1

Pseudo-Values in Continuous Time Survival Analysis

In this document, I discuss the use of pseudo-values for survival analysis. I assume you have read the section on survival analysis in Chapter 16 and are familiar with that material. **Pseudo-value**s are a data-transformation technique that converts censored time-to-event data into continuous "uncensored" outcomes. The transformations can focus on survival probabilities or restricted mean survival times. They permit researchers to bypass complex hazards models and directly apply standard regression tools.

The pseudo-value approach for survival analysis was developed by Andersen, Klein and Rosthøj (2003). It is based on traditional "leave-one-out" jackknife resampling theory, The program *Survival pseudo vars* on the *Programs* tab of my website generate variables with the relevant transformations for both RMSTs and for survival probabilities. In doing so, you need to specify the time horizons (tau) that you want to focus on.

Because pseudo-values are continuous and uncensored they can be used in regression modeling. The use of jack-knifing in their generation can introduce some dependencies in scores across individuals. The dependencies tend to be diffuse and lessen with increasing N to the point that many researchers treat them as “functionally independent.” Dependencies due to the jackknife tend to be minimal when N is approximately 200 or greater, though this can be undermined if the chosen tau is subject to few participants who remain at risk at tau. It also is advisable to use robust sandwich estimators (see Andersen & Perme, 2010 and Graw & Schumacher, 2009**).**

When modeling pseudo-value survival probabilities some of the individual scores can be less than 0 and/or greater than 1.00. This is not problematic and aligns with methods like multiple imputation or propensity score weighting. The pseudo values are linear extrapolations designed to reflect each individual’s leverage on the survival curve. With pseudo-values, the mathematical objective is for the average of the values to equal the unbiased non-parametric survival estimate. The mathematics of the transformations ultimately can push some individual’s score below 0 or above 1. This usually occurs when N is small or when an individual experiences an event or is censored at an influential time point. Allowing the possibility of some offending values at the individual level accomplishes this. Do not truncate the pseudo values to only occur between 0 and 1, Doing so can introduce bias.

Traditionally, pseudo-values are analyzed using GEE methods (see *Chapter 16* and *Chapter 25*). I illustrate such analyses first followed by FISEM analyses using Mplus.

GEE ANALYSIS

I use the program called *GEE cluster regression* on my website on the Programs tab. The tradition for pseudo values is to define each individual as a separate cluster by using their ID identifier as the cluster variable and then coupling this with an independence working correlation structure and a robust estimator. I used these options in my program for a tau 0f 24. Here are the core results for the RMST:

```
Coefficients:
            Estimate Std.err     Wald Pr(>|W|)
T             0.7240  0.4641    2.434    0.119
m1           -3.1390  0.1747  322.669   <2e-16 ***
m2           -3.0630  0.1717  318.302   <2e-16 ***

Approximate Confidence Intervals (normal theory)
             coeff   lower  upper
T            0.724 -0.1856  1.634
m1          -3.139 -3.4815 -2.796
m2          -3.063 -3.3995 -2.727
```

Compare these results to those reported in the main text using traditional time-to-event analyses:

```
Model summary (difference of RMST)
            coef se(coef)       z     p lower .95 upper .95
T          0.724    0.464   1.560 0.119    -0.186     1.634
m1        -3.139    0.175 -17.963 0.000    -3.481    -2.796
m2        -3.063    0.172 -17.841 0.000    -3.400    -2.727
```

The results are close to one another.

Here is the GEE analysis of the survival probabilities (note: I do not treat the outcome as binary; the pseudo value is continuous):

```
Coefficients:
             Estimate   Std.err    Wald Pr(>|W|)
T            0.029195  0.024485   1.422    0.233
m1          -0.143397  0.009490 228.347   <2e-16 ***
m2          -0.139622  0.009265 227.095   <2e-16 ***
```

```
Approximate Confidence Intervals (normal theory)
               coeff   lower    upper
T            0.02919 -0.0188  0.07718
m1          -0.14340 -0.1620 -0.12480
m2          -0.13962 -0.1578 -0.12146
```

For M1, for every one unit M1 increases, the survival probability is predicted to decrease by -0.14 ±0.02, holding the other predictors constant. For M2, for every one unit M2 increases, the survival probability is predicted to decrease by -0.14 ±0.02, holding the other predictors constant.

SEM ANALYSIS

The Mplus program is a straightforward application of RET syntax for a continuous outcome. In this case, I analyze the survival probabilities at tau = 24 by using the outcome variable I call psurv24 as generated by the aforementioned pseudo-values program on my website. To deal with dependences and non-normality, I use `TYPE = COMPLEX` and I identify a cluster variable as the participant id variable (see *Chapter 25*). I also make use of the `MODEL INDIRECT` command because as far as Mplus is concerned this is not a survival analysis so the complications of estimating the parameters of the baseline hazard issues are moot. For justification of the approach I outline here, see Wang and Logan (2024). Here is the relevant Mplus syntax:

```
1.  TITLE: Analysis of Pseudo Values ;
2.  DATA:
3.    FILE = pseudoM.txt;
4.  DEFINE:
5.    CENTER m1b m2b (GRANDMEAN) ;
6.  VARIABLE:
7.    NAMES =id time m1 m2 T m1b m2b event prmst6 psurv6
8.           prmst12 psurv12 prmst18 psurv18 prmst24 psurv24;
9.    USEVARIABLES = psurv24 m1 m2 T m1b m2b ;
10.   CLUSTER = id ;
11. ANALYSIS:
12.   TYPE = COMPLEX ;
13. MODEL:
14.   m1 ON T m1b (p1 b1) ;
15.   [m1] (am1) ;        !intercept of above equation
16.   m2 ON T m2b (p2 b2);
17.   [m2] (am2) ; ;      !intercept of above equation
18.   psurv24 ON m1 m2 T  (p3 p4 p5);
19.   [psurv24] (ay) ;   !intercept of above equation
20. MODEL INDIRECT:
```

```
21.   psurv24 IND T ;
22. OUTPUT: Samp StdYX Mod(4) Residual Cinterval Tech4 ;
```

All of the syntax should be familiar to you (again, see Chapter 25 for a discussion of clustering). Note also that when passing the survival probability pseudo-values to Mplus, I do not impose `CATEGORICAL` status on them. I treat the pseudo-variable as continuous. The use of `TYPE = COMPLEX` invokes a special form of the robust sandwich MLR estimator appropriate for the analysis.

The global model fit indices and localized fit indices appear as standard output. Here are the (edited) global fit indices all of which point to reasonable model fit:

```
MODEL FIT INFORMATION

Number of Free Parameters                       13

Chi-Square Test of Model Fit

          Value                             8.832*
          Degrees of Freedom                    5
          P-Value                          0.1160
          Scaling Correction Factor        0.9757
            for MLR

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                          0.023
          90 Percent C.I.                   0.000  0.047
          Probability RMSEA <= .05          0.973

CFI/TLI

          CFI                               0.996
          TLI                               0.990

SRMR (Standardized Root Mean Square Residual)

          Value                             0.014
```

The modification indices and the standardized residuals also were consistent with a reasonable fitting model.

Total Effect of the Intervention on Survival Probabilities

The FISEM estimated effect of the intervention on survival probabilities is located in the `MODEL INDIRECT` section. Here is the edited output:

```
TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

Effects from T to PSURV24

  Total                0.242      0.023     10.366      0.000
```

The estimated mean difference in survival probabilities for the intervention minus the control group was 0.24 ±0.05, CR = 10.34, $p < 0.05$. The survival probabilities for the separate groups can be obtained by adding the following `MODEL CONSTRAINT` commands just before Line 22:

```
MODEL CONSTRAINT:
NEW (tsmean csmean);
csmean = ay + p3*am1 + p4*am2 + p5*0 ;
tsmean = ay + p3*(am1+p1) + p4*(am2+p2) + p5*1 ;
```

Here is the relevant output:

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

New/Additional Parameters
    TSMEAN             0.441      0.018     23.886      0.000
    CSMEAN             0.199      0.014     13.967      0.000
```

The estimated survival probability at month 24 for the intervention group was 0.44 ±0.03 and for the control group it was 0.20 ±0.03.

Effect of the Intervention on the Mediators

Here is the relevant output for the estimated effects of the intervention on M1 and M2:

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
M1	ON				
T		-0.728	0.053	-13.816	0.000
M1B		0.204	0.026	7.834	0.000
M2	ON				
T		-0.775	0.054	-14.470	0.000
M2B		0.203	0.028	7.302	0.000

Here are the intercepts that reflect the covariate adjusted means for the control group:

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Intercepts				
M1	-0.051	0.037	-1.371	0.171
M2	0.010	0.037	0.269	0.788

Here are the intercepts for the reverse scored treatment dummy variable analysis, which yields the estimated covariate adjusted means for the intervention group:

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Intercepts				
M1	-0.779	0.038	-20.728	0.000
M2	-0.765	0.039	-19.747	0.000

For M1, the adjusted posttest mean for the intervention group was -0.78 ±0.08 and for the control group it was -0.051 ±0.08. The difference between the adjusted means was -0.73 ±0.11 (CR = 13.82, $p < 0.05$). For M2, the adjusted posttest mean for the intervention group was -0.77 ±0.08 and for the control group it was 0.01 ±0.08. The difference between the adjusted means was -0.77 ±0.11 (CR = 14.47, $p < 0.05$).

Effect of the Mediators on the Outcome

Here is the output for estimating the presumed impact of the mediators on the survival probabilities for the 24 month time horizon:

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

 PSURV24  ON
    M1              -0.143      0.009    -15.106      0.000
    M2              -0.140      0.009    -15.065      0.000
    T                0.029      0.024      1.191      0.234
```

For M1, holding M2 and the treatment condition constant, a one-unit increase in M1 is predicted to decrease the survival probability by -0.14 ±0.02 units (CR = 15.11, $p < 0.05$). Framing the result in accord with the intervention focus, decreasing M1 by one unit raises the survival probability by 0.14 ±0.02 units. For M2, holding M1 and the treatment condition constant, a one-unit increase in M2 is predicted to decrease the survival probability by -0.14 ±0.02 units (CR = 15.07, $p < 0.05$). Framing the result in accord with the intervention focus, decreasing M2 by one unit raises the survival probability by 0.14 ±0.02 units.

Omnibus Mediation

There are multiple strategies one can use to evaluate the indirect effect of a mediator on the outcome. A straightforward way to test the null hypothesis of no mediation for a given mediator is to use the joint significance test as outlined in Chapter 9. In the current example, both the link T→M and the link M→Y were statistically significant ($p < 0.05$) for both mediators which leads us to reject the null hypotheses of no mediation in each case.

An alternative strategy is to use the coefficient product approach. Here is the relevant (edited) output from the Mplus output:

```
TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

  Specific indirect 1
    PSURV24
    M1
    T                0.104      0.010     10.479      0.000

  Specific indirect 2
    PSURV24
    M2
    T                0.108      0.010     11.020      0.000
```

In both cases, the coefficient product of the relevant coefficients were statistically significant (for M1, coefficient product = 0.10 ±0.02, CR = 10.48, $p < 0.05$; for M2, coefficient product = 0.11 ±0.02, CR = 11.02, $p < 0.05$), leading to the rejection of the null hypothesis of no mediation for both M1 and M2, respectively. These results should be replicated using bootstrapping as discussed in prior chapters.

Profile Analyses

Profile analyses are straightforward using the `MODEL CONSTRAINT` commands. For example, I might want to compare profiles of two individuals, one at low risk of diet abandonment and one at high risk of diet abandonment. I define a high-risk individual as someone who did not receive the intervention (T = 0), and who has elevated M1 and M2 scores near their 80th quantiles (M1 = 0.50 and M2 = 0.50). I define a low-risk individual as someone who received the intervention (T = 1), and who had relatively low scores on M1 and M2, near their 20th quantiles (M1 = 1.30 and M2 = -1.30). Here are the relevant `MODEL CONSTRAINTS` commands that would be placed just before Line 22.

```
MODEL CONSTRAINT:
NEW(hirisk lowrisk contrast);
  hirisk =  ay + (p3 * .5) + (p4 * .5) + (p5 * 0);
  lowrisk = ay + (p3 * (-1.3)) + (p4 * (-1.3)) + (p5 * 1) ;
  contrast = lowrisk-hirisk ;
```

Here is the output:

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

New/Additional Parameters
    HIRISK             0.052      0.012      4.475      0.000
    LOWRISK            0.591      0.018     32.272      0.000
    CONTRAST           0.539      0.022     24.003      0.000
```

The survival probability for a high-risk individual is 0.54 ±0.04 units higher for the high-risk individual than for the low-risk individual.

CONCLUDING COMMENTS ON PSEUDO VALUES

If you want to use FISEM methods to analyze time-to event data with right censoring, the pseudo-value transformation strategy is a reasonable candidate. This is especially true if the proportional hazard assumptions of traditional methods are violated. You can explore

different values of tau in a FISEM context and analyze either RMSTs. survival probabilities or CIFs. The approach offers a great deal of flexibility.

REFERENCES

Andersen, P. K., Klein, J. P., & Rosthøj, S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, *90*(1), 15–27

Andersen, P. K., & Perme, M. P. (2010). Pseudo-observations in survival analysis. *Statistical Methods in Medical Research*, 19, 71-99.

Graw, F., Gerds, T. A., & Schumacher, M. (2009). On pseudo-values for supplementary survival analysis. *Biometrika*, 96, 503-513.

Wang, W., & Logan, B. (2024). Simplifying causal mediation analysis for time-to-event outcomes using pseudo-values. https://arxiv.org/pdf/2411.17533