svytmle:

Cross-fitted TMLE and AIPW for complex survey data in Stata

M. Ehsan Karim · School of Population and Public Health, University of British Columbia, Vancouver

Get svytmle
github.com/ehsanx/svytmle-stata

1 The gap

Causal effects from surveys such as NHANES need **confounding adjustment** and **design-based inference** (weights, strata, PSUs, design df) *together*. Official Stata's routes on the NHANES example:

Approach	Doubly robust	Strata & PSUs	Design df (t)	Time*
`svy: logit + margins, vce(unconditional)`	no	yes	yes	2.3 s
`teffects ipwra [pw], vce(cluster psu)`	yes	PSUs only	no (z)	28 s
`svy: teffects`	*not allowed (linearized): r(322)*			
`svy jackknife _b: teffects ipwra`	yes	yes	yes	88 min
`svytmle`	yes	yes	yes	**1.4 s**

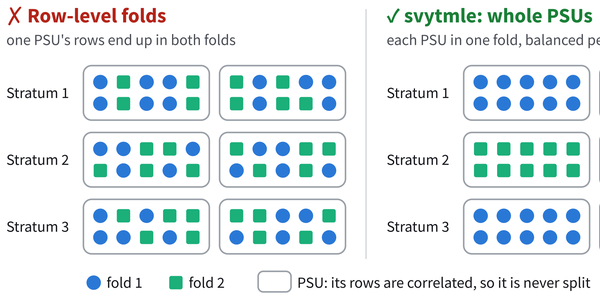
*One PC, Stata/SE 15.1; the jackknife refits `teffects` once per PSU (184 times). `teffects aipw` takes pweights only from StataNow 18 (r(101) before). `telasso` and `cate` cross-fit but take no pweights or `svy`.

2 One command, read from `svyset`

```
svyset SDMVPSU [pweight=WTMEC_POOLED], strata(SDMVSTRA)
svytmle Y, treatment(A) covariates(age i.sex i.race ///
    i.educ pir i.cycle i.smk i.married i.pa phq)  ///
    subpop(inpop) seed(20260607) rr or
```

- **`estimator()`** — `tmle` (default; targeted maximum likelihood) or `aipw` (augmented IPW: the DML score)
- **`subpop()`** — domains on the full design; `if` / `in` refused
- **`folds() seed()`** — whole-PSU folds within strata; also `foldvar()`, `strictfolds`
- **`nuisance()`** — `unweighted` (default; assumes non-informative sampling given the covariates) or `survey`

3 How it works



Cross-fitting by PSU. PSUs are the design's independent units; splitting one would let its correlated rows train the model that predicts for it.

1. **Folds:** whole PSUs to V folds within each stratum; V (default 5) drops to the fewest PSUs in any stratum (NHANES: 2–3, so V = 2).
2. **Nuisances:** logistic Q(A,W) and g(W) fit on the other folds, predicted for held-out PSUs.
3. **Targeting** (TMLE): survey-weighted logistic fluctuation on A/g and $(1-A)/(1-g)$.
4. **Inference:** weighted means; linearization of the influence function over strata and PSUs (with-replacement first stage); t on the design df.

4 NHANES 2007–2018: short sleep and obesity

Adults 20+, six pooled cycles: short sleep (< 7 h, 33%) vs 7–9 h; obesity BMI ≥ 30 (37%). Confounders: age, sex, race/ethnicity, education, income, cycle, smoking, marital status, activity, depression (PHQ-9 ≥ 10), singly imputed. *A methods illustration on cross-sectional data, not an epidemiologic finding.*

```
Survey-weighted cross-fitted TMLE                Number of strata  =          90
Estimand:  ATE (risk difference)                 Number of PSUs    =         184
Outcome:   Y                                     Number of obs     =      57,414
Treatment: A                                     Population size   = 309,033,079
Nuisance:  logit, unweighted                     Subpop. no. obs   =      30,983
Cross-fit: 2 folds of whole PSUs                 Subpop. size      = 212,636,714
           within strata, 4 strata uneven        Design df         =          94
Subpop.:   inpop

------------------------------------------------------------------------------
             |             Linearized
           Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ATE |   .0435657   .0080033     5.44   0.000     .0276749    .0594565
     POmean1 |   .4044401   .0074681    54.16   0.000     .3896119    .4192682
     POmean0 |   .3608744   .0060738    59.42   0.000     .3488148     .372934
------------------------------------------------------------------------------

Relative effects (delta method on the log scale, t with design df):
  Risk ratio     1.1207   [95% CI    1.0756,    1.1677]
  Odds ratio     1.2027   [95% CI    1.1247,    1.2861]
```

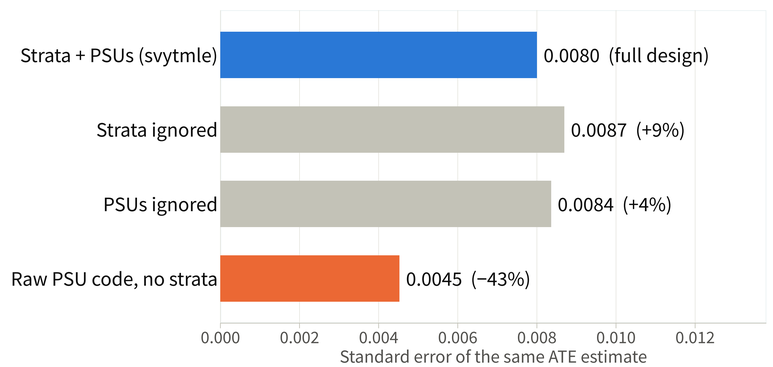
+4.4 pp
risk difference
95% CI 2.8 to 5.9

1.12
risk ratio
95% CI 1.08 to 1.17

0.15–0.78
out-of-fold propensity
range: good overlap

Same folds, same precision: TMLE 4.36 pp (SE 0.0080), AIPW 4.35 pp (SE 0.0080). `teffects ipwra` with pweights and PSU clusters: 4.40 pp (SE 0.0089), on z and without strata.

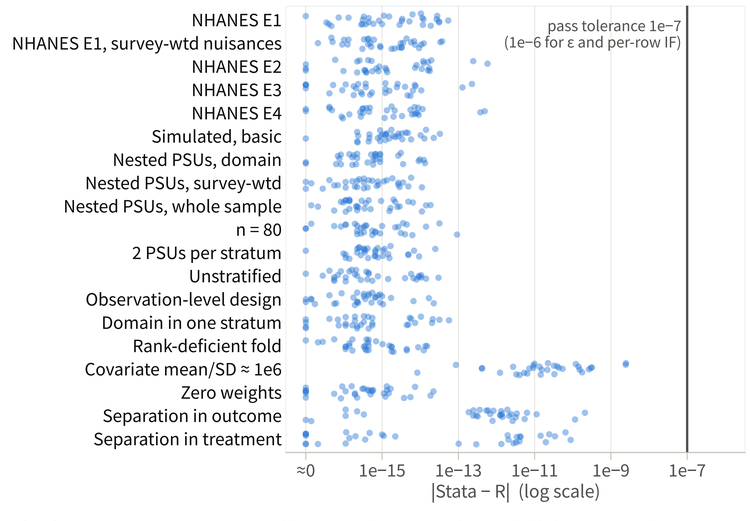
5 The design changes the uncertainty



Variance only: one estimate (4.36 pp) and its influence function (from `generate()`), summarised by `svy: mean` under four `svyset` declarations; the full design reproduces svytmle's SE exactly. Here, dropping strata or PSUs moved the SE modestly (+9%, +4%). NHANES reuses PSU codes 1–3 across strata, so `svyset` without `strata()` leaves 3 "clusters" and the SE falls **43%**. `svy` then uses t with 2 df, but a z interval on that SE, as from `vce(cluster SDMVPSU)`, is 44% too narrow: falsely precise.

Few PSUs, few df. The 2017–2018 cycle alone has 4,556 adults but 30 PSUs in 15 strata: df = 15, so its 95% CI is 8.7% wider than z.

6 Certified against the R package



All 909 checks pass. 855 compare Stata with R *svytmle* 0.1.2 on R's own folds in 19 cases: the 4 NHANES analyses (E1 also with survey-weighted nuisances) and 14 simulated stress designs; 54 test Stata-side invariants, postestimation and error paths. Over 627 numeric comparisons the median |Stata − R| is 7×10^{-16} and 90% are ≤ 10^{-12}; the largest, 2.5×10^{-9}, is one row's influence function in the mean/SD ≈ 10^{6} case. This checks the code; the estimator's properties are studied by simulation in Karim (2026).

7 Practical notes

- **Folds are random:** report `seed()`. Over 20 seeds the ATE varied with SD 0.0007 (0.09 SE); no repeated cross-fitting yet.
- **Fold balance** is reported (`e(fold_exact)`, `e(n_uneven)`); `strictfolds` enforces it.
- **Postestimation:** `lincom`, `test`, `nlcom` (with `df()`), `mi estimate:`; `generate()` saves g, Q and the influence function.
- **Limits:** logistic nuisance models only; binary treatment and outcome; no replicate weights yet.

8 Try it

```
net from https://raw.githubusercontent.com/ehsanx/svytmle-stata/main/
net install svytmle
webuse nhanes2
svytmle highbp, treatment(rural) covariates(age i.female i.race bmi) seed(1)
```

Take-home

1. Doubly robust ATEs from `svyset` data in one command, in seconds: TMLE or AIPW, with risk difference, risk ratio and odds ratio.
2. Strata, PSUs and design df change the uncertainty; svytmle derives them from `svyset`, as `svy` does, so declare the design correctly.
3. Pure Stata + Mata (Stata 15+); matches its R companion in all 19 certification cases.

References. Karim ME (2026). Cross-fitted survey-weighted TMLE with design-based variance for causal machine learning. arXiv:2606.30918. van der Laan MJ, Rose S (2011). *Targeted Learning*. Springer. Chernozhukov V et al. (2018). Double/debiased machine learning. *Econometrics J* 21:C1–C68. Binder DA (1983). *Int Stat Rev* 51:279–292. Data: NHANES 2007–2018, CDC/NCHS.
R package: github.com/ehsanx/svytmle (v0.1.2, commit 5534234).

ehsan.karim@ubc.ca