

Description and instructions for the use of SAS macros for the self-controlled case series analyses

Two SAS macros have been developed to analyse self-controlled case series (SCCS) data. The macro %SCCSDTA creates the analysis dataset (providing more flexibility than %SCCS) and %SCCS_MOD estimates the case series model for censored, perturbed, or curtailed post-event exposures on the basis of estimating equations with the sandwich variance estimates (Farrington 2009). The macros %SCCSDTA and %SCCS_MOD were developed in the context of a study assessing the association between a specific vaccine and the risk of solid organ transplant (SOT) rejection (Cohet et al. 2016), and adjusted for the time since transplantations (categorised). Therefore examples will be based on this study.

Description of %SCCSDTA

The macro %SCCSDTA creates the analysis dataset to be used with %POISREG or %SCCS_MOD. The output dataset of the macro includes several rows for each subject, each row corresponding to the time intervals defined by risk periods of the time-varying covariates, and including the number of events and the time of follow-up during each interval.

Example of input dataset

Input data are listed by Event, which includes both adverse events and exposures. An example of one subject in the analysis of the risk of SOT rejection within 30 days after vaccination, and adjusted for the time since transplantation (categorized in 0-90, 91-180 and >180 days) is given below. Event types in this study included the date of transplant `Transpl` (time since transplantation was included as a time-varying covariate), vaccine doses `Dose 1`, `Dose 2` the exposures, and solid organ transplant (SOT) rejections `Rejection` the adverse event of interest.

Patid	Event	Eventdate
1	Transpl	05AUG2009
1	Dose1	15OCT2009
1	Dose2	28NOV2009
1	Rejection	21JAN2010

Note: The observation period starts on 01-Oct-2009 (beginning of the vaccination campaign) and ends on 30-Sep-2010; one risk period of 90 days is associated with the transplantation; effect of all vaccine doses assumed to be the same

Example of output dataset

Patid	start	end	Dose1_30_	Dose2_30_	Transpl_90_	Transpl_180_	cnt_evt	fu_days	Dose
1	01OCT2009	14OCT2009	0	0	1	0	0	14	0
1	15OCT2009	03NOV2009	1	0	1	0	0	20	1
1	04NOV2009	14NOV2009	1	0	0	1	0	11	1
1	15NOV2009	27NOV2009	0	0	0	1	0	13	1
1	28NOV2009	28DEC2009	0	1	0	1	0	31	2
1	29DEC2009	01FEB2010	0	0	0	1	1	35	2
1	02FEB2010	30SEP2010	0	0	0	0	0	241	2

Dose1_30_ : flag for the period 0-30 days after the first dose of vaccine
Dose2_30_ : flag for the period 0-30 days after the second dose of vaccine

Transpl_90_: flag for the period 0-90 days after transplantation
Transpl_180_: flag for the period 91-180 days after transplantation
Cnt_evt: Number of rejections
Fu_days: Time in the period (days)
Dose: variable indicating which dose precedes the period

Macro parameters:

- in_dta*: name of the input dataset. This dataset should include one row per event or covariate of interest needed for the analyses. In the example, the dataset included the rejections (at least one per subject), the transplantation and the vaccination dates. The datasets should at minimum include the variables *patid* (subject identifier), *event* (type of event, including exposures), *eventdate* (start date of the event).
- sccs_evt*: outcome event. In the example it was the SOT rejections (*Rejection*).
- start*: start of the observation period. Only outcome events and follow-up time between *start* and *end* will be considered for the analyses. This day is included in the observation period. 01-Oct-2009 in the example
- end*: end of the observation period. This day is included in the observation period. 30-Sept-2010 in the example
- prior*: period before the observation period in which the exposures are assessed. For example, transplantations occurring less than 90 days before the start of the observation period have an associated risk period overlapping the observation period. In the example input data set, the transplantation occurred before the start of the observation period (5/8/2009), but the risk periods (0-90 and 91-180 days) end after the beginning of the observation period (1/10/2009).
- cens_sec*: Censoring at the second event (Y or N). The case series model for censored, perturbed, or curtailed post-event exposures allows only one outcome event per subject. In this example the follow-up of the subjects was censored at the subsequent event, but this censoring should only be used in rare situations because the end of the observation period must be independent of outcome event times. Default=Y, but most studies should set *cens_sec* = N.
- cond_var*: Event determining the period at risk of an outcome event. In the example, *cond_var*=Transpl because subjects are at risk of rejection only after the transplantation. In addition, subjects are censored at any new transplantation following the rejection considered for the analysis.
- Note: when the observation period is defined by the ages at the beginning and at the end, this variable can be the dates when the subject reached these ages. In this case the input dataset would include “age” events with the dates of ages at the beginning and at the end of the observation period.
- cond_main*: Y or N. If Y then the censoring according to *cond_var* is done according to the occurrence of the main exposure (“Vaccination” in the example). For instance, when *cond_main* =‘Y’ and the subject experienced several transplantations during the

observation period, only data from the last transplantation before vaccination to the next transplantation is considered (if it includes a rejection), even if there are previous rejections in the observation period. If *cond_main* = 'N', the period from the last transplantation before the first rejection to the next transplantation is considered, no matter whether there are vaccinations and rejections later. Default=N.

- rsk_lst*: List of independent variables to consider in the SCCS model, including main exposure. If there are several doses and the case series model for censored, perturbed, or curtailed post-event exposures will be used, then there should be one separate variable for each dose even if the effect is assumed to be the same for all doses. In the example: *rsk_lst*:Dose1 Dose2 Transpl
- rsk_time*: List of risk periods associated to each of the exposures or covariates in the same order as the independent variables listed in *rsk_lst*. When a covariate is associated with several risk periods, the ends of each of them is separated by '\$'. In the example: *rsk_time*=30 30 90\$180 (0-30 days after Dose 1, 0-30 days after Dose 2 and 0-90, 91-180 days after Transpl).
- Note: when gaps are needed (wash-out period for instance), the period of the gap can be created in the output dataset as any risk period and then removed.
- main_rsk*: Main exposure variables. One variable per dose for the case series model for censored, perturbed, or curtailed post-event exposures. In the example: *main_rsk*=Dose1 Dose2.
- end_cond*: Covariates for which the end date of the condition is taken into account for determining the risk period. For instance, if chemotherapy is one of the covariates, the risk period associated may include the period of treatment plus an additional following period (mentioned in *rsk_time*).
- end_var*: Variable with the end date of the covariate mentioned in *end_cond*. For instance end of the chemotherapy treatment.
- out*: Name of the output dataset for the SCCS analyses. Default=scsstd.

Description of %SCCS_MOD

The macro %SCCS_MOD fits the case series model for censored, perturbed, or curtailed post-event exposures, using the estimating equation approach and computing the sandwich variance estimates of the parameters. There is no limit for the number of doses.

Macro parameters:

- in_dta*: input dataset, created by %SCCSDTA. In addition to the variables needed for the macro %POISREG, the dataset must include a variable 'dose' identifying whether the period is before the first dose, between first and second dose, after second dose (in case of two doses but the macro can handle more than two doses).
- dep_v*: dependent variable: count of events during each period. Default=cnt_evt

- expos*: list of dummy variables for the risk periods associated with each dose. There should be one variable per dose. In the example: *dep_v=Dose1_30_ Dose2_30_*
- common*: common (=Y) or separate (=N) main exposure effect. Separate effect means that distinct relative incidences will be computed for dose 1 and dose 2. Default=Y
- covar*: covariates to include in the model. Only continuous variables are valid. Categorical variables should be handled through the use of dummy variables. In the example: *covar= Transpl_90_ Transpl_180_*
- start_b*: starting values for the parameters in the iteration process. The parameters of the doses should be the first one(s). There should be one parameter per dose and per covariate. When the effect of the doses is common for all doses (parameter *common=Y*), the parameters should be equal. In the example: *start_b=0 0 0 0*
- eps*: criteria for convergence. Default=1e-08
- alpha*: significance level for the confidence interval. Default=0.05
- min_cse*: Sometimes it can be planned a priori that if the sample size is too low, the model should not be estimated. This parameter is the minimum number of exposed cases to perform the analysis. Default=10
- min_cov*: minimum number of cases by covariate to perform the analysis. Default=5
- out_dta*: output dataset. Default=par_cov

Call of macros for example

```
%SCCSDTA(dta_in=sccs_1,
          sccs_evt=Rejection,
          start=01OCT2009,
          end=30SEP2010,
          prior=180,
          cens_sec=N,
          cond_var= Transpl,
          cond_main=N,
          rsk_lst= Dose1 Dose2 Transpl,
          rsk_time=30 30 90$180,
          main_rsk= Dose1 Dose2,
          end_cond=,
          end_var=,
          out=sccsstd);

%SCCS_MOD(in_dta=sccsstd,
          dep_v=cnt_evt,
          expos= Dose1_30_ Dose2_30_,
          common=Y,
          covar= Transpl_90_ Transpl_180_,
          start_b=0 0 0 0,
          eps=1e-08,
          alpha=0.05,
          min_cse=10,
          min_cov=5,
          out_dta=par_cov);
```

References

Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics* 1995;51:228-35.

Farrington CP, Whitaker HJ, Hocine MN. Case series analysis for censored, perturbed, or curtailed post-event exposures. *Biostatistics* 2009;10:3-16.

Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. *Stat Med* 2006;25:1768-97.

Cohet C. et al. Effect of the adjuvanted (AS03) A/H1N1 2009 pandemic influenza vaccine on the risk of rejection in solid organ transplant recipients in England: a self-controlled case series. *BMJ Open* 2016. Accepted for publication.