xtswitchdid — Difference in differences with switching treatments for panel data[+]

[+] This command is part of StataNow.

Description

`xtswitchdid` estimates event-study effects for panel data using the heterogeneous difference-in-differences (DID) estimator of de Chaisemartin and D'Haultfœuille (2026). These effects are dynamic treatment effects that are of interest in event studies, which explore how treatment effects change over time. Treatments can be binary or multivalued. Switching (nonabsorbing) treatments are also allowed—subjects may switch in and out of treatment or to different levels of treatment.

Quick start

Estimate event-study effects of treatment `treat` on outcome `y` with group `grpvar` on `xtset` data

```
xtswitchdid (y) (treat), group(grpvar)
```

Same as above, but model `y` using covariate `x`

```
xtswitchdid (y x) (treat), group(grpvar)
```

Same as above, but estimate effects for at most three periods of exposure

```
xtswitchdid (y x) (treat), group(grpvar) neffects(3)
```

Same as above, but request raw (nonnormalized) effects

```
xtswitchdid (y x) (treat), group(grpvar) neffects(3) raw
```

Same as above, but request effects only for switchers-in, and require the switchers to be common across all three effects

```
xtswitchdid (y x) (treat), group(grpvar) neffects(3) raw ///
    switchgroup(in) commonswitchers
```

Estimate path-specific treatment effects for a treatment path selected by indicator variable `pathvar`

```
xtswitchdid (y) (treat), group(grpvar) path(pathvar)
```

Menu

Statistics > Causal inference/treatment effects > Continuous outcomes > Heterogeneous DID > Panel-data DID with switching treatments

Syntax

```
xtswitchdid (ovar [omvarlist]) (tvar) [if] [in] [weight], group(groupvar) [options]
```

ovar is the outcome of interest.

omvarlist specifies the covariates in the outcome model and may contain factor variables; see **[U] 11.4.3 Factor variables**.

tvar indicates the treatment level for each observation and may be binary or multivalued.

groupvar is a categorical variable that indicates the group level at which the treatment occurs.

options	Description
Model	
* `group(`*groupvar*`)`	specify group variable
`supergroup(`*supergroupvar*`)`	specify supergroup variable
`switchgroup(`*sgtype*`)`	specify the type of switcher group; default is `switchgroup(all)`
`controlgroup(`*cgtype*`)`	specify the type of control group; default is `controlgroup(all)`
`path(`*pathspec*`)`	request path-specific treatment effects for the specified treatment path
`neffects(`#`)`	specify the number of event-study effects to estimate; default is all possible effects
`commonswitchers`	request that a common set of switchers be used for all event-study effects
`raw`	request event-study effects that have not been normalized
`nolistwise`	do not automatically drop observations with missing values in some variables
SE/Robust	
`vce(`*vcetype*`)`	*vcetype* may be `cluster` *clustvar*, `robust`, `bootstrap`, or `jackknife`
Reporting	
`level(`#`)`	set confidence level; default is `level(95)`
`showsupergroups`	show supergroups in the initial treatment table
display_options	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
`coeflegend`	display legend instead of statistics

*`group()` is required.

If the variable specified by `group()` is coarser than the panel variable, then *omvarlist* may not contain factor variables.

A panel variable and a time variable must be specified using `xtset`; see [XT] **xtset**.

`by`, `collect`, and `statsby` are allowed; see **[U] 11.1.10 Prefix commands**.

`iweights` are allowed; see **[U] 11.1.6 weight**. Weights must be constant within panels.

`coeflegend` does not appear in the dialog box.

See **[U] 20 Estimation and postestimation commands** for more capabilities of estimation commands.

sgtype	Description
`all`	include both switchers-in and switchers-out; the default
`in`	include only switchers-in
`out`	include only switchers-out

cgtype	Description
`all`	use both never-switchers and not-yet-switchers as controls; the default
`never`	use only never-switchers as controls

Options

Model

`group(`*groupvar*`)` specifies a group variable that indicates the group level at which treatment occurs. *groupvar* may be, for example, states, counties, or hospitals. `group()` also defines the clusters for the default cluster–robust standard errors. `group()` is required. *groupvar* may be the panel variable defined in `xtset`, which must in any case be nested within the levels defined by *groupvar*.

`supergroup(`*supergroupvar*`)` specifies a supergroup variable within levels of which groups can be compared with each other. *groupvar* must be nested in *supergroupvar*. For example, if *groupvar* specifies counties, *supergroupvar* may specify states. If no supergroup variable is specified, switchers are compared with all available control groups with the same initial treatment value. If a supergroup variable is specified, however, switchers are compared with only control groups that are also in the same supergroup.

`switchgroup(`*sgtype*`)` specifies the groups to use as switchers. Switchers are either switchers-in, meaning their first treatment different from their initial treatment is greater than the initial treatment, or switchers-out, meaning their first treatment different from their initial treatment is less than the initial treatment. *sgtype* can be one of `all` (referring to all switchers), `in` (referring to switchers-in), or `out` (referring to switchers-out). The default is `switchgroup(all)`.

`controlgroup(`*cgtype*`)` specifies the groups to use as controls. Control groups are either never-switchers, meaning that they keep the same value of the treatment variable until the last period, or not-yet-switchers, meaning that they keep their initial treatment value up to a specific period. *cgtype* can be `all` (referring to all possible control groups) or `never` (referring only to the never-switchers). The default is `controlgroup(all)`.

`path(`*pathspec*`)` specifies a treatment path for which treatment effects will be estimated. Only those switchers whose path matches the indicated path will be included in the estimation.

pathspec may be specified as *numlist*, indicating treatment values starting at the period before treatment. For example, `0 2 2 2` indicates a treatment path where groups have an initial treatment value of 0 and then stay at the treated value of 2 for at least 3 periods.

pathspec may alternatively be an indicator variable, where observations with the desired path have the value 1 (or any positive number) and other observations have the value 0. Path variables must indicate only one distinct path, which will be truncated at the length specified by `neffects()` if needed. If multiple groups have the same treatment path, the path variable need not indicate the path in every such group; an indicator for one such group is sufficient to define the path. Path variables can be generated using `estat paths` or constructed manually.

neffects(#) specifies the maximum number of event-study effects to estimate, starting from exposure $\ell = 1$. By default, xtswitchdid estimates and reports as many effects as are available in the data.

commonswitchers requests that each event-study effect be estimated using the same set of switchers. If commonswitchers is not specified, all available switchers are used for each effect, which typically means that shorter exposures are estimated using much more data than longer exposures. Specifying commonswitchers ensures that the same set of data is used for all estimated effects. Note that if commonswitchers and neffects() are specified together, the common set of switchers will be common up to the maximum number of effects specified in neffects().

raw disables the normalization of event-study effects. The default when this option is not specified is to normalize the event-study effect for exposure ℓ by the average incremental treatment dose experienced by switching groups, relative to their baseline treatment, over the ℓ periods starting from when they first switch.

nolistwise requests that the default behavior of dropping observations with missing values in any variables used for estimation be disabled. Instead, observations where the outcome or covariates are missing may still be used to infer treatment paths.

SE/Robust

vce(*vcetype*) specifies the type of standard error reported, which includes types that allow for intragroup correlation (cluster *clustvar*), that are robust to intragroup correlation among group variables (robust), and that use bootstrap or jackknife sampling done at the panel level (bootstrap, jackknife); see [R] ***vce_option***.

vce(cluster *clustvar*) with the variable specified in group(*groupvar*) is the default. If *clustvar* is specified as another variable, it must nest the variable specified in group(*groupvar*).

Specifying vce(robust) is equivalent to specifying vce(cluster *clustvar*), where *clustvar* is the variable specified in the group(*groupvar*) option.

Reporting

level(#); see [R] **Estimation options**.

showsupergroups requests that the initial treatment table, printed above the table of estimates, tabulate groups by initial treatment and supergroup, rather than only by initial treatment. If no variable is specified in supergroup(), this option has no effect.

display_options: noci, nopvalues, cformat(%*fmt*), pformat(%*fmt*), sformat(%*fmt*), and nolstretch; see [R] **Estimation options**.

The following option is available with xtswitchdid but is not shown in the dialog box:

coeflegend; see [R] **Estimation options**.

Remarks and examples

Researchers often want to study the effects of a treatment, such as a policy or intervention, on an outcome for a group by exploiting variation in the timing or presence of treatment. In some cases, these treatments are binary, meaning that each group is either treated or untreated at any point in time, and absorbing, meaning that a group that is treated remains treated in subsequent periods.

In many applications, however, the treatments that are of interest are multivalued. For example, a treatment variable may count the number of active interventions in a given jurisdiction and thus take any nonnegative integer value. In the examples below, we take the number of new safety rules in a minor hockey league, ranging from 0 to 3, as a treatment.

Treatments can also be switching rather than absorbing. A treated group, after entering treatment, may transition to a higher level of treatment or lower level of treatment or exit treatment altogether. A hockey league may introduce three new safety rules in one year but repeal one of them two years later.

Furthermore, groups need not be first observed as untreated. A hockey league may have always had one safety rule but within the period of our study decide to transition to two safety rules—or vice versa.

`xtswitchdid` estimates event-study effects for panel data using the heterogeneous DID estimator of de Chaisemartin and D'Haultfœuille (2026), which allows for multivalued and switching treatment variables and does not require that groups be first observed at treatment level 0.

These event-study effects are estimated at each of multiple exposures, where the ℓth effect is associated with having experienced a first treatment change ℓ periods ago. The effects can be compared with the dynamic treatment effects that are estimated by `estat aggregation, dynamic` after `xthdidregress`. They cannot, however, be given the same interpretation as a simple aggregation of average treatment effects on the treated (ATETs). In the setting of `xthdidregress`, treatments are binary and absorbing, so groups are distinguished from each other only by the timing of their treatment. In the more general setting of `xtswitchdid`, we must also consider the magnitude of treatments, the timing of switches between multiple treatment levels, and the initial level of treatment.

With `xtswitchdid`, the potential outcomes in the de Chaisemartin and D'Haultfœuille (2026) framework are defined for paths of treatment for groups. In contrast, `xtdidregress` estimates homogeneous treatment effects defined in terms of the potential outcome for group g being treated $Y_g(1)$ and the potential outcome for group g being untreated $Y_g(0)$. Also different is `xthdidregress`, which estimates heterogeneous treatment effects defined in terms of potential outcomes that are functions of treatment cohort, where treatment cohort refers to the point in time when a group is first exposed to treatment.

To illustrate, suppose treatment is binary and there are three time periods. We could consider the following potential outcomes: $Y_g(0,0,0)$, $Y_g(0,1,0)$, and $Y_g(0,1,1)$. Here $Y_g(0,0,0)$ is the potential outcome of having been exposed to treatment level 0 in all three periods. $Y_g(0,1,0)$ is the outcome of having been exposed to treatment level 0 in the first and third period and to treatment level 1 in the second period. Similarly, $Y_g(0,1,1)$ is the outcome of having been exposed to treatment level 0 in the first period and to treatment level 1 in the second and third periods. In settings like this, with only a few distinct treatment paths, we could use the `path()` option in `xtswitchdid` to isolate, say, the `0 1 1` treatment path, and estimate the effect $E\{Y_g(0,1,1) - Y_g(0,0,0)\}$. In this simple case, we can interpret the effect as an aggregation of ATETs, which is the same way we would interpret the results from `estat aggregation, dynamic` after `xthdidregress`.

In general, however, there can be too many paths and too little data for each path for this to be feasible. By default, `xtswitchdid` aggregates over groups that have the same initial treatment value, even if they have different treatment paths. For example, in the case above, we would aggregate the effects for the groups exposed to treatment path `0 1 0` together with those for the groups exposed to treatment path `0 1 1`, even though `0 1 0` is only "weakly stronger" as a treatment than the initial treatment value of 0.

In this way, we concern ourselves mainly with each group's "first switch" away from its initial treatment, while allowing the timing of the first switch to be staggered across groups. Accordingly, we can think of the treatment designs where `xtswitchdid` is applicable as staggered, first-switch designs.

For each group, we consider the effect of its observed treatment path relative to the counterfactual path where it had remained at its initial treatment value. In the simple case described above, the counterfactual path is `0 0 0`. Thus, we aggregate over the effects $E\{Y_g(0,1,0) - Y_g(0,0,0)\}$ and $E\{Y_g(0,1,1) - Y_g(0,0,0)\}$. Because we do not observe the counterfactual path for any group g, we rely on the presence of control groups—groups that start at the same initial value as the group of interest but that have experienced no treatment switch over the relevant period of time.

The ℓth event-study effect estimated by `xtswitchdid` can then be thought of as the average effect of being exposed to a treatment weakly stronger than a group's initial treatment state after ℓ periods, where the average is conditional on the data. By default, the estimated effects are normalized by the incremental treatment dose across the ℓ periods starting with the period where the treatment is first switched, facilitating comparisons across exposures. This normalization also ensures that the ℓth effect is a weighted average of the effects of the current treatment and the first through $\ell-1$th treatment lags on the outcome. For example, consider a group with a baseline treatment of 1, which experiences the path `1 3 2`. Here the first event-study effect, for $\ell = 1$, would be normalized by a constant of $(3-1) = 2$, because treatment in the first switched period, 3, is higher than the baseline treatment by 2. The second event-study effect would be normalized by a constant of $(3-1)+(2-1) = 3$ and would consist of a weighted average of the effect of the current treatment, weighted by $(2-1)/3 = 1/3$, and the effect of the first lagged treatment, weighted by $(3-1)/3 = 2/3$.

Because some groups switch "in", meaning their first treatment switch is to a level higher than their initial treatment level, and some groups switch "out", meaning their first treatment switch is to a level lower than their initial treatment level, the group-specific effect of a treatment switch is multiplied by -1 for switchers-out before aggregating across groups. In other words, we assume that a unit increase in the treatment variable has the opposite effect of a unit decrease in the treatment variable.

When aggregating to estimate effects, `xtswitchdid` excludes observations for groups starting from time periods where the group has experienced treatment levels both higher and lower than its initial treatment levels. In other words, a group's treatment paths can freely vary while remaining weakly higher than the initial treatment level (switchers-in) or freely vary at levels weakly lower than the initial treatment level (switchers-out), but it must not do both.

Data must be `xtset` before using `xtswitchdid`. When treatments and outcomes are observed periodically (such as every four years), the `delta()` option in `xtset` can be used to specify the relevant period.

For panel-data applications where treatments are binary and absorbing, `xthdidregress` can be used; see [CAUSAL] **xthdidregress**. When, in addition, treatment effects are thought not to vary across time or treatment cohort, `xtdidregress` can be used; see [CAUSAL] **didregress**.

▷ Example 1: Estimating event-study effects

We have fictional data on hockey teams in 30 minor hockey leagues in North America and are interested in the effect of the number of new safety rules, `safety`, on the number of player injury days, `injdays`, per team each season, normalized by season length. We hypothesize that exposure to these safety rules reduces `injdays`. Each safety rule is a rule change made at the league level. For example, one rule may require mouthguards, and another rule may set a minimum age for body checking.

Our panel has 462 teams across the 30 leagues, each of which is observed for some of or all the years between 2010 and 2025, with the exception of 2020, a pandemic year where outcomes and covariates are missing for all teams. We also observe the region in which each league is located, `region`, the team win percentages each season transformed by the logit function, `lwinpct`, and a measure of the intensity of

each team's schedule, `schedule`. `schedule` is a function of both the number of games played per season and a subjective measure of the competitiveness of each team's schedule and takes values between 1 and 6.

We can load our data and tabulate the treatment values by year:

```
. use https://www.stata-press.com/data/r19/hockey
(Fictional minor hockey team injury data)
. tabulate year safety
```

	Number of safety measures				
Year	0	1	2	3	Total
2010	213	121	41	0	375
2011	215	122	42	0	379
2012	43	30	306	0	379
2013	96	48	314	0	458
2014	100	34	328	0	462
2015	100	34	328	0	462
2016	100	34	328	0	462
2017	156	34	261	11	462
2018	156	34	261	11	462
2019	156	34	261	11	462
2020	141	34	233	11	419
2021	141	34	233	11	419
2022	123	34	85	177	419
2023	123	34	85	177	419
2024	123	34	85	177	419
2025	123	34	85	177	419
Total	2,109	729	3,276	763	6,877

`safety` takes four distinct values in the data: 0, 1, 2, and 3. In 2010, the first year of our panel, we observed 375 hockey teams, of which 213 had implemented no safety rules, 121 had implemented 1 safety rule, and 41 had implemented 2 safety rules. In subsequent years, leagues generally moved to higher levels of the treatment. Not easily observable here is that some groups in fact moved to lower levels of the treatment in some years. For example, we see that between 2016 and 2017, the number of teams implementing 0 safety rules rose from 100 to 156. We will use other tools to examine treatment paths in more detail.

Before obtaining event-study effects, we first `xtset` our data, with `team` as the panel variable and `year` as the time variable:

```
. xtset team year
Panel variable: team (unbalanced)
 Time variable: year, 2010 to 2025
         Delta: 1 unit
```

To obtain estimates of event-study effects using xtswitchdid, we use syntax similar to that of xthdidregress and other DID commands. Our outcome variable, injdays, is specified in the first set of parentheses, along with our covariates, lwinpct and schedule. The second set of parentheses contains the treatment variable, safety. Our panel and time variables are already set, but we must specify the group variable using the group() option. Here the group variable is the same as the panel variable, which in this case is team.

```
. xtswitchdid (injdays lwinpct schedule) (safety), group(team)
Treatment and time information
Panel variable:  team
Time variable:   year
Time interval:   2010 thru 2025
Treatment level: team
Control:         safety maintaining baseline
Treatment:       safety changing from baseline
```

Initial treatment	Groups	Obs	Switch out	Switch in	Absorbing
0	272	3,749		1 2 3	No
1	140	1,956		2 3	No
2	50	723		3	Yes

```
DID with switching treatments                 Number of obs    = 6,428
                                              Number of panels =   462
                                              Number of groups =   462
                                 (Std. err. adjusted for 462 clusters in team)
```

	Normalized effect	Robust std. err.	z	P>|z|	[95% conf.	interval]
Exposure						
1	-6.871766	2.679854	-2.56	0.010	-12.12418	-1.619348
2	-2.75066	1.380428	-1.99	0.046	-5.45625	-.0450695
3	-3.061524	.8162018	-3.75	0.000	-4.66125	-1.461798
4	-1.038777	.7065305	-1.47	0.141	-2.423551	.3459975
5	-.9033563	.5854862	-1.54	0.123	-2.050888	.2441756
6	-1.140814	.4573033	-2.49	0.013	-2.037112	-.2445157
7	-.9005077	.4021147	-2.24	0.025	-1.688638	-.1123773
8	-1.239846	.4123365	-3.01	0.003	-2.048011	-.4316814
9	.7121286	.9774432	0.73	0.466	-1.203625	2.627882
10	-1.615785	.3346074	-4.83	0.000	-2.271603	-.9599661
11	-1.405509	.457877	-3.07	0.002	-2.302931	-.5080862
12	-1.613272	.430911	-3.74	0.000	-2.457842	-.7687025
13	-1.213585	.3641668	-3.33	0.001	-1.927339	-.4998312
14	-.3144064	.3203691	-0.98	0.326	-.9423183	.3135055

```
Note: Effects computed using covariates.
```

The header gives information about the model and the treatment variable. The initial treatment table shows that there are three values of the treatment variable for which we observe teams with that value as their initial treatment. This is important because our estimator computes differences only between groups that share an initial treatment value.

Most teams start with zero safety rules implemented, as reflected in the Groups column. The number of observations reflects the fact that each team is observed at multiple time periods. Note that both switchers and groups that never switch, but are used as control groups, are counted in these summary statistics.

In this dataset, all treatment switches away from the initial treatment value are switches "in" to treatment, meaning to a higher level of treatment as opposed to a lower. The Switch in column lists the distinct treatment values that groups in each initial-treatment-value group are observed to eventually take. Among the groups whose initial treatment value is 0, there are teams whose first "switch" is to 1, others whose first switch is to 2, and others whose first switch is to 3.

The Absorbing column indicates whether all teams with a particular initial treatment have a treatment path that is absorbing. The Yes in the third row indicates that all teams with initial treatment 2 have absorbing treatment paths, while the Nos in the first and second rows indicate that there are at least some groups with initial treatment levels 0 and 1 whose treatment paths are not absorbing. Here an absorbing treatment path is one such that when the group has switched from its initial treatment, it does not again change its treatment to another value.

The table of results gives normalized event-study effects as defined in *Methods and formulas*. The effect at exposure 1 is the average effect on injdays of increasing the number of safety rules by one, as measured in the first treated period, relative to a counterfactual where the treatment had remained at the initial number of safety rules. Any team that changed its number of safety rules, whether to a treatment level higher or lower than its initial treatment level, experienced it for at least one period and was thus used to compute this effect.

The effect at exposure 5 is the average effect of additional safety rules five periods after the most recent preswitch period, relative to a counterfactual where the treatment never changed from the initial value. Any team that was observed for five periods starting with its treatment switch was included in the estimation of this effect, provided we have at least one control group with the same initial treatment value. For example, a team that began at zero safety rules and increased to two safety rules in 2012 and was observed until at least 2016 would be included in the computation of the effect for exposure 5.

We may be interested in a summary measure of the effect of safety on injdays, rather than in the effect of a certain exposure length. The event-study effects can be aggregated over exposures to produce an average total effect that we can interpret as the average effect of being exposed to a unit of treatment, conditional on the distribution of exposures we observe. We can compute an average total effect using the estat total postestimation command.

```
. estat total
Average total effect per unit of treatment             Number of obs    = 6,428
                                                       Number of panels =   462
                                                       Number of groups =   462
                                  (Std. err. adjusted for 462 clusters in team)
```

injdays	Effect	Robust std. err.	z	P>|z|	[95% conf.	interval]
safety	-8.65764	2.281467	-3.79	0.000	-13.12923	-4.186046

We see evidence for a negative effect of `safety` on `injdays`, where an additional unit of treatment corresponds to an average of 8.7 fewer player injury days per team in the seasons following the treatment change.

◁

▷ Example 2: Visualizing event-study effects

We can use the `estat eventplot` postestimation command to visualize normalized event-study effects by exposure:

```
. estat eventplot
```

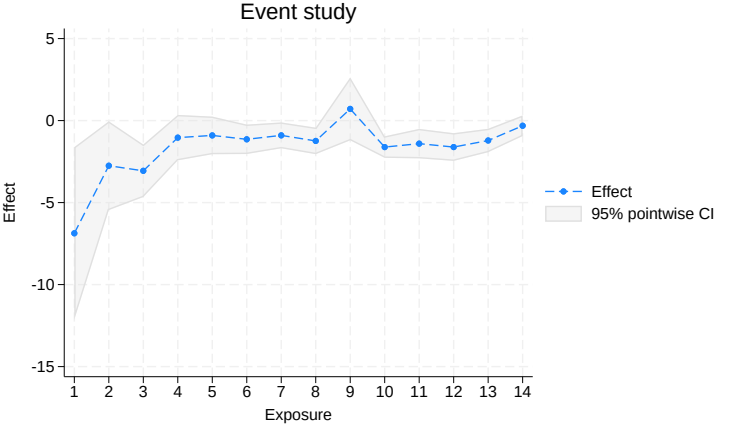

Figure 1. Event-study effects by exposure

In the first year of exposure to a treatment switch, we see an effect that is negative but imprecisely estimated. In exposures 2 through 8, we see an effect that declines in magnitude as teams are exposed to treatment for a longer period—maybe enforcement slips over time, or maybe teams gradually adjust their style of play in response to rule changes.

Note the imprecision in our estimate for the effect at exposure 9. This reflects the fact that, for most groups, the ninth period of exposure to switching from initial treatment is 2020, when no data points were observed.

◁

Example 3: Treatment paths

When treatments are multivalued or switching, interpretation becomes more challenging than in the standard case of binary and absorbing treatments. We may wish to isolate a single treatment path to measure its treatment effect, which can be interpreted directly. The postestimation utility `estat paths` lists the treatment paths observed in the data:

```
. estat paths
note: pathlen() not specified; using path length 14.
note: variable _did_paths, which encodes a unique value for each treatment
      path, was added to the dataset.

Distribution of treatment paths

Path length: 14, not counting initial treatment
Treatment:   team changing initial treatment level
```

Paths	Freq.	Percent	Cum.
0 1 1 2 2 2 2 2 2 . 2 2 2 2 2	14	4.56	4.56
0 2 . 2 2 2 2 2 2	1	0.33	4.89
0 2 2 2 . 2 2 2 2 . 2 3 3 3 3	3	0.98	5.86
0 2 2 2 2 2 0 0 . . 0 0 0 0 0	1	0.33	6.19
0 2 2 2 2 2 0 0 0 . 0 0 0 0 0	55	17.92	24.10
0 2 2 2 2 2 2 2 2 . . 2 3 3 3 3	1	0.33	24.43
0 2 2 2 2 2 2 2 2 . 2 3 3 3 3 .	1	0.33	24.76
0 2 2 2 2 2 2 2 2 2	11	3.58	28.34
0 2 2 2 2 2 2 2 2 2 . . 3 3 3 3	2	0.65	28.99
0 2 2 2 2 2 2 2 2 2 . 2 3 3 3 .	1	0.33	29.32
0 2 2 2 2 2 2 2 2 2 . 2 3 3 3 3	82	26.71	56.03
0 3 3 3 3	18	5.86	61.89
1 2 . 2 2 2 2 2 2 . 2 2 2 2 2	1	0.33	62.21
1 2 2 2 2 2 2 . 2 . 2 2 2 2 2	1	0.33	62.54
1 2 2 2 2 2 2 2 2 2	16	5.21	67.75
1 2 2 2 2 2 2 2 2 2 . 2 . 3 3 3	1	0.33	68.08
1 2 2 2 2 2 2 2 2 2 . 2 2 2 2 2	30	9.77	77.85
1 2 2 2 2 2 2 2 2 2 . 2 3 3 3 3	57	18.57	96.42
2 3 3 3 . 3 3 3 3 3	11	3.58	100.00
Total	307	100.00	

Below the command, a note informs us that a default path length has been set because we have not chosen one using the `pathlen()` option. In cases of long or numerous treatment paths, it may be useful to distinguish paths only up to a smaller number of periods. A second note tells us that a new variable has been created that records the treatment path for each observation.

The output shows us the treatment paths, starting from the last period before the treatment changed. Note specifically that the jth element of the path does not necessarily correspond to the jth period in the data or to any fixed period. It is possible for groups to have the same treatment path despite starting treatment at different times, though that does not occur in this dataset.

Missing values reflect time periods along the path where no treatment is observed or where observations have been dropped because of missing data. Every treatment path in this dataset happens to have at least one missing value corresponding to the year 2020, when no treatments were observed.

We saw earlier that there was at least one treatment path that was not absorbing, and now we can see them in the output. The fourth- and fifth-listed paths show there are 56 teams who switch into treatment (2 safety rules) and then, after 5 periods, switch back out of it.

We may be interested in event-study effects for these treatment paths. We can run `estat paths` with the `generate()` option to obtain path indicators that we can use in the `xtswitchdid` command with the `path()` option. (Alternatively, we could type the path as *numlist*.)

```
. quietly estat paths, generate(safetypath)
. describe safetypath*
```

Variable name	Storage type	Display format	Value label	Variable label
safetypath1	byte	%8.0g		_did_paths==0 1 1 2 2 2 2 2 2 . 2 2 2 2 2
safetypath2	byte	%8.0g		_did_paths==0 2 . 2 2 2 2 2 2 . . . . .
safetypath3	byte	%8.0g		_did_paths==0 2 2 2 . 2 2 2 2 . 2 3 3 3 3
safetypath4	byte	%8.0g		_did_paths==0 2 2 2 2 2 2 0 0 . . 0 0 0 0 0
safetypath5	byte	%8.0g		_did_paths==0 2 2 2 2 2 2 0 0 0 . 0 0 0 0 0
safetypath6	byte	%8.0g		_did_paths==0 2 2 2 2 2 2 2 2 . . 2 3 3 3 3
safetypath7	byte	%8.0g		_did_paths==0 2 2 2 2 2 2 2 2 . 2 3 3 3 3 .
safetypath8	byte	%8.0g		_did_paths==0 2 2 2 2 2 2 2 2 2 . . . . . .
safetypath9	byte	%8.0g		_did_paths==0 2 2 2 2 2 2 2 2 2 . . 3 3 3 3
safetypath10	byte	%8.0g		_did_paths==0 2 2 2 2 2 2 2 2 2 . 2 3 3 3 .
safetypath11	byte	%8.0g		_did_paths==0 2 2 2 2 2 2 2 2 2 . 2 3 3 3 3
safetypath12	byte	%8.0g		_did_paths==0 3 3 3 3 . . . .
safetypath13	byte	%8.0g		_did_paths==1 2 . 2 2 2 2 2 2 . 2 2 2 2 2
safetypath14	byte	%8.0g		_did_paths==1 2 2 2 2 2 2 . 2 . 2 2 2 2 2
safetypath15	byte	%8.0g		_did_paths==1 2 2 2 2 2 2 2 2 2 . . . . .
safetypath16	byte	%8.0g		_did_paths==1 2 2 2 2 2 2 2 2 2 . 2 . 3 3 3
safetypath17	byte	%8.0g		_did_paths==1 2 2 2 2 2 2 2 2 2 . 2 2 2 2 2
safetypath18	byte	%8.0g		_did_paths==1 2 2 2 2 2 2 2 2 2 . 2 3 3 3 3
safetypath19	byte	%8.0g		_did_paths==2 3 3 3 . 3 3 3 3 3 . . . . .

The fifth-listed path is generated as `safetypath5` and corresponds to the path starting with `0 2 2 2 2 2 0 0 0`. We can use this variable to estimate path-specific effects:

```
. xtswitchdid (injdays lwinpct schedule) (safety), group(team) path(safetypath5)
note: 4 groups with only a single available time period were dropped.
note: effect of exposure 9 cannot be estimated because of missing data.
DID with switching treatments                         Number of obs    = 2,222
                                                      Number of panels =   268
                                                      Number of groups =   268
Treatment path: (0 2 2 2 2 2 0 0 0 . 0 0 0 0 0)
Treatment type: Switcher-in, nonabsorbing
                                    (Std. err. adjusted for 268 clusters in team)
------------------------------------------------------------------------------
             | Normalized    Robust
             |    effect   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
    Exposure |
           1 |  -8.803413   3.036455    -2.90   0.004    -14.75476    -2.85207
           2 |  -3.423294   1.351155    -2.53   0.011    -6.071509   -.7750789
           3 |  -2.574106   .9883007    -2.60   0.009     -4.51114   -.6370724
           4 |  -1.090292   .8124545    -1.34   0.180    -2.682673      .50209
           5 |  -.3275366   .6412151    -0.51   0.609    -1.584295    .9292218
           6 |  -1.138047   .5870595    -1.94   0.053    -2.288663    .0125683
           7 |   -1.17035   .5815419    -2.01   0.044    -2.310152   -.0305493
           8 |  -1.320975   .5834758    -2.26   0.024    -2.464567   -.1773839
           9 |           0  (omitted)
          10 |  -.9877028   .5803887    -1.70   0.089    -2.125244    .1498381
          11 |  -.9415612    .927531    -1.02   0.310    -2.759489    .8763662
          12 |  -1.390374   .9659295    -1.44   0.150    -3.283561    .5028133
          13 |  -.3080592   .9232056    -0.33   0.739    -2.117509    1.501391
          14 |  -.8733448   .8459017    -1.03   0.302    -2.531282    .7845921
------------------------------------------------------------------------------
Note: Effects computed using covariates.
```

The output now omits the first table about initial treatment groups because we are looking only at one path. The header displays the path, informing us that groups with this treatment path are switchers-in and that the treatment is nonabsorbing.

For teams that follow this path, switching from 0 to 2 and then back to 0 after 5 periods, we expect an effect of 8.8 fewer injury days in the first period after the initial switch, 3.4 fewer injury days in the second period, and so on. The event-study effect for 9 periods of exposure is missing. This is again because all groups with this treatment path would have experienced that level of exposure in 2020, which we did not observe.

◁

▷ Example 4: Placebo estimators

We can use placebo estimators, as proposed in de Chaisemartin and D'Haultfœuille (2026), to test the assumptions of parallel trends and no anticipation. In our example, parallel trends is the assumption that, in the absence of safety rules, injuries for teams who implemented safety rules would have followed the same trend (growth, decline, or no change) as injuries for teams who did not implement safety rules. No anticipation is the assumption that safety rules have no effect on injuries in the years prior to their introduction; that is, teams do not change their injury rates (that is, their behavior) in anticipation of safety rules that will soon be introduced. See de Chaisemartin and D'Haultfœuille (2026) for details about the assumptions.

In this context, a placebo estimator is one that takes the same form as the estimator of interest (in this case, a normalized event-study effect estimator) but that measures the effect over a period when no treatment has occurred. Where the event-study effect estimator for exposure ℓ measures the difference between a period t and a later period $t+\ell$, the corresponding placebo estimator measures the difference between the same period t and an earlier period $t-\ell$.

We repeat the `xtswitchdid` command from example 1 and test the assumptions of parallel trends and no anticipation by using `estat ptrends`.

```
. quietly xtswitchdid (injdays lwinpct schedule) (safety), group(team)
. estat ptrends
Placebo estimates                                   Number of obs    = 6,428
                                                    Number of panels =   462
                                                    Number of groups =   462
                                (Std. err. adjusted for 462 clusters in team)
------------------------------------------------------------------------------
             | Normalized    Robust
     Placebo |     effect   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
          -6 |   1.181488   2.081433     0.57   0.570    -2.898045    5.261022
          -5 |   .4904605   2.581445     0.19   0.849    -4.569079        5.55
          -4 |  -.4558745   .6090937    -0.75   0.454    -1.649676    .7379272
          -3 |   1.016594   .7035227     1.45   0.148    -.3622855    2.395473
          -2 |  -.3029398    1.40519    -0.22   0.829    -3.057062    2.451182
          -1 |  -4.016654   3.185429    -1.26   0.207    -10.25998    2.226673
------------------------------------------------------------------------------

Test of parallel-trends and no-anticipation assumptions
H0: Placebos are zero for all exposures
    chi2(6) =   6.55
Prob > chi2 = 0.3647
```

The placebo estimate labeled `-6` is the difference between periods t and $t-6$. Similarly, the estimate labeled `-5` is the difference between periods t and $t-5$. The remaining placebo effects are interpreted similarly.

Only six placebo estimates are computed, reflecting the fact that most groups have only a small number of pretreatment periods available to observe. Note that placebo estimates are computed using only groups that actually experience switches of the corresponding exposure. For this reason, `estat ptrends` will not display more placebo estimates than there are event-study estimates in the main output.

We can plot the placebo estimates by exposure using the `graph` option. We run the command `quietly` to suppress the tabular output shown above.

```
. quietly estat ptrends, graph
```

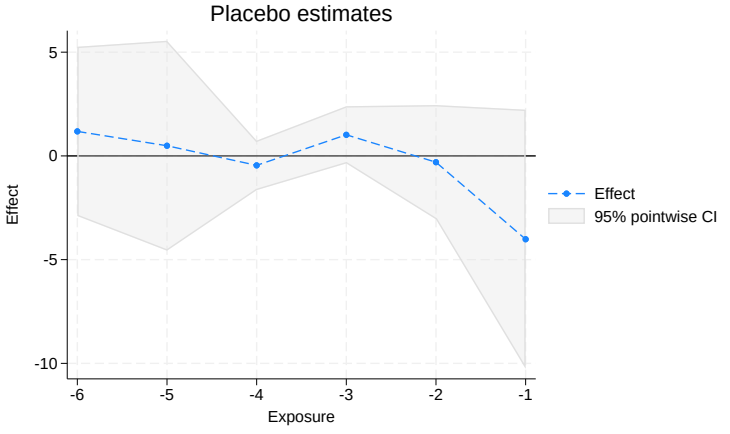

Figure 2. Placebo estimates by exposure

For estimation to be valid, we rely on a parallel-trends assumption, as described in de Chaisemartin and D'Haultfœuille (2026). All confidence intervals include zero. This plot gives evidence that, prior to treatment switches, there was no systematic difference between teams that experienced switches and those that did not. In other words, we have evidence that treated and controls were following the same trend prior to the intervention.

We can plot the main effect estimates on the same plot as our placebo estimates by adding the `alleffects` option:

```
. quietly estat ptrends, graph alleffects
```

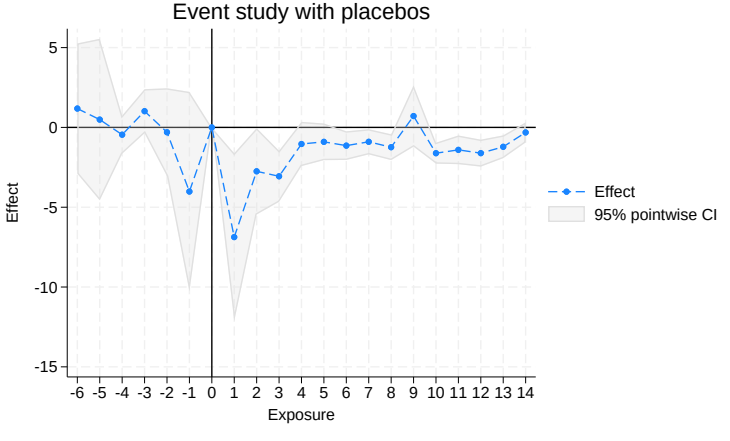

Figure 1. Event-study effect and placebo estimates by exposure

Example 5: Working with coarser treatment groups

While we have team-level data, treatment in our dataset takes place at the league level. We have a relatively small number of leagues (30), but we may want to cluster our standard errors by league:

```
. xtswitchdid (injdays lwinpct schedule) (safety), group(team)
> vce(cluster league)

Treatment and time information

Panel variable:  team
Time variable:   year
Time interval:   2010 thru 2025

Treatment level: team
Control:         safety maintaining baseline
Treatment:       safety changing from baseline
```

Initial treatment	Groups	Obs	Switch out	Switch in	Absorbing
0	272	3,749		1 2 3	No
1	140	1,956		2 3	No
2	50	723		3	Yes

DID with switching treatments

Number of obs = 6,428
Number of panels = 462
Number of groups = 462

(Std. err. adjusted for 30 clusters in league)

	Normalized effect	Robust std. err.	z	P>|z|	[95% conf.	interval]
Exposure						
1	-6.871766	2.731872	-2.52	0.012	-12.22614	-1.517395
2	-2.75066	.9620966	-2.86	0.004	-4.636334	-.864985
3	-3.061524	1.264431	-2.42	0.015	-5.539762	-.5832855
4	-1.038777	.6633062	-1.57	0.117	-2.338833	.2612794
5	-.9033563	.7267021	-1.24	0.214	-2.327666	.5209537
6	-1.140814	.3921218	-2.91	0.004	-1.909358	-.3722692
7	-.9005077	.213569	-4.22	0.000	-1.319095	-.4819201
8	-1.239846	.2984216	-4.15	0.000	-1.824742	-.6549505
9	.7121286	.8880673	0.80	0.423	-1.028451	2.452708
10	-1.615785	.6452509	-2.50	0.012	-2.880453	-.3511161
11	-1.405509	1.287507	-1.09	0.275	-3.928976	1.117959
12	-1.613272	1.357707	-1.19	0.235	-4.27433	1.047785
13	-1.213585	.9524794	-1.27	0.203	-3.08041	.6532404
14	-.3144064	.3302621	-0.95	0.341	-.9617082	.3328954

Note: Effects computed using covariates.

Compared with the results from clustering at the team level in example 1, some of our event-study effects are now estimated more precisely, while others show larger standard errors.

Our panel variable is team, and until now, we have also taken team as our group variable. However, xtswitchdid also allows for treatment groups that are coarser than panels, provided treatment is fixed within each group. In such cases, xtswitchdid first transforms outcomes and covariates to means within each group before estimation and weights each group by the number of panels within it. We specify group(league) to indicate that treatment was applied at the league level.

```
. xtswitchdid (injdays lwinpct schedule) (safety), group(league)

Treatment and time information

Panel variable:  team
Time variable:   year
Time interval:   2010 thru 2025

Treatment level: league
Control:         safety maintaining baseline
Treatment:       safety changing from baseline
```

Initial treatment	Groups	Obs	Switch out	Switch in	Absorbing
0	17	236		1 2 3	No
1	9	127		2 3	No
2	4	57		3	Yes

```
DID with switching treatments                   Number of obs    = 6,428
                                                Number of panels =   462
                                                Number of groups =    30
                                  (Std. err. adjusted for 30 clusters in league)
```

	Normalized effect	Robust std. err.	z	P>|z|	[95% conf.	interval]
Exposure						
1	-5.272442	2.456315	-2.15	0.032	-10.08673	-.4581535
2	-2.330813	.8055745	-2.89	0.004	-3.90971	-.7519158
3	-2.911389	1.066263	-2.73	0.006	-5.001226	-.8215528
4	-1.00979	.6159314	-1.64	0.101	-2.216994	.1974131
5	-.7906042	.6420756	-1.23	0.218	-2.049049	.4678408
6	-.9883763	.4369299	-2.26	0.024	-1.844743	-.1320093
7	-.7840238	.3195645	-2.45	0.014	-1.410359	-.157689
8	-.8602774	.4150936	-2.07	0.038	-1.673846	-.046709
9	.775594	.848628	0.91	0.361	-.8876864	2.438874
10	-1.339494	.4164269	-3.22	0.001	-2.155676	-.5233121
11	-1.249448	1.324925	-0.94	0.346	-3.846253	1.347358
12	-1.611414	1.386344	-1.16	0.245	-4.328599	1.105771
13	-1.067826	.9374129	-1.14	0.255	-2.905122	.7694693
14	-.2887542	.383987	-0.75	0.452	-1.041355	.4638464

```
Note: Effects computed using covariates.
```

Most of the resulting estimated effects are smaller in absolute value than those estimated with group(team). In general, effects are not equivalent across different specifications of the treatment group variable. Specifying a treatment group coarser than the panel variable may be appropriate when the parallel-trends assumption applied at the panel level is thought to be too strong.

◁

Example 6: Supergroups

By default, switchers are compared with control groups that share an initial treatment value when estimating event-study effects by DID. For example, switchers who start at treatment value 0 are compared only with control groups that also started at treatment value 0 (and either never switched or switched later than the switcher in question).

We may wish to further restrict the set of control groups. In our hockey dataset, we have a variable, `region`, that can be used for this purpose. By using the `supergroup()` option, we can require that switchers be compared with only control groups that are in the same region as the switcher, in addition to sharing the initial treatment value.

```
. xtswitchdid (injdays lwinpct schedule) (safety), group(team)
> supergroup(region)
note: 4 initial treatment and supergroup combinations dropped because of no
      switchers.
note: 3 initial treatment and supergroup combinations dropped because of no
      controls.
note: 1 initial treatment and supergroup combination dropped because of
      insufficient data.
note: effect of exposure 9 cannot be estimated because of missing data.

Treatment and time information

Panel variable:  team
Time variable:   year
Time interval:   2010 thru 2025

Supergroup:      region
Treatment level: team
Control:         safety maintaining baseline
Treatment:       safety changing from baseline
```

Initial treatment	Groups	Obs	Switch out	Switch in	Absorbing
0	204	2,198		1 2	No

```
DID with switching treatments                  Number of obs         = 2,198
                                               Number of panels      =   204
                                               Number of groups      =   204
                                               Number of supergroups =     3
                                 (Std. err. adjusted for 204 clusters in team)
```

	Normalized effect	Robust std. err.	z	P>|z|	[95% conf.	interval]
Exposure						
1	-6.155398	2.743241	-2.24	0.025	-11.53205	-.7787444
2	-2.284206	1.505154	-1.52	0.129	-5.234254	.6658426
3	-2.285472	1.047505	-2.18	0.029	-4.338544	-.2323989
4	-.8643861	.8303966	-1.04	0.298	-2.491934	.7631612
5	-1.110083	.6256107	-1.77	0.076	-2.336258	.1160909
6	-1.896084	.5303533	-3.58	0.000	-2.935557	-.8566108
7	-1.172279	.5187632	-2.26	0.024	-2.189036	-.1555216
8	-1.248924	.4757709	-2.63	0.009	-2.181418	-.3164301
9	0	(omitted)				
10	-1.792667	.4404417	-4.07	0.000	-2.655917	-.9294172

```
Note: Effects computed using covariates.
```

Because we are segmenting the data more finely, some combinations have been dropped, because there were no switchers, no controls, or both in the given subset of the data (see the notes at the top of the output). Furthermore, we can now estimate event-study effects only up to 10 years of exposure.

The `supergroup()` option is useful because it allows for weakening the standard parallel-trends assumption to hold only within specified supergroups, as discussed in de Chaisemartin et al. (2025). Here the event-study effects we find using `region` as a supergroup are slightly smaller in absolute value than in previous examples, which may suggest that the default assumption was too restrictive.

To see detailed information about supergroups in the first table, we use the `showsupergroups` option.

```
. xtswitchdid (injdays lwinpct schedule) (safety), group(team)
> supergroup(region) showsupergroups
note: 4 initial treatment and supergroup combinations dropped because of no
      switchers.
note: 3 initial treatment and supergroup combinations dropped because of no
      controls.
note: 1 initial treatment and supergroup combination dropped because of
      insufficient data.
note: effect of exposure 9 cannot be estimated because of missing data.

Treatment and time information

Panel variable:  team
Time variable:   year
Time interval:   2010 thru 2025

Supergroup:      region
Treatment level: team
Control:         safety maintaining baseline
Treatment:       safety changing from baseline
```

Initial treatment / supergroup	Groups	Obs	Switch out	Switch in	Absorbing
0 / 2	59	643		1 2	No
0 / 3	123	1,287		2	No
0 / 4	22	268		2	Yes

```
DID with switching treatments                 Number of obs         = 2,198
                                              Number of panels      =   204
                                              Number of groups      =   204
                                              Number of supergroups =     3
                                 (Std. err. adjusted for 204 clusters in team)
```

	Normalized effect	Robust std. err.	z	P>|z|	[95% conf.	interval]
Exposure						
1	-6.155398	2.743241	-2.24	0.025	-11.53205	-.7787444
2	-2.284206	1.505154	-1.52	0.129	-5.234254	.6658426
3	-2.285472	1.047505	-2.18	0.029	-4.338544	-.2323989
4	-.8643861	.8303966	-1.04	0.298	-2.491934	.7631612
5	-1.110083	.6256107	-1.77	0.076	-2.336258	.1160909
6	-1.896084	.5303533	-3.58	0.000	-2.935557	-.8566108
7	-1.172279	.5187632	-2.26	0.024	-2.189036	-.1555216
8	-1.248924	.4757709	-2.63	0.009	-2.181418	-.3164301
9	0	(omitted)				
10	-1.792667	.4404417	-4.07	0.000	-2.655917	-.9294172

```
Note: Effects computed using covariates.
```

Notice that the header table now gives information on initial treatment and supergroup combinations rather than only on initial treatment values.

◁

▷ Example 7: Common set of switchers

By default, `xtswitchdid` uses all switchers with available differences at each exposure to compute event-study effects. For example, if a given group has an observed treatment path of `0 0 1` and thus is observed for only one year starting from its first treatment switch, that team will be used to estimate the effect of exposure 1, even though it cannot be used for the effects of exposures 2 or higher.

If the `commonswitchers` option is specified, however, `xtswitchdid` will instead require that a common set of switchers be used for all exposures. If there are four exposures to be computed, `xtswitchdid` with `commonswitchers` will include only those switchers for which there are at least four periods since exposure to a switch. The group with the path `0 0 1` will be excluded from all effect estimation, as will any group observed for fewer than 4 periods starting with the first treatment switch. This may improve interpretability across exposures, but it will also tend to give less precise results as fewer groups are used to compute event-study effects, especially at smaller exposures for which there would otherwise be many groups available.

Note that while the `commonswitchers` option restricts the groups that are used as switchers for each exposure, it does not restrict which groups can be used as controls and in particular does not enforce a common set of control groups.

Here we estimate event-study effects of safety on injdays using the commonswitchers option. There are effects for 14 exposures, which means that only switchers observed for 14 periods after a switch are included in estimation.

```
. xtswitchdid (injdays lwinpct schedule) (safety), group(team)
> commonswitchers
note: 1 initial treatment value dropped because of no switchers in common set.
note: effect of exposure 9 cannot be estimated because of missing data.

Treatment and time information

Panel variable:  team
Time variable:   year
Time interval:   2010 thru 2025

Treatment level: team
Control:         safety maintaining baseline
Treatment:       safety changing from baseline
```

Initial treatment	Groups	Obs	Switch out	Switch in	Absorbing
0	272	3,749		1 2 3	No
1	140	1,956		2 3	No

DID with switching treatments — Number of obs = 5,705; Number of panels = 412; Number of groups = 412

(Std. err. adjusted for 412 clusters in team)

	Normalized effect	Robust std. err.	z	P>|z|	[95% conf.	interval]
Exposure						
1	-6.379932	3.026874	-2.11	0.035	-12.3125	-.4473692
2	-2.725799	1.54051	-1.77	0.077	-5.745143	.2935439
3	-3.271019	.9280748	-3.52	0.000	-5.090012	-1.452026
4	-1.152499	.775865	-1.49	0.137	-2.673166	.3681689
5	-.9635046	.5899073	-1.63	0.102	-2.119702	.1926925
6	-1.187439	.4627569	-2.57	0.010	-2.094426	-.2804524
7	-.9359829	.4124152	-2.27	0.023	-1.744302	-.1276639
8	-1.275036	.4203548	-3.03	0.002	-2.098917	-.4511563
9	0	(omitted)				
10	-1.604533	.3351186	-4.79	0.000	-2.261353	-.9477123
11	-1.386951	.4582902	-3.03	0.002	-2.285183	-.4887183
12	-1.623364	.4316211	-3.76	0.000	-2.469326	-.7774026
13	-1.217649	.3648081	-3.34	0.001	-1.932659	-.502638
14	-.3144064	.3203692	-0.98	0.326	-.9423185	.3135057

```
Notes: Effects computed using covariates.
       All effects estimated using a common set of switchers.
```

In the notes above the output, we are informed that an initial treatment group has been dropped. Groups for which the initial treatment level is 2 are not observed for 14 periods since switching, as would be required to be included in the common set of switchers. Specifically, the path 2 3 3 3 3 . 3 3 3 3 3 . . . . ., which is observed for 10 periods starting with the treatment switch, is excluded. The absence of groups with initial treatment 2 is reflected in the initial treatment table.

xtswitchdid also has an neffects() option, which specifies a maximum number of event-study effects to compute. Typically, this option is only cosmetic, because each effect is computed separately. But when commonswitchers is specified, specifying neffects() determines for which ex-

posures the common set of switchers must be valid. This is an important consideration because imposing `commonswitchers` in combination with a large value of `neffects()` can amount to discarding most of the switchers in the dataset. In this example, we may want to impose a maximum of five effects:

```
. xtswitchdid (injdays lwinpct schedule) (safety), group(team)
> commonswitchers neffects(5)

Treatment and time information

Panel variable:  team
Time variable:   year
Time interval:   2010 thru 2025

Treatment level: team
Control:         safety maintaining baseline
Treatment:       safety changing from baseline
```

Initial treatment	Groups	Obs	Switch out	Switch in	Absorbing
0	272	3,749		1 2 3	No
1	140	1,956		2 3	No
2	50	723		3	Yes

```
DID with switching treatments                   Number of obs    = 6,428
                                                Number of panels =   462
                                                Number of groups =   462
                                (Std. err. adjusted for 462 clusters in team)
```

	Normalized effect	Robust std. err.	z	P>|z|	[95% conf.	interval]
Exposure						
1	-6.291381	3.01167	-2.09	0.037	-12.19414	-.3886173
2	-2.375	1.540421	-1.54	0.123	-5.394169	.6441702
3	-2.954174	.9135166	-3.23	0.001	-4.744633	-1.163714
4	-1.064917	.787923	-1.35	0.177	-2.609217	.479384
5	-.9033563	.5854862	-1.54	0.123	-2.050888	.2441756

```
Notes: Effects computed using covariates.
       All effects estimated using a common set of switchers.
```

Note that initial treatment level 2 is no longer dropped and results are different from the case where `neffects()` was not specified.

◁

Stored results

xtswitchdid stores the following in e():

Scalars

e(N)	number of observations
e(N_clust)	number of clusters
e(N_g)	number of groups
e(N_exposure)	number of periods of exposure
e(N_panels)	number of panels
e(N_supergroups)	number of supergroups
e(neffects)	number of effects
e(tmin)	first time period
e(tmax)	last time period

Macros

e(cmd)	xtswitchdid
e(cmdline)	command as typed
e(depvar)	name of dependent variable
e(clustvar)	name of cluster variable
e(controlgroup)	control group
e(switchgroup)	switcher group
e(path)	name of path variable or *numlist* as specified
e(path_num)	treatment path
e(path_switch)	path switcher group
e(path_absorb)	path absorbing or nonabsorbing
e(supergroupvar)	name of supergroup variable
e(showsupergroups)	showsupergroups, if specified
e(raw)	raw, if specified
e(commonswitchers)	commonswitchers, if specified
e(wtype)	weight type
e(wexp)	weight expression
e(marginsnotok)	predictions disallowed by margins
e(listwise)	listwise or nolistwise
e(ivar)	name of panel variable
e(timevar)	name of time variable
e(groupvar)	name of group variable
e(treatment)	name of treatment variable
e(estat_cmd)	program used to implement estat
e(vce)	*vcetype* specified in vce()
e(vcetype)	title used to label Std. err.
e(properties)	b V

Matrices

e(b)	coefficient vector
e(V)	variance–covariance matrix of the estimators

Functions

e(sample)	marks estimation sample

In addition to the above, the following is stored in r():

Matrices

r(table)	matrix containing the coefficients with their standard errors, test statistics, *p*-values, and confidence intervals

Note that results stored in r() are updated when the command is replayed and will be replaced when any r-class command is run after the estimation command.

Methods and formulas

Methods and formulas are presented under the following headings:

Overview

We observe data $\{y_{gt}, d_{gt}\}$ for groups $g = 1, \ldots, G$, at periods $t = 1, \ldots, T$, where y_{gt} is the observed outcome and d_{gt} is the treatment variable. We let n_{gt} be the number of panels within each group. If $n_{gt} > 1$ so that the dataset is not aggregated to the group level, outcomes are transformed to group-level means. Treatments are required to be constant within group-time cells and are also collapsed to the group level. There may also be covariates $\{\mathbf{x}_{gt}\}$, but, for now, we will deal with the case in which there are no covariates.

Denote $y_{gt}(\mathbf{d})$ as the potential outcome for group g at time t if it experienced the treatment path given by the T-length vector $\mathbf{d}$. Let $\mathbf{d}_g = (d_{g1}, \ldots, d_{gT})$, and let $\mathbf{D} = (\mathbf{d}_1, \ldots, \mathbf{d}_G)$ be the vector of observed treatment paths across all groups.

We do not require initial treatment values d_{g1} to be zero or any other particular or common value. Groups for which treatment changes at least once are called switchers. If the first treatment change moves the treatment to a value higher than the initial treatment value, we call the group a switcher-in, and we call it a switcher-out if the first treatment change moves it to a lower value. s_g takes the value 1 if group g is a switcher-in, -1 if it is a switcher-out, and undefined otherwise. By default, all switchers are included when computing event-study effects. When the `switchgroup(in)` option is specified, only switchers with $s_g = 1$ are included, and only switchers with $s_g = -1$ are included when `switchgroup(out)` is specified.

In a group for which treatment both rises above and falls below the group's initial treatment value at some point in the observed treatment path for that group, s_g is undefined. Such groups are excluded from the estimation sample.

We let t_g denote the last period for which group g's treatment has not yet changed from d_{g1}, its initial treatment value. We let $T_g \geq t_g$ denote the last time period for which there exists some group g' for which $d_{g'1} = d_{g1}$ and $t_g < t_{g'} = T_g$. In other words, there exists a group that can serve as a control for group g if and only if $t_g < T_g$.

We let $\ell = 1, \ldots, L$ index the periods of exposure to treatment, relative to the last untreated period $t_g - 1$, where L is equal to the maximum value of $T_g - t_g$ across groups. n_ℓ is the number of groups for which $t_g + \ell \leq T_g$, meaning that control groups are available to estimate the ℓth event-study effect.

Our main estimand of interest is the following normalized event-study effect,

$$\delta^n_\ell = \frac{1}{\delta^d_\ell} \sum_{g:t_g+\ell \leq T_g} s_g n_{g,t_g+\ell} E\{y_{g,t_g+\ell} - y_{g,t_g+\ell}(d_{g,1}, \ldots, d_{g,1}) | \mathbf{D}\}$$

where $\delta^d_{g,\ell} = \sum_{k=0}^{\ell-1}(d_{g,t_g+k} - d_{g,1})$ is the total "dose" of treatment received by a group g from t_g until $t_g + \ell$ and $\delta^d_\ell = \sum_{g:t_g+\ell \leq T_g} |\delta^d_{g,\ell}|$ is the total of the absolute values of such doses over each group for which the ℓth event-study effect can be estimated.

This event-study effect aggregates effects across groups conditional on the observed treatment data. For each group, we are concerned with the difference between the actual treatment path the group experienced up to period $t_g - 1 + \ell$ and a counterfactual path where the group remained at their initial treatment level.

We cannot observe the counterfactual object $y_{g,t_g+\ell}(d_{g1}, \ldots, d_{g1})$ for switchers and thus estimate these effects by the principle of difference in differences, where the differences in outcomes for switchers before and after a treatment change and the differences in outcomes that occurred at the same time for comparable never-switchers or not-yet-switchers are themselves differenced to generate a plausible estimate of the difference of interest. The estimator is given by

$$\hat{\delta}^n_\ell = \frac{1}{\delta^d_\ell} \sum_{g:t_g+\ell \leq T_g} s_g \left\{ n_{g,t_g+\ell}(y_{g,t_g+\ell} - y_{g,t_g}) - \frac{1}{n^g_{t_g+\ell}} \sum_{\substack{g': d_{g'1}=d_{g1} \\ t_{g'}>t_g+\ell}} n_{g',t_g+\ell}(y_{g',t_g+\ell} - y_{g',t_g}) \right\}$$

where for any period t, $n^g_t = \sum_{g'} \mathbf{1}(d_{g'1} = d_{g1}, t_{g'} + 1 > t)$.

When the `controlgroup(never)` option is specified, the control groups for group g are defined the same way, except that groups whose treatments ever change from their initial treatments are excluded as controls.

The raw event-study effect δ_ℓ and its estimator $\hat{\delta}_\ell$ are identical to their normalized equivalents with δ^d_ℓ replaced by n_ℓ.

Standard errors

By default, the analytic standard errors given in de Chaisemartin and D'Haultfœuille (2026) and its companion technical documentation (de Chaisemartin et al. 2025) are computed. Note that these standard errors can be conservative, particularly when there are few groups associated with each initial treatment level.

For the purpose of computing standard errors, writing the event-study estimator as an aggregation over group-specific quantities, without secondary sums over all other groups g', is convenient. We can write $\hat{\delta}^n_\ell = (1/\delta^d_\ell) \sum_{g=1}^G \mathbf{1}(\ell \leq T_g - 1)\widehat{\Delta}_{g,l}$, where

$$\widehat{\Delta}_{g,l} = \frac{1}{n_\ell} \sum_{t=\ell+1}^{T_g} \left\{ \frac{n^{\text{in}}_{gt\ell} - n^{\text{out}}_{gt\ell}}{n^g_t} \mathbf{1}(t_g + 1 > t) + s_g \mathbf{1}(t_g + \ell = t) \right\} n_{gt}(y_{gt} - y_{gt-\ell})$$

Here

$$n^{\text{in}}_{gt\ell} = \sum_{\substack{s_{g'}=1 \\ t_{g'}+\ell=t \\ g': d_{g'1}=d_{g1} \\ n^{g'}_t>0}} n_{g't}$$

and $n^{\text{out}}_{gt\ell}$ is defined equivalently with the condition $s_{g'} = -1$. An estimator of the variance of the estimator is constructed similarly using the sum of squared deviations from estimated means. We first define

$$\widehat{\Delta}^{V,\text{in}}_{g,l} = \frac{1}{n_\ell} \sum_{t=\ell+1}^{T_g} \left\{ \mathbf{1}(t_g + \ell = t) - \frac{n^{\text{in}}_{gt\ell}}{n^g_t} \mathbf{1}(t_g + 1 > t, S_g = 1) \right\} n_{gt} r^{\text{in}}_{gtl} (y_{gt} - y_{gt-\ell} - \widehat{m}^{\text{in}}_{gt\ell})$$

$$\widehat{\Delta}^{V,\text{out}}_{g,l} = \frac{1}{n_\ell} \sum_{t=\ell+1}^{T_g} \left\{ -\mathbf{1}(t_g + \ell = t) + \frac{n^{\text{out}}_{gt\ell}}{n^g_t} \mathbf{1}(t_g + 1 > t, S_g = -1) \right\} n_{gt} r^{\text{out}}_{gtl} (y_{gt} - y_{gt-\ell} - \widehat{m}^{\text{out}}_{gt\ell})$$

where $\widehat{m}^{\text{in}}_{gt\ell}$ and $\widehat{m}^{\text{out}}_{gt\ell}$ are sample means of ℓ-length differences among a particular cohort and r^{in}_{gtl} and r^{outl}_{gtl} are adjustment factors. Here we define $\widehat{m}^{\text{in}}_{gt\ell}$ and r^{in}_{gtl}, but $\widehat{m}^{\text{out}}_{gt\ell}$ and r^{out}_{gtl} are similarly defined.

For periods t, where group g has not yet experienced a switch ($t \leq t_g$), $\widehat{m}^{\text{in}}_{gt\ell}$ is computed in terms of not-yet-switched group cohorts defined by initial treatment d_{g1},

$$\widehat{m}^{\text{in}}_{gt\ell} = \frac{\sum_{g':d_{g'1}=d_{g1}} n_{g't}(y_{g't} - y_{g',t-\ell})}{\sum_{g':d_{g'1}=d_{g1}} n_{g't}}$$

provided there is at least one other group not yet switched sharing that value of the initial treatment d_{g1}. Otherwise, $\widehat{m}_{gt\ell}$ is computed in terms of group cohorts defined by initial treatment d_{g1}, including both not-yet-switched groups and switching-in groups for which $t = t_g + \ell$.

For periods t in which group g experienced its first switch from its baseline treatment value in period $t-\ell+1$ ($t = t_g + \ell$), $\widehat{m}^{\text{in}}_{gt\ell}$ is computed in terms of cohorts defined by initial treatment d_{g1}, latest period where no change had occurred t_g, and treatment value at first treatment change d_{g,t_g+1}, which we can call path cohorts. We have

$$\widehat{m}^{\text{in}}_{gt\ell} = \sum_{\substack{d_{g'1}=d_{g1} \\ g': \ t'_g=t_g \\ d_{g',t'_g+1}=d_{g,t_g+1}}} n_{g't}(y_{g't} - y_{g',t-\ell}) \Bigg/ \sum_{\substack{d_{g'1}=d_{g1} \\ g': \ t'_g=t_g \\ d_{g',t'_g+1}=d_{g,t_g+1}}} n_{g't}$$

provided there is at least one other group in the path cohort. Otherwise, $\widehat{m}^{\text{in}}_{gt\ell}$ is computed in terms of a cohort made up of any nonswitchers or switchers-in sharing the value of the initial treatment $D_{g,1}$.

r^{in}_{gtl} takes the value $\sqrt{n_{\text{coh}}/(n_{\text{coh}}-1)}$, where n_{coh} is the number of groups in the cohort used to compute $\widehat{m}^{\text{in}}_{gt\ell}$.

The variance estimator for $\hat{\delta}^n_\ell$ can then be written as

$$\hat{\sigma}^2_{\ell,n} = \frac{n^2_\ell}{(\delta^d_\ell)^2} \sum_{g=1}^{G} \mathbf{1}(\ell \leq T_g - 1)(\widehat{\Delta}^V_{g,l})^2$$

where $\widehat{\Delta}^V_{g,l} = \widehat{\Delta}^{V,\text{in}}_{g,l} + \widehat{\Delta}^{V,\text{out}}_{g,l}$. The estimated covariance between $\hat{\delta}^n_\ell$ and $\hat{\delta}^n_{\ell'}$ is

$$\hat{\sigma}^n_{\ell,\ell'} = \frac{n_\ell n_{\ell'}}{\delta^d_\ell \delta^d_{\ell'}} \sum_{g=1}^{G} \mathbf{1}(\ell \leq T_g - 1)\mathbf{1}(\ell' \leq T_g - 1)\widehat{\Delta}^V_{g,l}\widehat{\Delta}^V_{g,l'}$$

For both estimators, the raw equivalent is obtained by replacing δ_ℓ^d by n_ℓ for each exposure time ℓ.

If clustered variance–covariance estimators are requested, the size of the cohorts used to construct $\widehat{m}^{\text{in}}_{gt\ell}$, $\widehat{m}^{\text{out}}_{gt\ell}$, r^{in}_{gtl}, and r^{out}_{gtl} is determined by the number of clusters represented, rather than the number of groups in each cohort. The clustered variance estimator can then be written

$$\hat{\sigma}^2_{\ell,n} = \frac{n_\ell^2}{(\delta_\ell^d)^2} \sum_{c=1}^{C} \left\{ \sum_{g \in g_c} \mathbf{1}(\ell \leq T_g - 1) \widehat{\Delta}^V_{c,l} \right\}^2$$

where $c = 1, \ldots, C$ indexes clusters and g_c is the set of groups in cluster c.

Covariates

Covariates $\mathbf{x}_{gt}$ may be included in the model. As with the other variables, if covariates are not aggregated up to the group level, means are taken so that there is one observation for each (g, t) cell.

$\hat{\delta}^{X,n}_\ell$ is unmodified relative to $\hat{\delta}^n_\ell$, except that differences $y_{g,t_g+\ell} - y_{g,t_g}$ are replaced by the differences $y_{g,t_g+\ell} - y_{g,t_g} - (\mathbf{x}_{g,t_g+\ell} - \mathbf{x}_{g,t_g})' \hat{\pi}_{d_{g1}}$ for each group g.

$\hat{\pi}_{d_{g1}}$ is a vector of coefficients from regression of the first differences of y_{gt} on first differences of $\mathbf{x}_{gt}$, entering both linearly and interacted with indicator variables corresponding to each distinct value of d_{g1}, after partialing out the full set of time fixed effects (also included both linearly and as interactions with indicators for distinct values of d_{g1}), fit on the subset of group-time cells, where $t_g \geq t$, and with importance weights given by n_{gt}.

To write down a variance estimator, let $\tilde{\mathbf{x}}_{gt}$ be the vector of covariates projected off the time fixed effects, or in other words, the residuals from the partialing-out regression. Also let $S_{XX,d}$ be the mean sum-of-squares matrix of the projected covariates $\tilde{\mathbf{x}}_{gt}$ for a given subset of group-time cells (g, t), where $d_{g1} = d$ and $t_g \geq t$, and let $\widehat{\Delta} y_{gt}$ be the fitted values (linear predictions) for the first differences of y_{gt}. We can then write

$$\hat{\sigma}^2_{\ell,n,X} = \frac{n_\ell^2}{(\delta_\ell^d)^2} \sum_{g=1}^{G} (\widehat{\Delta}^{V,X}_{g,\ell} + e_{g,\ell})^2$$

Here $\widehat{\Delta}^{V,X}_{g,\ell}$ is equivalent to $\widehat{\Delta}^V_{g,\ell}$, except that differences $y_{g,t_g+\ell} - y_{g,t_g}$ are replaced by the differences $y_{g,t_g+\ell} - y_{g,t_g} - (\mathbf{x}_{g,t_g+\ell} - \mathbf{x}_{g,t_g})' \hat{\pi}_{d_{g1}}$ for each group g, including in the definitions of $\widehat{m}^{\text{in}}_{gtl}$ and $\widehat{m}^{\text{out}}_{gtl}$. The term e_{gl} corrects for the fact that the differences were constructed using the estimated quantities $\hat{\pi}_{d_{g1}}$ and is given as

$$e_{g,\ell} = \sum_{d \in \mathcal{D}_1} M_{d\ell} \left\{ S^{-1}_{XX,d} \frac{1}{n_d} \sum_{t=2}^{t_g} n_{gt} \tilde{\mathbf{x}}_{gt} r^X_d (\Delta y_{gt} - \widehat{\Delta} y_{gt}) \mathbf{1}(d_{g1} = d, t_g \geq 2) - \hat{\pi}_d \right\}$$

where

$$M_{d\ell} = \sum_{\substack{g: \\ t_g+\ell \leq T_g}}^{\substack{d_{g1}=d}} n_{g,t_g+\ell} \left\{ (\mathbf{x}_{g,t_g+\ell} - \mathbf{x}_{g,t_g})' - \sum_{\substack{g': \\ t_{g'}+1 > t_g+\ell}}^{\substack{d_{g'1}=d_{g1}}} \frac{n_{g',t_g+\ell}}{n^g_{t_g+\ell}} (\mathbf{x}_{g',t_g+\ell} - \mathbf{x}_{g',t_g})' \right\}$$

Further, $r^X_d = \sqrt{n_{\text{coh}}/(n_{\text{coh}} - 1)}$, where n_{coh} is the number of group-time cells for which $d_{g1} = d$.

In some cases, there may be insufficient data to compute $\hat{\pi}_{d_{g1}}$ and adjust for covariates for certain values of the initial treatment d_{g1}, even if there is enough information to compute DID estimators. In these cases, a note is printed and estimation proceeds without covariate adjustment for those values of d_{g1}.

Missingness

By default, `xtswitchdid` handles missing values the way most other estimation commands, including `hdidregress` and `xthdidregress`, handle them. Observations for which one of the variables specified by the user is missing are dropped at the outset, and estimation proceeds using the remaining sample. Time of first treatment switch for a group is determined by the first observed treatment value different from the observed baseline, even if there are intervening gaps in the time series that may have arisen from missing covariates or other values.

The `nolistwise` option still drops observations with missing group, panel, or time variables but otherwise handles missingness on the fly throughout the estimation procedure. Observations with missing outcomes or covariates but observed treatments are used to infer time of treatment switch but are excluded from effect estimation.

As a general principle, `xtswitchdid` with the `nolistwise` option attempts to maximize what can be estimated given the nonmissing data without imposing strong assumptions. If missing values of the treatment variable prevent inferring when a group had its first treatment switch, observations after those missing values are discarded. This corresponds to the "conservative" option for dealing with missing treatments discussed in de Chaisemartin et al. (2025). Any missing treatment values in periods prior to $t_g + \ell$ for a given group g and exposure ℓ are imputed as equal to the value of the treatment at the time of the first treatment switch, as in the same article. This means that these missing treatment values are assumed not to violate the requirement that treatments cannot both rise above and fall below the initial treatment level. It also allows computing $\widehat{\delta}^d_\ell$ for the normalized estimators and average total effect estimator. This imputation does not affect inference of an individual event-study effect being nonzero but may affect interpretation; the `raw` option may be more appropriate with the `nolistwise` option.

Note that whether `nolistwise` is specified or not, `xtswitchdid` drops data that cannot be used to compute differences. In particular, groups that are singletons, where only one time period is observed, are excluded from the estimation sample. Initial treatment groups for which there are no switchers or no available controls are also dropped.

Acknowledgments

We thank Clément de Chaisemartin of Sciences Po and Xavier D'Haultfœuille of CREST-ENSAE for their significant contributions to the literature on DID methods and for helpful conversations and advice. Their community-contributed command `did_multiplegt_dyn` implements a full-featured version of the method in this entry. We also thank their capable research assistants—with special thanks to David Arboleda, Diego Ciccia, Felix Knau, and Anzony Quispe for their feedback.

References

de Chaisemartin, C., D. Ciccia, F. Knau, M. Malézieux, D. Sow, D. Arboleda, R. Angotti, X. D'Haultfœuille, B. Li, H. Fabre, and A. Quispe. 2025. Using did_multiplegt_dyn to estimate event-study effects in complex designs: Overview, and four examples based on real datasets. Unpublished manuscript. https://doi.org/10.2139/ssrn.5337463.

de Chaisemartin, C., and X. D'Haultfœuille. 2026. Difference-in-differences estimators of intertemporal treatment effects. *Review of Economics and Statistics* 108: 863–880. https://doi.org/10.1162/rest_a_01414.

Also see

[CAUSAL] **xtswitchdid postestimation** — Postestimation tools for xtswitchdid+

[CAUSAL] **DID intro** — Introduction to difference-in-differences estimation

[CAUSAL] **didregress** — Difference-in-differences estimation+

[CAUSAL] **hdidregress** — Heterogeneous difference in differences+

[CAUSAL] **xthdidregress** — Heterogeneous difference in differences for panel data+

[U] 20 Estimation and postestimation commands