g\right]-\mathbb{E}\left[\left(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\right)|G=k\right]\\
&=\int_{d_L}^{d_U}\left(\mathbb{P}(D\geq l|G=g)-\mathbb{P}(D\geq l|G=k)\right)\frac{\partial\mathbb{E}[\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}|G=g,D=l]}{\partial l}\,dl\\
&-\left\{\int_{d_L}^{d_U}\mathbb{P}(D\geq l|G=k)\left(\frac{\partial\mathbb{E}[\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}|G=k,D=l]}{\partial l}-\frac{\partial\mathbb{E}[\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}|G=g,D=l]}{\partial l}\right)dl\right.\\
&\qquad+d_L\frac{\mathbb{E}[\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}|G=k,D=d_L]-\mathbb{E}[\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}|D=0]}{d_L}\\
&\qquad\left.-d_L\frac{\mathbb{E}[\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}|G=g,D=d_L]-\mathbb{E}[\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}|D=0]}{d_L}\right\}.
\end{aligned}
$$

*Proof.* First, by adding and subtracting terms

$$
\begin{aligned}
&\mathbb{E}\left[\left(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\right)|G=g\right]-\mathbb{E}\left[\left(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\right)|G=k\right]\\
&\quad=\mathbb{E}\left[\left(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\right)|G=g\right]-\mathbb{E}\left[\left(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\right)|D=0\right]\\
&\qquad-\left(\mathbb{E}\left[\left(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\right)|G=k\right]-\mathbb{E}\left[\left(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\right)|D=0\right]\right).
\end{aligned}
$$

Then, using similar arguments as in Lemma S8 above, one can show that

$$
\mathbb{E}\left[\left(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\right)|G=g\right]-\mathbb{E}\left[\left(\bar{Y}^{POST(k)}-\bar{Y}^{PRE(g)}\right)|D=0\right]
$$

$$= \int_{d_L}^{d_U} \mathbb{P}(D \geq l | G = g) \frac{\partial \mathbb{E}[\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)} | G = g, D = l]}{\partial l} \, dl$$
$$+ d_L \frac{\mathbb{E}[\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)} | G = g, D = d_L] - \mathbb{E}[\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)} | D = 0]}{d_L}$$

and that

$$\mathbb{E}\left[ \left(\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)}\right) | G = k \right] - \mathbb{E}\left[ \left(\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)}\right) | D = 0 \right]$$
$$= \int_{d_L}^{d_U} \mathbb{P}(D \geq l | G = k) \frac{\partial \mathbb{E}[\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)} | G = k, D = l]}{\partial l} \, dl$$
$$+ d_L \frac{\mathbb{E}[\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)} | G = k, D = d_L] - \mathbb{E}[\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)} | D = 0]}{d_L}$$

Then, the result holds by adding and subtracting $\int_{d_L}^{d_U} \mathbb{P}(D \geq l | G = k) \frac{\partial \mathbb{E}[\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)} | G = g, D = l]}{\partial l} \, dl$ and combining terms. □

**Proof of Part (1) of Theorem S3**

*Proof.* Starting from the result in Proposition S1, the expression for $\delta^{WITHIN}(g)$ comes from its definition, the result in Lemma S7, and the definition of the weights $w_1^{within}(g, l)$. The expression for $\delta^{MID,PRE}(g, k)$ comes from its definition, the result in Lemma S8, and the definitions of $w_1(g, l)$ and $w_0(g)$. The expression for $\delta^{POST,MID}(g, k)$ comes from combining its definition with the result in Lemma S9, and the definitions of $w_1(k, l)$ and $w_0(k)$. Finally, the expression for $\delta^{POST,PRE}(g, k)$ comes from its definition, the result in Lemma S10, and the definitions of $w_1^{across}(g, k, l)$, $\tilde{w}_1^{across}(g, k, l)$, and $\tilde{w}_0^{across}(g, k)$.

That $w_1^{within}(g, d) \geq 0$, $w_1(g, d) \geq 0$, $w_0(g) \geq 0$ for all $g \in \mathcal{G}$ and $d \in \mathcal{D}_+^c$ all hold immediately from the definitions of the weights. That $\int_{d_L}^{d_U} w_1^{within}(g, l) \, dl = 1$, $\int_{d_L}^{d_U} w_1(g, l) \, dl + w_0(g) = 1$, and $\int_{d_L}^{d_U} w_1^{across}(g, k, l) \, dl = 1$ hold from the same sorts of arguments used to show that the weights integrate to 1 in the proof of Theorem 3.4(a). □

Notice that none of the previous results have invoked any sort of parallel trends assumption. Next, we push forward the previous results once a researcher invokes parallel trends assumptions; in Theorem S3, we consider the case where the researcher invoked Assumption 5-MP-Extended(a), but here we handle both that assumption and Assumption 4-MP-Extended(a) (as in Theorem S3-Extended). To further understand this, for $1 \leq t_1 < t_2 \leq T$ define

$$\bar{Y}_i^{(t_1,t_2)}(g, d) = \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} Y_{i,t}(g, t, d)$$

which averages potential outcomes from time periods $t_1$ to $t_2$ for unit $i$ if they were in group $g$ and experienced dose $d$. Note that $\bar{Y}_i^{(t_1,t_2)} = \bar{Y}_i^{(t_1,t_2)}(G_i, D_i)$. Next, for $t_1 \leq t_2$, define

$$\overline{ATT}^{(t_1,t_2)}(g, d | g, d) = \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} ATT(g, t, d | g, d)$$

which is the average treatment effect experienced by units in group $g$ who experienced dose $d$ averaged across periods from $t_1$ to $t_2$. Likewise, define

$$\overline{ATE}^{(t_1,t_2)}(g,d) = \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} ATE(g,t,d)$$

which is the average treatment effect of dose $d$ among all units in group $g$ averaged across periods from $t_1$ to $t_2$. An alternative expression for $\overline{ATT}^{(t_1,t_2)}(g,d|g,d)$ is given by

$$\overline{ATT}^{(t_1,t_2)}(g,d|g,d) = \mathbb{E}\left[\bar{Y}^{(t_1,t_2)}(g,d) - \bar{Y}^{(t_1,t_2)}(0)|G=g, D=d\right]$$

which holds by the definition of $ATT(g,t,d|g,d)$ and changing the order of the expectation and the average over time periods; here, $\mathbb{E}[\bar{Y}^{(t_1,t_2)}(0)|G=g, D=d]$ is the average outcome that units in group $g$ that experienced dose $d$ would have experienced if they had not participated in the treatment between time periods $t_1$ and $t_2$. Similarly, for $\overline{ATE}^{(t_1,t_2)}(g,d)$,

$$\overline{ATE}^{(t_1,t_2)}(g,d) = \mathbb{E}\left[\bar{Y}^{(t_1,t_2)}(g,d) - \bar{Y}^{(t_1,t_2)}(0)|G=g\right]$$

In addition, define

$$\overline{ACRT}^{(t_1,t_2)}(g,d|g,d) = \frac{\partial \overline{ATT}(g,l|g,d)}{\partial l}\Big|_{l=d} \quad \text{and} \quad \overline{ACR}^{(t_1,t_2)}(g,d) = \frac{\partial \overline{ATE}(g,d)}{\partial d}$$

which are the average causal response to a marginal increase in the dose among units in group $g$ conditional on having dose experienced dose $d$ (for $\overline{ACRT}(g,d|g,d)$) and the average causal response to a marginal increase in the dose among all units in group $g$.

The next result connects derivatives of conditional expectations to $ACRT$ and $ACR$ parameters under parallel trends assumptions. This is similar to Theorems 3.2 and 3.3 in the main text and to Theorem S2 above.

**Lemma S11.** *Under Assumptions 1-MP, 2-MP, and 3-MP, and for* $1 \leq t_1 \leq t_2 < g \leq t_3 \leq t_4 \leq T$ *(i.e.,* $t_1$ *and* $t_2$ *are pre-treatment periods for group* $g$*, and* $t_3$ *and* $t_4$ *are post-treatment periods for group* $g$*), and for* $d \in \mathcal{D}_+^c$*,*

*(1) If, in addition, Assumption 4-MP-Extended(a) holds, then*

$$\frac{\partial \mathbb{E}\left[\bar{Y}^{(t_3,t_4)} - \bar{Y}^{(t_1,t_2)}|G=g, D=d\right]}{\partial d} = \overline{ACRT}^{(t_3,t_4)}(g,d|g,d) + \underbrace{\frac{\partial \overline{ATT}^{(t_3,t_4)}(g,d|g,l)}{\partial l}\Big|_{l=d}}_{\text{selection bias}}$$

*(2) If, in addition, Assumption 5-MP-Extended(a) holds, then*

$$\frac{\partial \mathbb{E}\left[\bar{Y}^{(t_3,t_4)} - \bar{Y}^{(t_1,t_2)}|G=g, D=d\right]}{\partial d} = \overline{ACR}^{(t_3,t_4)}(g,d)$$

*Proof.* For part (1), notice that, for $1 \leq t_1 \leq t_2 < g \leq t_3 \leq t_4 \leq T$ (i.e., for group $g$, $t_1$ and $t_2$ are pre-treatment time periods while $t_3$ and $t_4$ are post-treatment time periods), we can write

$$\begin{aligned}\mathbb{E}\left[\bar{Y}^{(t_3,t_4)} - \bar{Y}^{(t_1,t_2)}|G=g, D=d\right] &= \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(g,d) - \bar{Y}^{(t_1,t_2)}(0)|G=g, D=d\right] \\ &= \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(g,d) - \bar{Y}^{(t_3,t_4)}(0)|G=g, D=d\right]\end{aligned}$$

$$
\begin{aligned}
&\quad - \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(0) - \bar{Y}^{(t_1,t_2)}(0)|G=g, D=d\right] \\
&= \overline{ATT}^{(t_3,t_4)}(g,d|g,d) \\
&\quad - \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(0) - \bar{Y}^{(t_1,t_2)}(0)|G=g, D=d\right]
\end{aligned}
$$

where the first equality holds by writing observed outcomes in terms of their corresponding potential outcomes, the second equality holds by adding and subtracting $\mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(0)|G=g, D=d\right]$, and the last equality holds by the definition of $\overline{ATT}^{(t_3,t_4)}(g,d|g,d)$.

This equation looks very similar to DiD-type equations in simpler cases such as when there are two periods and two groups. The left-hand side is immediately identified. The right-hand side involves a causal effect parameter of interest and an unobserved path of untreated potential outcomes that would typically be handled using a parallel trends assumption.

In particular, under Assumption 4-MP-Extended(a) (though notice that Assumption 4-MP-Extended(b) and (c) are not generally strong enough here),

$$
\mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(0) - \bar{Y}^{(t_1,t_2)}(0)|G=g, D=d\right] = \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(0) - \bar{Y}^{(t_1,t_2)}(0)|D=0\right]
$$

which, importantly, does not vary across $d$ or $g$.

This suggests that, under Assumption 4-MP-Extended(a),

$$
\mathbb{E}\left[\bar{Y}^{(t_3,t_4)} - \bar{Y}^{(t_1,t_2)}|G=g, D=d\right] = \overline{ATT}^{(t_3,t_4)}(g,d|g,d) - \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(0) - \bar{Y}^{(t_1,t_2)}(0)|D=0\right]
$$

Taking derivatives of both sides of the previous equation with respect to $d$ implies the result.

For part (2), notice that,

$$
\begin{aligned}
\mathbb{E}\left[\bar{Y}^{(t_3,t_4)} - \bar{Y}^{(t_1,t_2)}|G=g, D=d\right] &= \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(g,d) - \bar{Y}^{(t_1,t_2)}(0)|G=g, D=d\right] \\
&= \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(g,d) - \bar{Y}^{(t_1,t_2)}(0)|G=g\right] \\
&= \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(g,d) - \bar{Y}^{(t_3,t_4)}(0)|G=g\right] \\
&\quad + \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(0) - \bar{Y}^{(t_1,t_2)}(0)|G=g\right] \\
&= \overline{ATE}^{(t_3,t_4)}(g,d) + \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(0) - \bar{Y}^{(t_1,t_2)}(0)|D=0\right]
\end{aligned}
$$

where the first equality holds by writing observed outcomes in terms of their corresponding potential outcomes, the second equality holds by Assumption 5-MP-Extended(a), the third equality holds by adding and subtracting $\mathbb{E}[\bar{Y}^{(t_3,t_4)}(0)|G=g]$, and the last equality holds by the definition of $\overline{ATE}^{(t_3,t_4)}(g,d)$ and by Assumption 5-MP-Extended(a). Taking derivatives of both sides implies the result for part (2). □

The result in Lemma S11 says that, under Assumption 4-MP-Extended(a), the derivative of the path of outcomes (averaged over some post-treatment periods) relative to some pre-treatment periods corresponds to averaging $ACRT(g,t,d|g,d)$ across post-treatment time periods plus the derivative of an averaged selection bias-type across some post-treatment time periods for group $g$. Similarly, under Assumption 5-MP-Extended(a), the derivative of the path of average outcomes in some post-treatment periods relative to average outcomes in some pre-treatment periods corresponds to an average of $ACR(g,d)$ across the same post-treatment time periods.

**Lemma S12.** *Under Assumptions 1-MP, 2-MP, and 3-MP, and for* $1 \leq t_1 \leq t_2 < g \leq t_3 \leq t_4 < k$ *(i.e.,* $t_1$ *and* $t_2$ *are pre-treatment periods for both groups* $g$ *and* $k$*, group* $g$ *is treated before group* $k$*, and* $t_3$ *and* $t_4$ *are post-treatment periods for group* $g$ *but pre-treatment periods for group* $k$*),*

*(1) If, in addition, Assumption 4-MP-Extended(a) holds, then*

$$d_L \frac{\mathbb{E}\left[\bar{Y}^{(t_3,t_4)} - \bar{Y}^{(t_1,t_2)}|G=g, D=d_L\right] - \mathbb{E}\left[\bar{Y}^{(t_3,t_4)} - \bar{Y}^{(t_1,t_2)}|G=k\right]}{d_L} = d_L \frac{\overline{ATT}^{(t_3,t_4)}(g, d_L|g, d_L)}{d_L}$$

*(2) If, in addition, Assumption 5-MP-Extended(a) holds, then*

$$d_L \frac{\mathbb{E}\left[\bar{Y}^{(t_3,t_4)} - \bar{Y}^{(t_1,t_2)}|G=g, D=d_L\right] - \mathbb{E}\left[\bar{Y}^{(t_3,t_4)} - \bar{Y}^{(t_1,t_2)}|G=k\right]}{d_L} = d_L \frac{\overline{ATE}^{(t_3,t_4)}(g, d_L)}{d_L}$$

*Proof.* For part (1), notice that

$$\begin{aligned}
&\mathbb{E}\left[\bar{Y}^{(t_3,t_4)} - \bar{Y}^{(t_1,t_2)}|G=g, D=d_L\right] - \mathbb{E}\left[\bar{Y}^{(t_3,t_4)} - \bar{Y}^{(t_1,t_2)}|G=k\right] \\
&\quad = \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(g,d_L) - \bar{Y}^{(t_1,t_2)}(0)|G=g, D=d_L\right] - \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(0) - \bar{Y}^{(t_1,t_2)}(0)|G=k\right] \\
&\quad = \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(g,d_L) - \bar{Y}^{(t_3,t_4)}(0)|G=g, D=d_L\right] \\
&\qquad\quad + \left\{\mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(0) - \bar{Y}^{(t_1,t_2)}(0)|G=g, D=d_L\right] - \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(0) - \bar{Y}^{(t_1,t_2)}(0)|G=k\right]\right\} \\
&\quad = \overline{ATT}^{(t_3,t_4)}(g,d_L)
\end{aligned}$$

where the first equality holds by writing observed outcomes in terms of their corresponding potential outcomes, the second equality holds by adding and subtracting $\mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(0)|G=g, D=d_L\right]$, and the last equality holds by the definition of $\overline{ATT}^{(t_3,t_4)}(g,d_L)$ and because the difference between the two terms involving paths of untreated potential outcomes on the second line of the previous equality is equal to 0 under Assumption 4-MP-Extended(a). Then, the result holds by multiplying and dividing by $d_L$.

For part (2),

$$\begin{aligned}
&\mathbb{E}\left[\bar{Y}^{(t_3,t_4)} - \bar{Y}^{(t_1,t_2)}|G=g, D=d_L\right] - \mathbb{E}\left[\bar{Y}^{(t_3,t_4)} - \bar{Y}^{(t_1,t_2)}|G=k\right] \\
&\quad = \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(g,d_L) - \bar{Y}^{(t_1,t_2)}(0)|G=g, D=d_L\right] - \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(0) - \bar{Y}^{(t_1,t_2)}(0)|G=k\right] \\
&\quad = \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(g,d_L) - \bar{Y}^{(t_1,t_2)}(0)|G=g\right] - \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(0) - \bar{Y}^{(t_1,t_2)}(0)|G=k\right] \\
&\quad = \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(g,d_L) - \bar{Y}^{(t_3,t_4)}(0)|G=g\right] \\
&\qquad\quad + \left\{\mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(0) - \bar{Y}^{(t_1,t_2)}(0)|G=g\right] - \mathbb{E}\left[\bar{Y}^{(t_3,t_4)}(0) - \bar{Y}^{(t_1,t_2)}(0)|G=k\right]\right\} \\
&\quad = \overline{ATE}^{(t_3,t_4)}(g,d_L)
\end{aligned}$$

where the first equality holds by writing observed outcomes in terms of their corresponding potential outcomes, the second equality holds by Assumption 5-MP-Extended(a), the third equality holds by adding and subtracting $\mathbb{E}[\bar{Y}^{(t_3,t_4)}(0)|G=g]$, and the last equality holds by Assumption 5-MP-Extended(a). The result holds by multiplying and dividing by $d_L$. ☐

### Proof of Part (2) of Theorem S3

*Proof.* The result holds immediately by using the results of Lemmas S11 and S12 in each of the expressions for $\delta^{WITHIN}(g)$, $\delta^{MID,PRE}(g,k)$, $\delta^{POST,MID}(g,k)$, and $\delta^{POST,PRE}(g,k)$ in part (1) of Theorem S3. $\square$

### Proof of Proposition S2

*Proof.* For part (a), we consider the nuisance term involving $\pi^{POST(k),PRE(g)}(g)-\pi^{MID(g,k),PRE(g)}(g)$ in the expression for $\delta^{POST,MID}(g,k)$ in part (2) of Theorem S3. Then, using similar arguments as in Lemma S8 and then under Assumption 5-MP-Extended(a), it follows that

$$\begin{aligned}\pi^{POST(k),PRE(g)}(g) &= \mathbb{E}\left[\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)}|G=g\right] - \mathbb{E}\left[\bar{Y}^{POST(k)} - \bar{Y}^{PRE(g)}|D=0\right] \\ &= \int_{d_L}^{d_U} \mathbb{P}(D \geq l|G=g)\overline{ACR}^{POST(k)}(g,l)\,dl + d_L\frac{\overline{ATE}^{POST(k)}(g,d_L)}{d_L}\end{aligned}$$

and that

$$\begin{aligned}\pi^{MID(g,k),PRE(g)}(g) &= \mathbb{E}\left[\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}|G=g\right] - \mathbb{E}\left[\bar{Y}^{MID(g,k)} - \bar{Y}^{PRE(g)}|D=0\right] \\ &= \int_{d_L}^{d_U} \mathbb{P}(D \geq l|G=g)\overline{ACR}^{MID(g,k)}(g,l)\,dl + d_L\frac{\overline{ATE}^{MID(g,k)}(g,d_L)}{d_L}\end{aligned}$$

Under Assumption S1(a), $ACR(g,t,d)$ and $ATE(g,t,d_L)$ do not vary over time which implies that, for all $g \in \mathcal{G}$ and $k \in \mathcal{G}$ with $k > g$, $\overline{ACR}^{POST(k)}(g,l) = \overline{ACR}^{MID(g,k)}(g,l)$ for all $l \in \mathcal{D}_+^c$ and $\overline{ATE}^{POST(k)}(g,d_L) = \overline{ATE}^{MID(g,k)}(g,d_L)$. This implies that $\pi^{POST(k),PRE(g)}(g) = \pi^{MID(g,k),PRE(g)}(g)$ which implies the result for part (a).

For part (b), we consider the two nuisance terms in the expression for $\delta^{POST,PRE}(g,k)$ in part (2) of Theorem S3. For the first one, notice that, under Assumption 5-MP-Extended(a),

$$\begin{aligned}\frac{\partial \pi_D^{POST(k),PRE(g)}(k,l)}{\partial l} - \frac{\partial \pi_D^{POST(k),PRE(g)}(g,l)}{\partial l} &= \overline{ACR}^{POST(k)}(k,l) - \overline{ACR}^{POST(k)}(g,l) \\ &= 0\end{aligned}$$

for $l \in \mathcal{D}_+^c$ and where the second equality holds by Assumption S1(b) (which implies that, for a particular time period, $ACR(g,t,d)$ does not vary across groups).

For the second nuisance term, the same sort of arguments imply that

$$\begin{aligned}\frac{\pi_D^{POST(k),PRE(g)}(k,d_L) - \pi_D^{POST(k),PRE(g)}(g,d_L)}{d_L} &= \frac{\overline{ATE}^{POST(k)}(k,d_L) - \overline{ATE}^{POST(k)}(g,d_L)}{d_L} \\ &= 0\end{aligned}$$

under Assumption S1(b).

Finally, for part (c), under Assumption S1(a), (b), and (c), $ACR(g,t,d)$ does not vary across groups, time periods, or dose; since this does not vary, we denote it by $ACR$ for the remainder of the proof. Moreover, from Theorem S3, we have that $\int_{d_L}^{d_U} w_1^{within}(g,l)\,dl = 1$, $\int_{d_L}^{d_U} w_1(g,l)\,dl + w_0(g) = 1$, and that $\int_{d_L}^{d_U} w_1^{across}(g,k,l) = 1$. From the first two parts of the current result, we also have that the nuisance paths of outcomes in $\delta^{POST,MID}(g,k)$ and $\delta^{POST,PRE}(g,k)$ are both equal to 0 under Assumption S1(a) and (b). This implies that, under the conditions for part (c), $\delta^{WITHIN}(g) = \delta^{MID,PRE}(g,k) = \delta^{POST,MID}(g,k) = \delta^{POST,PRE}(g,k) = ACR$. Finally, from Proposition S1, we have that $\beta^{twfe}$ is a weighted average of $\delta^{MID,PRE}(g,k)$, $\delta^{POST,MID}(g,k)$, $\delta^{POST,MID}(g,k)$, and $\delta^{POST,PRE}(g,k)$. That these are all equal to each other implies that $\beta^{twfe} = ACR = ACR^o$. □

**Proof of Theorem S3-Extended**

*Proof.* The result holds immediately by plugging in the results of part (1) of Lemmas S11 and S12 for $\delta^{WITHIN}(g)$, $\delta^{MID,PRE}(g,k)$, $\delta^{POST,MID}(g,k)$, and $\delta^{POST,PRE}(g,k)$ in part (1) of Theorem S3. □

# SC Additional Theoretical Results

This appendix provides (and proves) a number of additional results that were referred to in the main text.

## SC.1 No Untreated Units

This section considers the causal interpretation of comparisons of paths of outcomes across dose groups in settings with no untreated units under different versions of the parallel trends assumption.

**Proposition S3.** *Under Assumptions 1, 2, 3, and 4,*[9] *and for* $(h,l) \in \mathcal{D}_+ \times \mathcal{D}_+$*,*

$$\mathbb{E}[\Delta Y|D=h] - \mathbb{E}[\Delta Y|D=l] = ATT(h|h) - ATT(l|l)$$

*Proof.* Notice that

$$\begin{aligned}
\mathbb{E}[\Delta Y|D=h] - \mathbb{E}[\Delta Y|D=l] &= \mathbb{E}[Y_{t=2}(h) - Y_{t=1}(0)|D=h] - \mathbb{E}[Y_{t=2}(l) - Y_{t=1}(0)|D=l] \\
&= \mathbb{E}[Y_{t=2}(h) - Y_{t=2}(0)|D=h] - \mathbb{E}[Y_{t=2}(l) - Y_{t=2}(0)|D=l] \\
&\quad + \Big(\mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D=h] - \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D=l]\Big) \\
&= ATT(h|h) - ATT(l|l)
\end{aligned}$$

where the first equality holds by plugging in potential outcomes for observed outcomes, the second equality holds by adding and subtracting $\mathbb{E}[Y_{t=2}(0)|D=h]$ and $\mathbb{E}[Y_{t=2}(0)|D=l]$, and the last equality holds by the definition of $ATT(d|d)$ and by Assumption 4. □

The result in Proposition S3 is the same as in Theorem 3.2(b) in the main text though the proof technique is different here as there does not exist an untreated comparison group in the setting considered here.

[9] To be fully precise, Assumption 2 needs to be modified here to allow for no untreated units. Likewise, the parallel trends assumption in Assumption 4 does not immediately apply to this setting because $\mathbb{P}(D=0)=0$ here. Instead, by parallel trends, we mean that $\mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D=d] = \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)]$ which says that the path of untreated potential outcomes is the same across all dose groups. We do not state this as a separate assumption partly for brevity but also because, in a setting where $\mathbb{P}(D=0)>0$, the condition here is simply an alternative way to write Assumption 4.

Next, we provide an analogous result under strong parallel trends.

**Proposition S4.** *Under Assumptions 1, 2, 3, and 5, and for* $(h, l) \in \mathcal{D}_+ \times \mathcal{D}_+$*,*

$$\mathbb{E}[\Delta Y|D=h] - \mathbb{E}[\Delta Y|D=l] = ATE(h) - ATE(l)$$

*Proof.* Notice that

$$\begin{aligned}
\mathbb{E}[\Delta Y|D=h] - \mathbb{E}[\Delta Y|D=l] &= \mathbb{E}[Y_{t=2}(h) - Y_{t=1}(0)|D=h] - \mathbb{E}[Y_{t=2}(l) - Y_{t=1}(0)|D=l] \\
&= \mathbb{E}[Y_{t=2}(h) - Y_{t=1}(0)] - \mathbb{E}[Y_{t=2}(l) - Y_{t=1}(0)] \\
&= \mathbb{E}[Y_{t=2}(h) - Y_{t=2}(0)] - \mathbb{E}[Y_{t=2}(l) - Y_{t=2}(0)] \\
&= ATE(h) - ATE(l)
\end{aligned}$$

where the first equality holds by replacing observed outcomes with corresponding potential outcomes, the second equality holds by Assumption 5, the third equality holds by canceling the $\mathbb{E}[Y_{t=1}(0)]$ terms from the previous line and by adding and subtracting $\mathbb{E}[Y_{t=2}(0)]$, and the last equality holds by the definition of $ATE(d)$. □

## SC.2 Additional TWFE Decomposition Results

This section provides some extensions and additional details related to the TWFE decompositions discussed in Section 3.3 in the main text.

### Additional Results for TWFE Levels Decomposition

This first part of this section derives the expression for $\beta^{twfe}$ in Equation (3.1) in the main text which relates $\beta^{twfe}$ to a weighted average of "more treated" units (units that experienced a dose larger than $\mathbb{E}[D]$) relative to "less treated" units (units that were untreated or experienced a dose smaller than $\mathbb{E}[D]$) scaled by a weighted average of the difference in treatment experienced by these two groups. Recalling that Theorem 3.4(b) in the main text showed that the "weights" integrated to 0, the second part of this section integrates separately the positive and negative parts of those weights (which are separated on the basis of whether or not $d$ is greater than the mean dose $\mathbb{E}[D]$). The takeaway is that the positive weights do not integrate to 1 (nor do the negative weights integrate to $-1$), but rather they integrate to the reciprocal of the weighted distance between the effective treated and effective comparison group discussed in the main text. This provides an explicit connection between the levels decomposition in Theorem 3.4 and the alternative expression for $\beta^{twfe}$ provided in Equation (3.1) in the main text.

**Corollary S1.** *Under Assumptions 1, 2(a), and 3,*

$$\beta^{twfe} = \frac{\mathbb{E}\Big[w_1^{bin}(D)\Delta Y\Big|D > \mathbb{E}[D]\Big] - \mathbb{E}\Big[w_0^{bin}(D)\Delta Y\Big|D \le \mathbb{E}[D]\Big]}{\mathbb{E}\Big[w_1^{bin}(D)D\Big|D > \mathbb{E}[D]\Big] - \mathbb{E}\Big[w_0^{bin}(D)\Delta D\Big|D \le \mathbb{E}[D]\Big]}. \tag{S17}$$

*If, in addition, Assumption 4 also holds, then*

$$\beta^{twfe} = \frac{\mathbb{E}\Big[w_1^{bin}(D)ATT(D|D)\Big|D > \mathbb{E}[D]\Big] - \mathbb{E}\Big[w_0^{bin}(D)ATT(D|D)\Big|D \leq \mathbb{E}[D]\Big]}{\mathbb{E}\Big[w_1^{bin}(D)D\Big|D > \mathbb{E}[D]\Big] - \mathbb{E}\Big[w_0^{bin}(D)\Delta D\Big|D \leq \mathbb{E}[D]\Big]}. \tag{S18}$$

*where*

$$w_1^{bin}(d) := \frac{\big|d - \mathbb{E}[D]\big|}{\mathbb{E}\Big[\big|D - \mathbb{E}[D]\big|\Big|D > \mathbb{E}[D]\Big]}$$

$$w_0^{bin}(d) := \frac{\big|d - \mathbb{E}[D]\big|}{\mathbb{E}\Big[\big|D - \mathbb{E}[D]\big|\Big|D \leq \mathbb{E}[D]\Big]}$$

*which satisfy* $\mathbb{E}\big[w_1^{bin}(D)\big|D > \mathbb{E}[D]\big] = \mathbb{E}\big[w_0^{bin}\big|D \leq \mathbb{E}[D]\big] = 1$.

*Proof.* To start with, recall that

$$\beta^{twfe} = \frac{\mathbb{E}[(D - \mathbb{E}[D])\Delta Y]}{\mathrm{Var}(D)} =: \frac{\beta_{num}}{\beta_{den}}$$

where we consider the numerator and denominator separately below. Next, notice that

$$\begin{aligned} 0 &= \mathbb{E}[(D - \mathbb{E}[D])] \\ &= \mathbb{E}\Big[(D - \mathbb{E}[D])\Big|D \leq \mathbb{E}[D]\Big]\mathbb{P}(D \leq \mathbb{E}[D]) + \mathbb{E}\Big[(D - \mathbb{E}[D])\Big|D > \mathbb{E}[D]\Big]\mathbb{P}(D > \mathbb{E}[D]) \end{aligned}$$

where the second equality holds by the law of iterated expectation. Rearranging the previous expression we have that

$$\mathbb{E}\Big[|D - \mathbb{E}[D]|\Big|D \leq \mathbb{E}[D]\Big]\mathbb{P}(D \leq \mathbb{E}[D]) = \mathbb{E}\Big[|D - \mathbb{E}[D]|\Big|D > \mathbb{E}[D]\Big]\mathbb{P}(D > \mathbb{E}[D]) =: \delta$$

where the equality uses that the sign of $(D - \mathbb{E}[D])$ is fully determined in both conditional expectations.

Next, similar to above, split the numerator of $\beta^{twfe}$ on the basis of whether or not $D > \mathbb{E}[D]$:

$$\beta_{num} = \mathbb{E}\Big[(D - \mathbb{E}[D])\Delta Y\Big|D > \mathbb{E}[D]\Big]\mathbb{P}(D > \mathbb{E}[D]) + \mathbb{E}\Big[(D - \mathbb{E}[D])\Delta Y\Big|D \leq \mathbb{E}[D]\Big]\mathbb{P}(D \leq \mathbb{E}[D])$$

and, now consider,

$$\begin{aligned} \frac{\beta_{num}}{\delta} &= \mathbb{E}\left[\frac{|D - \mathbb{E}[D]|}{\mathbb{E}\big[|D - \mathbb{E}[D]|\big|D > \mathbb{E}[D]\big]}\Delta Y\Big|D > \mathbb{E}[D]\right] - \mathbb{E}\left[\frac{|D - \mathbb{E}[D]|}{\mathbb{E}\big[|D - \mathbb{E}[D]|\big|D \leq \mathbb{E}[D]\big]}\Delta Y\Big|D \leq \mathbb{E}[D]\right] \\ &= \mathbb{E}\Big[w_1^{bin}(D)\Delta Y\Big|D > \mathbb{E}[D]\Big] - \mathbb{E}\Big[w_0^{bin}(D)\Delta Y\Big|D \leq \mathbb{E}[D]\Big] \end{aligned} \tag{S19}$$

which uses the two different expressions for $\delta$ given above. Also, notice that it also immediately follows that $\mathbb{E}[w_1^{bin}(D)|D > \mathbb{E}[D]] = \mathbb{E}[w_0^{bin}|D \leq \mathbb{E}[D]] = 1$. Thus, $\beta_{num}/\delta$ can be thought of as a weighted average of the change in outcomes for units with $D > \mathbb{E}[D]$ relative to a weighted average of the change in outcomes for units with $D \leq \mathbb{E}[D]$, where the weights are larger for units with values of $D$ further away from $\mathbb{E}[D]$.

Similarly, since $\mathrm{Var}(D) = \mathbb{E}[(D - \mathbb{E}[D])D]$ we can apply the same argument to the denominator, and show that

$$\frac{\beta_{den}}{\delta} = \mathbb{E}\Big[w_1^{bin}(D)D\Big|D > \mathbb{E}[D]\Big] - \mathbb{E}\Big[w_0^{bin}(D)D\Big|D \leq \mathbb{E}[D]\Big] \tag{S20}$$

This can be thought of as a weighted average of $D$ for units with $D > \mathbb{E}[D]$ relative to units with $D \leq \mathbb{E}[D]$, or, in other words, the distance between the mean of $D$ for the "effective" treated group relative to the "effective" comparison group given the weighting scheme discussed above. Taking the ratio of Equations S19 and S20 completes the proof for the expression in Equation (S17). That the weights are positive and have mean one follows immediately from their definitions. The result in Equation (S18) holds because

$$\begin{aligned}\mathbb{E}\Big[w_1^{bin}(D)\Delta Y\Big|D > \mathbb{E}[D]\Big] &= \mathbb{E}\Big[w_1^{bin}(D)\mathbb{E}[\Delta Y|D]\Big|D > \mathbb{E}[D]\Big] \\ &= \mathbb{E}\Big[w_1^{bin}(D)\Big(\mathbb{E}[\Delta Y|D] - \mathbb{E}[\Delta Y|D=0]\Big)\Big|D > \mathbb{E}[D]\Big] + \mathbb{E}[\Delta Y|D=0] \\ &= \mathbb{E}\Big[w_1^{bin}(D)ATT(D|D)\Big|D > \mathbb{E}[D]\Big] + \mathbb{E}[\Delta Y|D=0] \qquad \text{(S21)}\end{aligned}$$

where the first equality holds by the law of iterated expectations, the second equality holds by adding and subtracting $\mathbb{E}[\Delta Y|D=0]$ and because $\mathbb{E}[\Delta Y|D=0]$ is non-random and $\mathbb{E}\Big[w_1^{bin}(D)\Big|D > \mathbb{E}[D]\Big]$ has mean one, and the last equality holds under Assumption 4. The same sort of argument can be used to show that

$$\mathbb{E}\Big[w_0^{bin}(D)\Delta Y\Big|D \leq \mathbb{E}[D]\Big] = \mathbb{E}\Big[w_0^{bin}(D)ATT(D|D)\Big|D \leq \mathbb{E}[D]\Big] + \mathbb{E}[\Delta Y|D=0] \qquad \text{(S22)}$$

where, by construction, $ATT(0|0) = 0$. Taking the difference between the expressions in Equations (S21) and (S22) and then combining these expressions with the above results for Equation (S17) completes the proof for the expression in Equation (S18).[10] □

**Corollary S2.** *Under Assumptions 1, 2(a), 3, and 4,*

$$w_0^{lev} + \int_{d_L}^{\mathbb{E}[D]} w_1^{lev}\, dl = \int_{\mathbb{E}[D]}^{d_U} w_1^{lev}(l)\, dl = \frac{1}{\mathbb{E}\Big[w_1^{bin}(D)D\Big|D > \mathbb{E}[D]\Big] - \mathbb{E}\Big[w_0^{bin}(D)D\Big|D \leq \mathbb{E}[D]\Big]}$$

*where $w_1^{bin}$ and $w_0^{bin}$ are defined in Corollary S1.*

*Proof.* We showed that $w_0^{lev} + \int_{d_L}^{\mathbb{E}[D]} w_1^{lev}(l)\, dl = \int_{\mathbb{E}[D]}^{d_U} w_1^{lev}(l)\, dl$ in Theorem 3.4(b). Therefore, consider

$$\begin{aligned}\int_{\mathbb{E}[D]}^{d_U} w_1^{lev}(l)\, dl &= \int_{\mathbb{E}[D]}^{d_U} \frac{(l - \mathbb{E}[D])}{\mathrm{Var}(D)} f_D(l)\, dl \\ &= \frac{\mathbb{E}\Big[|D - \mathbb{E}[D]|\Big|D > \mathbb{E}[D]\Big]\mathbb{P}(D > \mathbb{E}[D])}{\mathrm{Var}(D)} \\ &= \frac{\delta}{\beta_{den}} \\ &= \frac{1}{\mathbb{E}\Big[w_1^{bin}(D)D\Big|D > \mathbb{E}[D]\Big] - \mathbb{E}\Big[w_0^{bin}(D)D\Big|D \leq \mathbb{E}[D]\Big]}\end{aligned}$$

where the first equality holds by the definition of $w_1^{lev}(l)$, the second equality holds by the law of iterated expectations and because $(D - \mathbb{E}[D])$ is positive conditional on $D > \mathbb{E}[D]$, the third equality holds from the expressions for $\delta$ and $\beta_{den}$ in the proof of Corollary S1, and the last equality holds by

[10] Notice that if we were to invoke Assumption 5, a result analogous to the one in Equation (S18) holds with $ATE(D)$ replacing $ATT(D|D)$.

Equation (S20) above. This completes the proof. □

**Scaled-Levels Decomposition for Fixed Dose**

Next, we consider interpreting $\beta^{twfe}$ as $ATT(d)/d$ for some particular fixed value of $d$. This is similar to the scaled-level effects discussed in Section 3.3 in the main text except for that we fix $d$ instead of relating $\beta^{twfe}$ to a weighted average of this type of scaled level effect across all values of the dose.

In this section and the next, we define the following weights

$$w^{diff}(d_1, d_2) := \frac{1}{d_2 - d_1}$$
$$w_1^{s,+}(d) := \frac{d - d_L}{d}$$

Also, recall that we defined $m_\Delta(d) = \mathbb{E}[\Delta Y | D = d]$ in the main text—we use this shorthand notation in the results below.

**Proposition S5.** *Under Assumptions 1, 2(a), 3, and 4,*

$$\frac{ATT(d|d)}{d} - \beta^{twfe} = \Big(1 - w_1^{s,+}(d)\Big)\frac{ATT(d_L|d_L)}{d_L}\underbrace{\left(1 - \frac{w_0^{acr}}{\big(1 - w_1^{s,+}(d)\big)}\right)}$$
$$+ \int_{d_L}^{d_U} w_1^{s,+}(d) w^{diff}(d, d_L) m'_\Delta(l) \underbrace{\Big(1 - d w_1^{acr}(l)\Big)} \, dl$$
$$- \left\{ \int_d^{d_U} m'_\Delta(l) w_1^{acr}(l) \, dl \right\}$$

*where* $m'_\Delta(l) = ACRT(l|l) + \frac{\partial ATT(l|h)}{\partial h}\Big|_{h=l}$.

*If Assumption 5 holds instead of Assumption 4, then the same sort of result holds with $ATE(d)$ replacing $ATT(d|d)$ on the LHS of the previous equation and with $m'_\Delta(l) = ACR(l)$ on the RHS of the previous equation.*

*Proof.* To start with, consider the path of outcomes experienced by dose group $d$ relative to the untreated group scaled by $d$:

$$\begin{aligned}
\frac{m_\Delta(d) - m_\Delta(0)}{d} &= \frac{m_\Delta(d) - m_\Delta(d_L)}{d} + \frac{m_\Delta(d_L) - m_\Delta(0)}{d} \\
&= \frac{(d - d_L)}{d}\frac{m_\Delta(d) - m_\Delta(d_L)}{d - d_L} + \frac{d_L}{d}\frac{m_\Delta(d_L) - m_\Delta(0)}{d_L} \\
&= \frac{(d - d_L)}{d}\frac{\int_{d_L}^d m'_\Delta(l)\,dl}{d - d_L} + \frac{d_L}{d}\frac{m_\Delta(d_L) - m_\Delta(0)}{d_L} \\
&= w_1^{s,+}(d)\int_{d_L}^d w^{diff}(d, d_L) m'_\Delta(l)\,dl + \Big(1 - w_1^{s,+}(d)\Big)\frac{m_\Delta(d_L) - m_\Delta(0)}{d_L}
\end{aligned} \tag{S23}$$

where the first equality holds by adding and subtracting $m_\Delta(d_L)/d$, the second equality holds by multiplying and dividing the first term by $(d - d_L)$ and the second term by $d_L$, the third equality holds by the fundamental theorem of calculus, and the last line holds by the definitions of $w^{diff}$ and

$w_1^{s,+}$. Further, notice that the weights integrate/sum to 1:

$$w_1^{s,+}(d)\int_{d_L}^{d} w^{diff}(d,d_L)\,dl + \Big(1-w_1^{s,+}(d)\Big) = \frac{(d-d_L)}{d}\underbrace{\frac{1}{d-d_L}\int_{d_L}^{d} dl}_{=1} + \frac{d_L}{d} = 1$$

which suggests interpreting $(m_\Delta(d)-m_\Delta(0))/d$ as an average of derivative-type terms. Then, using a similar argument for $\beta^{twfe}$ as the one used in Equation (S26) below and combining it with the expression in Equation (S23), we have that

$$\begin{aligned}\frac{m_\Delta(d)-m_\Delta(0)}{d} - \beta^{twfe} &= \Big(1-w_1^{s,+}(d)\Big)\frac{(m_\Delta(d_L)-m_\Delta(0))}{d_L}\left(1-\frac{w_0^{acr}}{\big(1-w_1^{s,+}(d)\big)}\right)\\ &\quad + \int_{d_L}^{d_U} w_1^{s,+}(d) w^{diff}(d,d_L) m'_\Delta(l)\Big(1-d w_1^{acr}(l)\Big)\,dl\\ &\quad - \left\{\int_{d}^{d_U} m'_\Delta(l) w_1^{acr}(l)\,dl\right\}\end{aligned}$$

As in Theorem 3.1, under Assumption 4, $m_\Delta(d)-m_\Delta(0) = ATT(d|d)$, and, as in Theorem 3.2, $m'_\Delta(l) = ACRT(l|l) + \frac{\partial ATT(l|h)}{\partial h}\Big|_{h=l}$ (notice that this term includes selection bias). Under Assumption 5, $m_\Delta(d)-m_\Delta(0) = ATE(d_2)$ and $m'_\Delta(l) = ACR(l)$. This completes the proof. ☐

In other words, in general, $\beta^{twfe}$ is not equal to $ATT(d|d)/d$ (under parallel trends) or $ATE(d)/d$ (under strong parallel trends) for two reasons: (i) they put different weights on the same effects (the underlined terms above), and (ii) the value of $\beta^{twfe}$ additionally depends on effects of the treatment for doses greater than $d$ (the third term, in brackets, in the proposition).

### Scaled-2 × 2 Decomposition for Fixed Doses

Finally, we consider interpreting $\beta^{twfe}$ as $\frac{ATT(d_2|d_2)-ATT(d_1|d_1)}{d_2-d_1}$ or $\frac{ATE(d_2)-ATE(d_1)}{d_2-d_1}$ for two fixed doses $d_1$ and $d_2$ that satisfy $d_L < d_1 < d_2 < d_U$. This is similar to the scaled $2\times 2$ effects discussed in Section 3.3 except for that here we fix the values of $d_1$ and $d_2$ rather than relating $\beta^{twfe}$ to a weighted average of all possible scaled $2\times 2$ effects.

**Proposition S6.** *Under Assumptions 1, 2(a), 3, and 4 and for $d_L < d_1 < d_2 < d_U$,*

$$\begin{aligned}&\frac{ATT(d_2|d_2)-ATT(d_1|d_1)}{d_2-d_1} - \beta^{twfe}\\ &= \int_{d_1}^{d_2} w^{diff}(d_1,d_2) m'_\Delta(l)\underbrace{\Big(1-(d_2-d_1)w_1^{acr}(l)\Big)}\,dl\\ &\quad - \left\{\int_{d_L}^{d_1} m'_\Delta(l) w_1^{acr}(l)\,dl + \int_{d_2}^{d_U} m'_\Delta(l) w_1^{acr}(l)\,dl + w_0^{acr}\frac{(m_\Delta(d_L)-m_\Delta(0))}{d_L}\right\}\end{aligned}$$

*where $m'_\Delta(l) = ACRT(l|l) + \frac{\partial ATT(l|h)}{\partial h}\Big|_{h=l}$.*

*If Assumption 5 holds instead of Assumption 4, then the same sort of result holds with $ATE(d_2) - ATE(d_1)$ replacing $ATT(d_2|d_2) - ATT(d_1|d_1)$ on the LHS of the previous equation and with $m'_\Delta(l) = ACR(l)$ on the RHS of the previous equation.*

*Proof.* To start with, consider the path of outcomes under dose $d_2$ relative to the path of outcomes under dose $d_1$ scaled by $(d_2 - d_1)$, and notice that

$$\frac{m_\Delta(d_2) - m_\Delta(d_1)}{d_2 - d_1} = \int_{d_1}^{d_2} \frac{1}{d_2 - d_1} m'_\Delta(l)\, dl = \int_{d_1}^{d_2} w^{diff}(d_1, d_2) m'_\Delta(l)\, dl \tag{S24}$$

where the first equality holds by the law of iterated expectations, and the second equality by the definition of $w^{diff}$. In addition, notice that the "weights" here integrate to one:

$$\int_{d_1}^{d_2} w^{diff}(d_1, d_2)\, dl = \frac{1}{d_2 - d_1} \int_{d_1}^{d_2} dl = 1$$

Now, move to considering $\beta^{twfe}$. From Equation (B.6) in the proof of Theorem 3.4 in the main text, we have that

$$\beta^{twfe} = \mathbb{E}\left[\frac{(D - \mathbb{E}[D])}{\mathrm{Var}(D)}(m_\Delta(D) - m_\Delta(d_L))\Big| D > 0\right] \mathbb{P}(D > 0) + \mathbb{E}\left[\frac{(D - \mathbb{E}[D])}{\mathrm{Var}(D)}(m_\Delta(d_L) - m_\Delta(0))\Big| D > 0\right] \mathbb{P}(D > 0)$$

Focusing on the first term in the above expression, and, again, from the proof of Theorem 3.4, we have that

$$\begin{aligned}
&\mathbb{E}\left[\frac{(D - \mathbb{E}[D])}{\mathrm{Var}(D)}(m_\Delta(D) - m_\Delta(d_L))\Big| D > 0\right] \mathbb{P}(D > 0) \\
&\qquad = \int_{d_L}^{d_U} m'_\Delta(l) w_1^{acr}(l)\, dl \\
&\qquad = \left\{ \int_{d_L}^{d_1} m'_\Delta(l) w_1^{acr}(l)\, dl + \int_{d_1}^{d_2} m'_\Delta(l) w_1^{acr}(l)\, dl + \int_{d_2}^{d_U} m'_\Delta(l) w_1^{acr}(l)\, dl \right\}
\end{aligned} \tag{S25}$$

where the second equality just splits the integral into three parts and, as in the main text, $w_1^{acr}(l) = \frac{(\mathbb{E}[D|D\geq l] - \mathbb{E}[D])\mathbb{P}(D\geq l)}{\mathrm{Var}(D)}$. Taking the difference between the expressions in Equations S24 and S25, we have that

$$\begin{aligned}
\frac{m_\Delta(d_2) - m_\Delta(d_1)}{d_2 - d_1} - \beta^{twfe} &= \int_{d_1}^{d_2} w^{diff}(d_1, d_2) m'_\Delta(l) \underbrace{\Big(1 - (d_2 - d_1) w_1^{acr}(l)\Big)}\, dl \\
&\quad - \left\{ \int_{d_L}^{d_1} m'_\Delta(l) w_1^{acr}(l)\, dl + \int_{d_2}^{d_U} m'_\Delta(l) w_1^{acr}(l)\, dl + w_0^{acr} \frac{(m_\Delta(d_L) - m_\Delta(0))}{d_L} \right\}
\end{aligned} \tag{S26}$$

where, as in the main text, $w_0^{acr} = \frac{(\mathbb{E}[D|D>0] - \mathbb{E}[D])\mathbb{P}(D>0) d_L}{\mathrm{Var}(D)}$.

As in Theorem 3.2, under Assumption 4, $m_\Delta(d_2) - m_\Delta(d_1) = ATT(d_2|d_2) - ATT(d_1|d_1) = \mathbb{E}[Y_{t=2}(d_2) - Y_{t=2}(d_1)|D = d_2] + \Big(ATT(d_1|d_2) - ATT(d_2|d_2)\Big)$ and $m'_\Delta(l) = ACRT(l|l) + \frac{\partial ATT(l|h)}{\partial h}\Big|_{h=l}$ (notice that both of these expressions also include selection bias). Under Assumption 5, $m_\Delta(d_2) - m_\Delta(d_1) = ATE(d_2) - ATE(d_1)$ and $m'_\Delta(l) = ACR(l)$. This completes the proof. □

This shows that, in general, $\beta^{twfe}$ will be different from $\frac{ATT(d_2|d_2) - ATT(d_1|d_1)}{d_2 - d_1}$ (under parallel trends) or $\frac{ATE(d_2) - ATE(d_1)}{d_2 - d_1}$ (under strong parallel trends) due to (i) different weights on underlying derivative terms (i.e., $m'_\Delta(l)$) for values of $l$ between $d_1$ and $d_2$ (this is the underlined term in the expression in the proposition), and (ii) because $\beta^{twfe}$ additionally depends on effects of the treatment for values outside of $[d_1, d_2]$ (this is the second term in curly brackets in the expression in the proposition).

## SC.3 TWFE Decomposition with a Multi-Valued Discrete Treatment

The following theorem provides the discrete analog of Theorem 3.4 from the main text. The weights in the decomposition are the same ones as those used in the main text which are reported in Table 1 in the main text with the exception that $f_D(l)$ should be understood as $p_l$ in the discrete case. In this section, we continue to use the notation $m_\Delta(d) = \mathbb{E}[\Delta Y | D = d]$.

**Theorem S4.** *Under Assumptions 1, 2(b), 3, and 4, $\beta^{twfe}$ can be decomposed in the following ways:*

*(a) Causal Response Decomposition:*

$$\beta^{twfe} = \sum_{d_j \in \mathcal{D}_+^{mv}} w_1^{acr}(d_j)(d_j - d_{j-1}) \left\{ \frac{ACRT(d_j|d_j)}{d_j - d_{j-1}} + \underbrace{\frac{\Big(ATT(d_{j-1}|d_j) - ATT(d_{j-1}|d_{j-1})\Big)}{d_j - d_{j-1}}}_{\text{selection bias}} \right\}$$

*where the weights, $w_1^{acr}(d_j)(d_j - d_{j-1})$ are always positive and sum to 1.*

*(b) Levels Decomposition:*

$$\beta^{twfe} = \sum_{d_j \in \mathcal{D}_+^{mv}} w_1^{lev}(d_j) ATT(d_j|d_j)$$

*where $w_1^{lev}(d_j) \lesseqgtr 0$ for $d_j \lesseqgtr \mathbb{E}[D]$, and $\sum_{d_j \in \mathcal{D}_+^{mv}} w_1^{lev}(d_j) + w_0^{lev} = 0$.*

*(c) Scaled Levels Decomposition:*

$$\beta^{twfe} = \sum_{d_j \in \mathcal{D}_+^{mv}} w^s(d_j) \frac{ATT(d_j|d_j)}{d_j},$$

*where $w^s(d_j) \lesseqgtr 0$ for $d_j \lesseqgtr \mathbb{E}[D]$, and $\sum_{d_j \in \mathcal{D}_+^{mv}} w^s(d_j) = 1$.*

*(d) Scaled $2 \times 2$ Decomposition*

$$\beta^{twfe} = \sum_{l \in \mathcal{D}} \sum_{h \in \mathcal{D}, h > l} w_1^{2\times 2}(l, h) \left\{ \underbrace{\frac{\mathbb{E}[Y_{t=2}(h) - Y_{t=2}(l) | D = h]}{h - l}}_{\text{causal response}} + \underbrace{\frac{\Big(ATT(l|h) - ATT(l|l)\Big)}{h - l}}_{\text{selection bias}} \right\}$$

*where the weights are always positive and sum to 1.*

*If one imposes Assumption 5 instead of Assumption 4, then the selection bias terms from part (a) and part (d) become zero, and the remainder of the decompositions remain true, except one needs to replace $ACRT(d_j|d_j)$ with $ACR(d_j)$ in part (a), $ATT(d_j|d_j)$ with $ATE(d_j)$ in parts (b) and (c), and $\mathbb{E}[Y_{t=2}(h) - Y_{t=2}(l)|D = h]$ with $\mathbb{E}[Y_{t=2}(h) - Y_{t=2}(l)]$ in part (d).*

## Proof of Theorem S4

We follow the same proof strategy as for the continuous case in the main text and mainly emphasize the parts of the proof that are different from those in the continuous case. As in the continuous case, our strategy is to provide a mechanical decomposition in terms of $m_\Delta(d) = \mathbb{E}[\Delta Y|D = d]$. Then, given those results, the main results in the theorem hold because, under Assumption 4

- $m_\Delta(d_j) - m_\Delta(0) = ATT(d_j|d_j)$
- $m_\Delta(d_j) - m_\Delta(d_{j-1}) = ACRT(d_j|d_j) + \underbrace{\Big(ATT(d_{j-1}|d_j) - ATT(d_{j-1}|d_{j-1})\Big)}_{\text{selection bias}}$
- For $(h, l) \in \mathcal{D}_+^{mv} \times \mathcal{D}_+^{mv}$, $m_\Delta(h) - m_\Delta(l) = ATT(h|h) - ATT(l|l) = \mathbb{E}[Y_{t=2}(h) - Y_{t=2}(l)|D = h] + \underbrace{\Big(ATT(l|h) - ATT(l|l)\Big)}_{\text{selection bias}}$

or, when Assumption 5 holds,

- $m_\Delta(d_j) - m_\Delta(0) = ATE(d_j)$
- $m_\Delta(d_j) - m_\Delta(d_{j-1}) = ACR(d_j)$
- For $(h, l) \in \mathcal{D}_+^{mv} \times \mathcal{D}_+^{mv}$, $m_\Delta(h) - m_\Delta(l) = ATE(h) - ATE(l) = \mathbb{E}[Y_{t=2}(h) - Y_{t=2}(l)]$

## Proof of Theorem S4(a)

*Proof.* Notice that,

$$\begin{aligned}
\beta^{twfe} &= \mathbb{E}\left[\frac{(D - \mathbb{E}[D])}{\text{Var}(D)}(m_\Delta(D) - m_\Delta(0))\right] \\
&= \frac{1}{\text{Var}(D)} \sum_{d \in \mathcal{D}} (d - \mathbb{E}[D])(m_\Delta(d) - m_\Delta(0)) p_d \\
&= \frac{1}{\text{Var}(D)} \sum_{d \in \mathcal{D}} (d - \mathbb{E}[D]) p_d \sum_{d_j \in \mathcal{D}_+^{mv}} \mathbf{1}\{d_j \leq d\}(m_\Delta(d_j) - m_\Delta(d_{j-1})) \\
&= \frac{1}{\text{Var}(D)} \sum_{d_j \in \mathcal{D}_+^{mv}} (m_\Delta(d_j) - m_\Delta(d_{j-1})) \sum_{d \in \mathcal{D}} (d - \mathbb{E}[D]) \mathbf{1}\{d \geq d_j\} p_d \\
&= \sum_{d_j \in \mathcal{D}_+^{mv}} (m_\Delta(d_j) - m_\Delta(d_{j-1})) \frac{(\mathbb{E}[D|D \geq d_j] - \mathbb{E}[D])\mathbb{P}(D \geq d_j)}{\text{Var}(D)} \\
&= \sum_{d_j \in \mathcal{D}_+^{mv}} w_1^{acr}(d_j)(d_j - d_{j-1}) \frac{(m_\Delta(d_j) - m_\Delta(d_{j-1}))}{(d_j - d_{j-1})}
\end{aligned}$$

where the second equality holds by writing the expectation as a summation, the third equality holds by adding and subtracting $m_\Delta(d_j)$ for all $d_j$'s between 0 and $d$, the fourth equality holds by changing the order of the summations, the fifth equality writes the second summation as an expectation, and the last equality holds by the definition of the weights and by multiplying and dividing by $(d_j - d_{j-1})$.

That $w_1^{acr}(d_j)(d_j - d_{j-1}) > 0$ holds immediately since $w_1^{acr}(d_j) \geq 0$ for all $d_j \in \mathcal{D}_+^{mv}$ and $d_j > d_{j-1}$. Further,

$$\begin{aligned}
&\sum_{d_j \in \mathcal{D}_+^{mv}} w_1^{acr}(d_j)(d_j - d_{j-1}) \\
&= \left( \sum_{d_j \in \mathcal{D}_+^{mv}} \mathbb{E}[\mathbf{1}\{D \geq d_j\}D](d_j - d_{j-1}) - \mathbb{E}[D] \sum_{d_j \in \mathcal{D}_+^{mv}} \mathbb{P}(D \geq d_j)(d_j - d_{j-1}) \right) \Big/ \mathrm{Var}(D) \\
&= (A - B)/\mathrm{Var}(D)
\end{aligned}$$

We consider each of these terms in turn:

$$\begin{aligned}
A &= \sum_{d_j \in \mathcal{D}_+^{mv}} \sum_{d_k \in \mathcal{D}} \mathbf{1}\{d_k \geq d_j\} d_k p_{d_k} (d_j - d_{j-1}) \\
&= \sum_{d_k \in \mathcal{D}} p_{d_k} d_k \sum_{d_j \in \mathcal{D}_+^{mv}, d_j \leq d_k} (d_j - d_{j-1}) \\
&= \sum_{d_k \in \mathcal{D}} p_{d_k} d_k (d_k - 0) \\
&= \mathbb{E}[D^2]
\end{aligned}$$

where the first equality holds by writing the expectation for Term A as a summation, the second equality holds by re-ordering the summations, the third equality holds by canceling all the duplicate $d_j$ terms across summations (and because $d_0 = 0$), and the last equality holds by the definition of $\mathbb{E}[D^2]$.

Next,

$$\begin{aligned}
B &= \mathbb{E}[D] \sum_{d_j \in \mathcal{D}_+^{mv}} \sum_{d_k \in \mathcal{D}} \mathbf{1}\{d_k \geq d_j\} p_{d_k} (d_j - d_{j-1}) \\
&= \mathbb{E}[D] \sum_{d_k \in \mathcal{D}} p_{d_k} \sum_{d_j \in \mathcal{D}_+^{mv}, d_j \leq d_k} (d_j - d_{j-1}) \\
&= \mathbb{E}[D] \sum_{d_k \in \mathcal{D}} d_k p_{d_k} \\
&= \mathbb{E}[D]^2
\end{aligned}$$

where the first equality holds by writing the expectation for Term B as a summation, the second equality holds by re-ordering the summations, the third equality holds by canceling all the duplicate $d_j$ terms across summations (and because $d_0 = 0$), and the last equality holds by the definition of $\mathbb{E}[D]$.

This implies that $A - B = \mathrm{Var}(D)$, which implies that the weights sum to 1. □

## Proof of Theorem S4(b)

*Proof.* The proof is analogous to the continuous case in Theorem 3.4(b) in the main text except for replacing the integral with a summation and $f_D(l)$ with $p_l$. Then the result holds by the definition of $w^{lev}$. □

### Proof of Theorem S4(c)

*Proof.* The proof is analogous to the continuous case in Theorem 3.4(c) in the main text except for replacing the integral with a summation and $f_D(l)$ with $p_l$. Then the result holds by the definition of $w^s$. □

### Proof of Theorem S4(d)

*Proof.* Up to Equation (B.14) in the main text, the steps of the proof of Theorem 3.4(d) for the continuous case carry over to the discrete case. Under Assumption 2(b),

$$\text{Equation (B.14)} = \frac{1}{\text{Var}(D)} \sum_{l \in \mathcal{D}} \sum_{h \in \mathcal{D}, h > l} (h-l)^2 \frac{(m_\Delta(h) - m_\Delta(l))}{(h-l)} p_h p_l$$

which holds immediately from Equation (B.14) and then the result holds by the definition of $w_1^{2\times 2}$. That the weights are positive and sum to 1 holds by the same type of argument as used in the continuous case. □

# SD Additional Details about Comparing Parallel Trends Assumptions

## SD.1 Parallel Trends and Strong Parallel Trends Are Non-Nested

This section considers in more detail the differences between Assumption 4 and Assumption 5. In this section, we show that Assumption 5 is not strictly stronger than Assumption 4 though it is likely to be *stronger in practice* in most applications. Here, we maintain Assumption 3, so $Y_{t=1}(d) = Y_{t=1}(d') = Y_{t=1}(0)$ for any $(d, d') \in \mathcal{D} \times \mathcal{D}$.

To see that Assumption 5 is not strictly stronger, consider the case where there are two doses $d_1$ and $d_2$. In this case, Assumption 4 is equivalent to the following conditions

$$\mathbb{E}[Y_{t=2}(0) - Y_{t-1}(0)|D = d_1] = \mathbb{E}[Y_{t=2}(0) - Y_{t-1}(0)|D = d_2] = \mathbb{E}[Y_{t=2}(0) - Y_{t-1}(0)|D = 0] \quad \text{(S27)}$$

while Assumption 5 is equivalent to

$$\mathbb{E}[Y_{t=2}(0) - Y_{t-1}(0)] = \mathbb{E}[Y_{t=2}(0) - Y_{t-1}(0)|D = 0] \quad \text{(S28)}$$

$$\mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)] = \mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = d_1] \quad \text{(S29)}$$

$$\mathbb{E}[Y_{t=2}(d_2) - Y_{t=1}(0)] = \mathbb{E}[Y_{t=2}(d_2) - Y_{t=1}(0)|D = d_2]. \quad \text{(S30)}$$

Assumption 4 does not place any restrictions on any potential outcomes besides untreated potential outcomes, and therefore the "extra" conditions in Equations (S29) and (S30) imply that Assumption 5 is not weaker than Assumption 4.

On the other hand, Equation (S28) does not imply Equation (S27); rather, it implies that

$$\mathbb{E}[Y_{t=2}(0) - Y_{t-1}(0)|D = 0] = \mathbb{E}[Y_{t=2}(0) - Y_{t-1}(0)|D = d_1]\frac{p_{d_1}}{p_{d_1} + p_{d_2}} + \mathbb{E}[Y_{t=2}(0) - Y_{t-1}(0)|D = d_2]\frac{p_{d_2}}{p_{d_1} + p_{d_2}}$$

In other words, the trend in untreated potential outcomes does not have to be exactly the same for all doses, but, instead, they have to be the same on average.

In practice, this potentially allows for some units to select their amount of dose on the basis of the path of their untreated potential outcomes (which is not allowed under the standard parallel trends assumption in Assumption 4), but that the amount of selection has to average out across doses to be equal to zero. It seems hard to think of realistic cases where Assumption 5 would be practically weaker than Assumption 4, though.

## SD.2 Proofs of Results from Appendix C

This section provides the proof of Theorem C.1 in the main text which compared different versions of parallel trends assumptions along with their relationship to restrictions on treatment effect heterogeneity.

### Proof of Theorem C.1(a)

*Proof.* Notice that

$$\begin{aligned}
ATT^o &= \mathbb{E}\Big[ATT(D|D)\Big|D > 0\Big] \\
&= \mathbb{E}\Big[\mathbb{E}[Y_{t=2}(D) - Y_{t=2}(0)|D]\Big|D > 0\Big] \\
&= \mathbb{E}\Big[\mathbb{E}[\mathbf{1}\{D > 0\}(Y_{t=2}(D) - Y_{t=2}(0))|D]\Big|D > 0\Big] \\
&= \mathbb{E}\left[\frac{\mathbb{E}[\mathbf{1}\{D > 0\}(Y_{t=2}(D) - Y_{t=2}(0))|D, D > 0]}{\mathbb{P}(D > 0|D)}\Big|D > 0\right] \\
&= \mathbb{E}\Big[\mathbb{E}[Y_{t=2}(D) - Y_{t=2}(0)|D, D > 0]\Big|D > 0\Big] \\
&= \mathbb{E}\Big[Y_{t=2}(D) - Y_{t=2}(0)\Big|D > 0\Big] \\
&= \mathbb{E}\Big[Y_{t=2}(D) - Y_{t=1}(0)\Big|D > 0\Big] - \mathbb{E}\Big[Y_{t=2}(0) - Y_{t=1}(0)\Big|D > 0\Big] \\
&= \mathbb{E}\Big[Y_{t=2}(D) - Y_{t=1}(0)\Big|D > 0\Big] - \mathbb{E}\Big[Y_{t=2}(0) - Y_{t=1}(0)\Big|D = 0\Big] \\
&= \mathbb{E}[\Delta Y|D > 0] - \mathbb{E}[\Delta Y|D = 0]
\end{aligned}$$

where the first equality holds by the definition of $ATT^o$, the second equality holds by the definition of $ATT(d|d)$, the third equality holds because $\mathbf{1}\{D > 0\} = 1$ conditional on $D > 0$, the fourth equality holds by the law of iterated expectations, the fifth equality holds because, conditional on $D > 0$, $\mathbf{1}\{D > 0\} = 1$ and $\mathbb{P}(D > 0|D) = 1$, the sixth equality holds by the law of iterated expectations, the seventh equality holds by adding and subtracting $\mathbb{E}[Y_{t=1}(0)|D > 0]$, the eighth equality holds by aggregate parallel trends, and the last equality holds by replacing potential outcomes with their corresponding observed outcomes. □

### Proof of Theorem C.1(b)

*Proof.* This is a restatement of Theorem 3.1 and holds from that result. □

**Proof of Theorem C.1(c)**

*Proof.* This is a restatement of Theorem 3.3 and holds from that result. □

**Proof of Theorem C.1(d)**

*Proof.* Start with the definition of $ATE(d)$

$$\begin{aligned}
ATE(d) &= \mathbb{E}[Y_{t=2}(d) - Y_{t=2}(0)] \\
&= \mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)] - \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)] \\
&= \mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)|D = d] - \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d] \\
&= \mathbb{E}[Y_{t=2}(d) - Y_{t=2}(0)|D = d] \\
&= ATT(d|d)
\end{aligned}$$

where the third equality uses strong parallel trends on the first term and parallel trends on the second term (by itself, strong parallel trends would imply that the second term is equal to $\mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = 0]$ only). That $ATT(d|d) = \mathbb{E}[\Delta Y|D = d] - \mathbb{E}[\Delta Y|D = 0]$ holds by Theorem 3.1 in the main text. This completes the proof. □

**Proof of Theorem C.1(e)**

*Proof.* We start by showing that $ATE(d) = ATT(d|d')$ for any $d'$. Starting with the definition of $ATE(d)$,

$$\begin{aligned}
ATE(d) &= \mathbb{E}[Y_{t=2}(d) - Y_{t=2}(0)] \\
&= \mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)] - \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)] \\
&= \mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)|D = d'] - \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d'] \\
&= \mathbb{E}[Y_{t=2}(d) - Y_{t=2}(0)|D = d'] \\
&= ATT(d|d')
\end{aligned}$$

where the second equality holds by adding and subtracting $\mathbb{E}[Y_{t=1}(0)]$, the third equality holds by Assumption 5-Alt (also, notice that this equality does not hold under Assumption 4 or Assumption 5), the fourth equality holds by canceling the two $\mathbb{E}[Y_{t=1}(0)|D = d']$ terms, and the remaining term in that equality is $ATT(d|d')$.

The previous argument holds for any $d'$, including $d' = d$; therefore, $ATE(d) = ATT(d|d') = ATT(d|d)$. Finally, notice that Assumption 5-Alt implies Assumption 4 by taking $d = 0$ in the statement of the assumption; thus, $ATT(d|d) = \mathbb{E}[\Delta Y|D = d] - \mathbb{E}[\Delta Y|D = 0]$ by Theorem 3.1. This completes the proof. □

**Proof of Theorem C.1(f)**

*Proof.* Consider the case with two doses, $d_1$ and $d_2$. We will start by showing that $ATT(d_1|d_1)$ is not identified under aggregate parallel trends alone. Notice that

$$\begin{aligned}ATT(d_1|d_1) &= \mathbb{E}[Y_{t=2}(d_1) - Y_{t=2}(0)|D = d_1]\\ &= \mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = d_1] - \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d_1]\\ &= \mathbb{E}[\Delta Y|D = d_1] - \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d_1]\end{aligned}$$

so that identifying $ATT(d_1|d_1)$ comes down to identifying $\mathbb{E}[Y_{t=2}(0)-Y_{t=1}(0)|D = d_1]$. In the setting with dose groups 0, $d_1$, and $d_2$, the only restriction from aggregate parallel trends is that

$$\mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = 0] = \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D > 0]$$

which is equivalent to

$$\begin{aligned}\mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = 0] &= \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d_1]\frac{p_{d_1}}{p_{d_1} + p_{d_2}}\\ &\quad + \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d_2]\frac{p_{d_2}}{p_{d_1} + p_{d_2}}\end{aligned}$$

In this equation, $\mathbb{E}[Y_{t=2}(0)-Y_{t=1}(0)|D = 0]$ and the terms involving probabilities are identified by the sampling process, but $\mathbb{E}[Y_{t=2}(0) - Y_{t-1}(0)|D = d_1]$ and $\mathbb{E}[Y_{t=2}(0) - Y_{t-1}(0)|D = d_2]$ are not identified by the sampling process, and any combination of these two conditional expectations that satisfies the previous equation also satisfies the aggregate parallel trends assumption. This implies that aggregate parallel trends does not lead to point identification of $ATT(d_1|d_1)$ and completes the proof.

Next, we show that aggregate parallel trends does not identify $ATE(d_1)$. Notice that

$$\begin{aligned}ATE(d_1) &= \mathbb{E}[Y_{t=2}(d_1) - Y_{t=2}(0)]\\ &= \mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)] - \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)]\end{aligned}$$

where the second equality holds by adding and subtracting $\mathbb{E}[Y_{t=1}(0)]$. The first term gives the average path of potential outcomes under dose $d_1$ while the second term involves the path of untreated potential outcomes. Focusing on the first term, we have that

$$\begin{aligned}\mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)] &= \mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = 0]p_0\\ &\quad + \mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = d_1]p_{d_1}\\ &\quad + \mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = d_2]p_{d_2}\end{aligned}$$

In this expression, $\mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = d_1]$ and the probabilities are identified by the sampling process. The other terms are not identified by the sampling process, nor does aggregate parallel trends restrict the path of potential outcomes under dose $d_1$. Therefore, $ATE(d_1)$ is not identified under aggregate parallel trends. ☐

**Proof of Theorem C.1(g)**

*Proof.* To show the result, consider the case with two doses $d_1$ and $d_2$. We will show that $ATE(d_1)$ is not identified under parallel trends. The argument is very similar to the proof of the second part of part (f). In particular,

$$ATE(d) = \mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)] - \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)]$$

And, if we focus on the first term, we have that

$$\begin{aligned}\mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)] &= \mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = 0]p_0 \\ &+ \mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = d_1]p_{d_1} \\ &+ \mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = d_2]p_{d_2}\end{aligned}$$

where $\mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = d_1]$ and the probabilities are identified by the sampling process, but the conditional expectations $\mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = 0]$ and $\mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = d_2]$ are not identified by the sampling process. Since parallel trends does not restrict the path of outcomes under dose $d_1$, it implies that $ATE(d_1)$ is not identified under parallel trends. □

**Proof of Theorem C.1(h)**

*Proof.* Consider the case with two doses, $d_1$ and $d_2$. We will show that $ATT(d_1|d_1)$ is not point-identified under strong parallel trends alone. As in the proof of part (f), identifying $ATT(d_1|d_1)$ comes down to identifying $\mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d_1]$ The only restriction on the path of untreated potential outcomes from strong parallel trends is that

$$\mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)] = \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = 0]$$

In the current setting where the dose groups are $0, d_1, d_2$, this is equivalent to

$$\begin{aligned}\mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = 0] &= \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d_1]\frac{p_{d_1}}{p_{d_1} + p_{d_2}} \\ &+ \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d_2]\frac{p_{d_2}}{p_{d_1} + p_{d_2}}\end{aligned}$$

which holds by applying the law of iterated expectations to the LHS of the previous equation and then re-arranging. In the previous expression, the probabilities are identified from the sampling process, but $\mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d_1]$ and $\mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d_2]$ are not. And any values of these two terms that satisfies the previous equation additionally satisfies strong parallel trends. This implies that $ATT(d_1|d_1)$ is not identified under strong parallel trends alone.

□

**Proof of Theorem C.1(i)**

*Proof.* Consider the same case as in Part (f) with two doses, $d_1$ and $d_2$. Since parallel trends holds for this part, it implies that $ATT(d_1|d_1)$ and $ATT(d_2|d_2)$ are identified. We will show that $ATT(d_1|d_2)$

is not identified under the combination of parallel trends and strong parallel trends. Notice that

$$\begin{aligned}
ATT(d_1|d_2) &= \mathbb{E}[Y_{t=2}(d_1) - Y_{t=2}(0)|D = d_2] \\
&= \mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = d_2] - \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d_2] \\
&= \mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = d_2] - \mathbb{E}[\Delta Y|D = 0]
\end{aligned}$$

where the first line is the definition of $ATT(d_1|d_2)$, the second line adds and subtracts $\mathbb{E}[Y_{t=1}(0)|D = d_2]$, and third line holds by parallel trends. The equation above implies that identifying $ATT(d_1|d_2)$ comes down to identifying $\mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = d_2]$. Parallel trends places no restrictions on this term. The only restriction from strong parallel trends on the path of outcomes under does $d_1$ is that

$$\mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)] = \mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = d_1]$$

In the current setting where the dose groups are $0, d_1, d_2$, this is equivalent to

$$\begin{aligned}
\mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = d_1] &= \mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = 0]\frac{p_0}{p_0 + p_{d_2}} \\
&\quad + \mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = d_2]\frac{p_{d_2}}{p_0 + p_{d_2}}
\end{aligned}$$

which holds by applying the law of iterated expectations to the LHS of the previous equation and then re-arranging. In the previous expression, the probabilities are identified from the sampling process, but $\mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = 0]$ and $\mathbb{E}[Y_{t=2}(d_1) - Y_{t=1}(0)|D = d_2]$ are not. And any values of these two terms that satisfies the previous equation additionally satisfies strong parallel trends. This implies that $ATT(d_1|d_2)$ is not identified under the combination of parallel trends and strong parallel trends. □

**Proof of Theorem C.1(j)**

*Proof.* Given the result in part (b), if aggregate parallel trends implied parallel trends, then this would imply that $ATT(d|d)$ could be recovered under aggregate parallel trends, but this contradicts part (f) above. Similarly, given the result in part (c), if aggregate parallel trends implied strong parallel trends, then this would imply that $ATE(d)$ could be recovered under aggregate parallel trends, but this also contradicts part (f) above. □

**Proof of Theorem C.1(k)**

*Proof.* We start by showing that parallel trends implies aggregate parallel trends. Notice that

$$\begin{aligned}
\mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D > 0] &= \mathbb{E}\Big[\mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D, D > 0]\Big|D > 0\Big] \\
&= \mathbb{E}\Big[\mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D]\Big|D > 0\Big] \\
&= \mathbb{E}\Big[\mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = 0]\Big|D > 0\Big] \\
&= \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = 0]
\end{aligned}$$

where the first equality holds by the law of iterated expectations, the second equality holds using an argument similar to the one provided in part (a), the third equality holds by parallel trends, and the last equality holds because $\mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = 0]$ is non-random.

Next, we show that strong parallel trends implies aggregate parallel trends. Notice that

$$\begin{aligned}\mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D > 0] &= \frac{\mathbb{E}[Y_{t=2}(0) - Y_{t=2}(0)] - \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = 0]p_0}{\mathbb{P}(D > 0)} \\ &= \frac{\mathbb{E}[Y_{t=2}(0) - Y_{t=2}(0)|D = 0] - \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = 0]p_0}{1 - p_0} \\ &= \mathbb{E}[Y_{t=2}(0) - Y_{t=2}(0)|D = 0]\end{aligned}$$

where the first equality holds by the law of iterated expectations (and re-arranging terms), the second equality holds by strong parallel trends and because $\mathbb{P}(D > 0) = 1 - \mathbb{P}(D = 0)$, and the last equality holds by combining terms and canceling terms. □

**Proof of Theorem C.1(l)**

*Proof.* Given the results in parts (b), if strong parallel trends implied parallel trends, then this would imply that would imply that $ATT(d|d)$ could be recovered under strong parallel trends, which contradicts part (h) above. Similarly, given the result in part (c), if parallel trends implied strong parallel trends, then this would imply that $ATE(d)$ could be recovered under parallel trends, which contradicts part (g) above. Therefore, these are non-tested assumptions. See also the discussion above in Appendix SD.1 for an example. □

**Proof of Theorem C.1(m)**

*Proof.* That alternative strong parallel trends implies parallel trends holds immediately from the definition of alternative strong parallel trends by taking $d = 0$.

To show that it also implies strong parallel trends, starting with the LHS of strong parallel trends assumption (Assumption 5), notice that for any $d$

$$\begin{aligned}\mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)] &= \int_{\mathcal{D}} \mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)|D = l]\, dF_D(l) \\ &= \mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)|D = d] \int_{\mathcal{D}} dF_D(l) \\ &= \mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)|D = d]\end{aligned}$$

where the first equality holds by the law of iterated expectations, the second equality uses alternative strong parallel trends, and the last line holds immediately. This shows that alternative strong parallel trends implies strong parallel trends. □

**Proof of Theorem C.1(n)**

*Proof.* Start with the LHS of the strong parallel trends assumption

$$\begin{aligned}\mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)] &= \mathbb{E}[Y_{t=2}(d) - Y_{t=2}(0)] + \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)] \\ &= ATE(d) + \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d] \\ &= ATT(d|d) + \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d]\end{aligned}$$

$$= \mathbb{E}[Y_{t=2}(d) - Y_{t=2}(0)|D = d] + \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d]$$
$$= \mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)|D = d]$$

where the second equality uses the definition of $ATE(d)$ and the parallel trends assumption, and the third equality uses that $ATE(d) = ATT(d|d)$. □

**Proof of Theorem C.1(o)**

*Proof.* Start with the LHS of alternative strong parallel trends

$$\begin{aligned}
\mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)|D = l] &= \mathbb{E}[Y_{t=2}(d) - Y_{t=2}(0)|D = l] + \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = l] \\
&= ATT(d|l) + \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d] \\
&= ATT(d|d) + \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d] \\
&= \mathbb{E}[Y_{t=2}(d) - Y_{t=2}(0)|D = d] + \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = d] \\
&= \mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)|D = d]
\end{aligned}$$

where the second equality uses parallel trends and the third equality uses $ATT(d|d') = ATT(d|d)$. This shows the result. □

# SE Relaxing Strong Parallel Trends

In this section, we provide more details about the three possible ideas to weaken the strong parallel trends assumption that were discussed in Section 5 in the main text.

## SE.1 Partial Identification

To start with, we consider the case where a researcher only wishes to invoke parallel trends (Assumption 4) but is willing to assume that the sign of the selection bias is known. Without loss of generality, we focus on the case where there is positive selection bias in the sense that, for dose $d$ and any two dose groups $l$ and $h$ with $l < h$, we have that $ATT(d|l) \leq ATT(d|h)$—this is positive selection bias in that the $ATT$ of any dose is higher for the high dose group, $h$, relative to the low dose group, $l$. The following result shows that, under this sort of condition, we can construct bounds on differences between causal effect parameters at different values of the dose.

**Proposition S7.** *Under Assumptions 1 to 4 and suppose without loss of generality that for any $d \in \mathcal{D}_+$ and $l < h$, $ATT(d|l) < ATT(d|h)$, then the following results hold*

*(1)* $\mathbb{E}[Y_{t=2}(h) - Y_{t=2}(l)|D = h] \leq \mathbb{E}[\Delta Y|D = h] - \mathbb{E}[\Delta Y|D = l] = ATT(h|h) - ATT(l|l)$

*(2)* $ACRT(d|d) \leq \frac{\partial \mathbb{E}[\Delta Y|D=d]}{\partial d}$

*Proof.* For part (1), from Theorem 3.2(b) in the main text, we have that, under Assumption 4,

$$\mathbb{E}[\Delta Y|D = h] - \mathbb{E}[\Delta Y|D = l] = ATT(h|h) - ATT(l|l)$$

$$= \mathbb{E}[Y_{t=2}(h) - Y_{t=2}(l)|D = h] + \underbrace{\Big(ATT(l|h) - ATT(l|l)\Big)}_{\geq 0}$$

$$\geq \mathbb{E}[Y_{t=2}(h) - Y_{t=2}(l)|D = h]$$

where the last inequality holds due to the positive selection bias.

For part (2), from Theorem 3.2(b) in the main text, we have that

$$\frac{\partial \mathbb{E}[\Delta Y|D = d]}{\partial d} = ACRT(d|d) + \underbrace{\frac{\partial ATT(d|l)}{\partial l}\bigg|_{l=d}}_{\geq 0}$$

$$\geq ACRT(d|d)$$

where the last inequality holds due to the positive selection bias. □

Part (1) of Proposition S7 says that, given positive selection bias, the average causal response of the high dose, $h$, relative to the low dose, $l$, for the high dose group is bounded by comparing the average path of outcomes over time for the high dose group relative to the low dose group. Part (2) says that, under positive selection bias, the $ACRT(d|d)$ is bounded by the derivative of $\mathbb{E}[\Delta Y|D = d]$ with respect to $d$.

## SE.2 Local Strong Parallel Trends

In this section, we consider a local strong parallel trends assumption where, as discussed in the main text, strong parallel trends holds in some sub-region $\mathcal{D}_s \subseteq \mathcal{D}_+$. As discussed in the main text, we focus on identifying a local version of $ATE(d)$ given by $\mathbb{E}[Y_{t=2}(d) - Y_{t=2}(0)|D \in \mathcal{D}_s]$. This is the average treatment effect of dose $d$ across all dose groups that experienced a treatment in $\mathcal{D}_s$. We consider the following assumption

**Assumption S2.** *For all $d \in \mathcal{D}_s \subseteq \mathcal{D}_+$,*

$$\mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)|D \in \mathcal{D}_s] = \mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)|D = d]$$

This is an analogous assumption to Assumption 5 from the main text with the exception that it holds locally to the sub-region $\mathcal{D}_s$ rather than for all $\mathcal{D}$. Next, we show that $\mathbb{E}[Y_{t=2}(d) - Y_{t=2}(0)|D \in \mathcal{D}_s]$ is identified under this assumption.[11]

**Proposition S8.** *Under Assumptions 1 to 4 and S2, and for $d \in \mathcal{D}_s$*

$$\mathbb{E}[Y_{t=2}(d) - Y_{t=2}(0)|D \in \mathcal{D}_s] = \mathbb{E}[\Delta Y|D = d] - \mathbb{E}[\Delta Y|D = 0]$$

[11] In the proposition, we use both Assumption S2 and Assumption 4 from the main text. The latter assumption is used to deal with the path of untreated potential outcomes that shows up in the proof of the proposition. This is slightly different from how we used Assumption 5 to identify $ATE(d)$ in the main text. It is possible to slightly adjust Assumption S2 to include untreated potential outcomes and then have the proof use the analogous steps to the ones for Theorem 3.3 in the main text. Instead, in our view, the combination of Assumption S2 and Assumption 4 could be somewhat more attractive in empirical applications.

*Proof.* For any $d \in \mathcal{D}_s$, we have that

$$\begin{aligned}\mathbb{E}[Y_{t=2}(d) - Y_{t=2}(0)|D \in \mathcal{D}_s] &= \mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)|D \in \mathcal{D}_s] - \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D \in \mathcal{D}_s] \\ &= \mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)|D = d] - \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|D = 0] \\ &= \mathbb{E}[\Delta Y|D = d] - \mathbb{E}[\Delta Y|D = 0]\end{aligned}$$

where the first equality holds by adding and subtracting $\mathbb{E}[Y_{t=1}(0)|D \in \mathcal{D}_s]$, the second equality uses Assumption S2 for the first term and Assumption 4 for the second term, and the last equality holds by replacing potential outcomes with their observed counterparts. □

An immediate corollary to the previous result is that a local version of the $ACR$ is also identified: $\dfrac{\partial \mathbb{E}[Y_{t=2}(d) - Y_{t=2}(0)|D \in \mathcal{D}_s]}{\partial d} = \dfrac{\partial \mathbb{E}[\Delta Y|D = d]}{\partial d}$ for $d$ in the interior of $\mathcal{D}_s$—notice that there are no selection bias terms in this expression which is due to this being a version of strong parallel trends.[12]

## SE.3 Strong Parallel Trends Conditional-on-Covariates

In this section, we consider a conditional-on-covariates version of strong parallel trends that can be used to recover conditional $ATE$ parameters. We target $ATE_x(d) := \mathbb{E}[Y_{t=2}(d) - Y_{t=2}(0)|X = x]$. We consider the following assumption

**Assumption S3.** *For all $d \in \mathcal{D}$,*

$$\mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)|X = x] = \mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)|X = x, D = d]$$

This is a conditional-on-covariates version of strong parallel trends. The following result shows that $ATE_x(d)$ is identified under this condition.

**Proposition S9.** *Under Assumptions 1 to 4 and S3,*[13]

$$ATE_x(d) = \mathbb{E}[\Delta Y|X = x, D = d] - \mathbb{E}[\Delta Y|X = x, D = 0]$$

*Proof.* For any $d \in \mathcal{D}$, we have that

$$\begin{aligned}ATE_x(d) &= \mathbb{E}[Y_{t=2}(d) - Y_{t=2}(0)|X = x] \\ &= \mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)|X = x] - \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|X = x] \\ &= \mathbb{E}[Y_{t=2}(d) - Y_{t=1}(0)|X = x, D = d] - \mathbb{E}[Y_{t=2}(0) - Y_{t=1}(0)|X = x, D = 0] \\ &= \mathbb{E}[\Delta Y|X = x, D = d] - \mathbb{E}[\Delta Y|X = x, D = 0],\end{aligned}$$

where the first equality holds by the definition of $ATE_x(d)$, the second equality holds by adding and subtracting $\mathbb{E}[Y_{t=1}(0)|X = x]$, the third equality holds by Assumption S3, and the last equality by replacing potential outcomes with their observed counterparts. □

[12] As an interesting side-comment, unlike the level-effect parameter in Proposition S8, notice that the causal response-type parameter here does not require parallel trends for untreated potential outcomes (i.e., it does not require Assumption 4); this argument is the same as in Appendix SC.1 for the case with no untreated units.

[13] Assumptions 1 and 2 need to be slightly modified so that $X_i$ is included in the random sample and clarity regarding the support of $D$ conditional on $X = x$. We omit writing these as formal conditions for the sake of brevity.

An immediate corollary to the previous result is that the conditional on covariates version of $ACR$ is also identified. In particular, $ACR_x(d) := \frac{\partial ATE_x(d)}{\partial d} = \frac{\partial \mathbb{E}[\Delta Y | X=x, D=d]}{\partial d}$. Notice that there is no selection bias term in this expression.

## References


Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2023). "Revisiting event study designs: Robust and efficient estimation". *Review of Economic Studies* Forthcoming.

Callaway, Brantly (2023). "Difference-in-differences for policy evaluation". *Handbook of Labor, Human Resources and Population Economics.* Ed. by Klaus F. Zimmermann. Springer International Publishing, pp. 1–61.

Callaway, Brantly and Pedro H. C. Sant'Anna (2021). "Difference-in-differences with multiple time periods". *Journal of Econometrics* 225.2, pp. 200–230.

de Chaisemartin, Clement and Xavier D'Haultfœuille (2020). "Two-way fixed effects estimators with heterogeneous treatment effects". *American Economic Review* 110.9, pp. 2964–2996.

Goodman-Bacon, Andrew (2021). "Difference-in-differences with variation in treatment timing". *Journal of Econometrics* 225.2, pp. 254–277.

Marcus, Michelle and Pedro H. C. Sant'Anna (2021). "The role of parallel trends in event study settings: An application to environmental economics". *Journal of the Association of Environmental and Resource Economists* 8.2, pp. 235–275.