

# Results schema of the SelfControlledCaseSeries package

Martijn J. Schuemie

2024-08-15

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Exposures, covariates of interest, and controls . . . . .	1
1.2	Exposures-outcome-sets, analysis IDs and models . . . . .	2
1.3	Fields with minimum values . . . . .	2
<b>2</b>	<b>Tables</b>	<b>2</b>
2.1	Table <code>sccs_age_spanning</code> . . . . .	2
2.2	Table <code>sccs_analysis</code> . . . . .	2
2.3	Table <code>sccs_attrition</code> . . . . .	2
2.4	Table <code>sccs_calendar_time_spanning</code> . . . . .	3
2.5	Table <code>sccs_censor_model</code> . . . . .	3
2.6	Table <code>sccs_covariate</code> . . . . .	3
2.7	Table <code>sccs_covariate_analysis</code> . . . . .	3
2.8	Table <code>sccs_covariate_result</code> . . . . .	4
2.9	Table <code>sccs_diagnostics_summary</code> . . . . .	4
2.10	Table <code>sccs_era</code> . . . . .	4
2.11	Table <code>sccs_event_dep_observation</code> . . . . .	4
2.12	Table <code>sccs_exposure</code> . . . . .	5
2.13	Table <code>sccs_exposures_outcome_set</code> . . . . .	5
2.14	Table <code>sccs_likelihood_profile</code> . . . . .	5
2.15	Table <code>sccs_result</code> . . . . .	5
2.16	Table <code>sccs_spline</code> . . . . .	6
2.17	Table <code>sccs_time_to_event</code> . . . . .	6
2.18	Table <code>sccs_time_trend</code> . . . . .	6

## 1 Introduction

This document describes the data model of the output of the SelfControlledCaseSeries (SCCS) package, generated by the `exportToCsv()` function. This vignette assumes you are already familiar with the SelfControlledCaseSeries package, and have read all other vignettes.

### 1.1 Exposures, covariates of interest, and controls

As described in the ‘Single studies using the SelfControlledCaseSeries package’ vignette, **eras** are cohorts or drug eras extracted from the database. **Covariates** can either be splines, for example representing age or season, or era covariates, derived from eras. When defining covariates using the `createEraCovariateSettings()` function we can either use verbatim era IDs (e.g. cohort IDs), or we can reference a variable (typically called ‘exposureId’). When defining **exposures** using the `exposure()` function, we can define different era IDs to be used for this variable, thereby using the same analysis settings for different exposures and outcomes. For each exposure we can set the `trueEffectSize` if known. Any exposure with known true effect size is considered a

`control`, and will be used for empirical calibration. Some of our covariates can be marked as **covariates of interest** by setting `exposureOfInterest = TRUE` when calling `createEraCovariateSettings()`. This is especially relevant for the results model, since these covariates will be reported in the `sccs_result` table.

## 1.2 Exposures-outcome-sets, analysis IDs and models

Using the `createExposuresOutcome()` function we can define an outcome with one or more exposures, since an SCCS model can have multiple exposures (e.g. we could have separate exposures for the first and second dose of a vaccine). With the `createSccsAnalysis()` function we can create a set of settings for analysis describing which data to extract from the database, how to transform that data including which covariates to construct, and how to fit the SCCS model. Each analysis setting has a unique analysis ID. Each combination of an exposures-outcome-set and an analysis setting will correspond to one SCCS model. A model can have multiple covariates, and each covariates can be based on multiple eras.

## 1.3 Fields with minimum values

Some fields contain patient counts or fractions that can easily be converted to patient counts. To prevent identifiability, these fields are subject to a minimum value. When the value falls below this minimum, it is replaced with the negative value of the minimum. For example, if the minimum subject count is 5, and the actual count is 2, the value stored in the data model will be -5, which could be represented as '<5' to the user. Note that the value 0 is permissible, as it identifies no persons. These fields are identified below as having `Min. count = 'Yes'`.

# 2 Tables

In this section you will find the list of tables and their fields.

## 2.1 Table `sccs_age_spanning`

Field	Type	Key	Min. count	Description
<code>analysis_id</code>	int	Yes	No	A foreign key referencing the <code>sccs_analysis</code> table.
<code>exposures_outcome_set_id</code>	int	Yes	No	A foreign key referencing the <code>sccs_exposures_outcome_set</code> table.
<code>database_id</code>	varchar	Yes	No	Foreign key referencing the database.
<code>age_month</code>	int	Yes	No	Age in months since birth.
<code>cover_before_after_subjects</code>	int	No	Yes	Number of subjects whose observation period covers this month.

## 2.2 Table `sccs_analysis`

Field	Type	Key	Min. count	Description
<code>analysis_id</code>	int	Yes	No	A unique identifier for an analysis.
<code>description</code>	varchar	No	No	A description for an analysis, e.g. 'Correcting for age and season'.
<code>definition</code>	varchar	No	No	A JSON object specifying the analysis.

## 2.3 Table `sccs_attrition`

Field	Type	Key	Min. count	Description
<code>sequence_number</code>	int	Yes	No	The place in the sequence of steps defining the final analysis.
<code>description</code>	varchar	No	No	A description of the last restriction, e.g. "Removing persons with missing data".
<code>analysis_id</code>	int	Yes	No	A foreign key referencing the <code>sccs_analysis</code> table.

Field	Type	Key	Min. count	Description
exposures_outcome_set_id	int	Yes	No	A foreign key referencing the sccs_exposures_outcome_set table.
covariate_id	int	Yes	No	A foreign key referencing the sccs_covariate table. The identifier for a covariate.
database_id	varchar	Yes	No	Foreign key referencing the database.
outcome_subjects	int	No	Yes	The number of subjects with at least one outcome.
outcome_events	int	No	Yes	The number of outcome events.
outcome_observation_periods	int	No	Yes	The number of observation periods containing at least one outcome event.
observed_days	bigint	No	Yes	The number of days subjects were observed.

## 2.4 Table sccs\_calendar\_time\_spanning

Field	Type	Key	Min. count	Description
analysis_id	int	Yes	No	A foreign key referencing the sccs_analysis table.
exposures_outcome_set_id	int	Yes	No	A foreign key referencing the sccs_exposures_outcome_set table.
database_id	varchar	Yes	No	Foreign key referencing the database.
calendar_year	int	Yes	No	Calendar year (e.g. 2022)
calendar_month	int	Yes	No	Calendar month (e.g. 1 is January).
cover_before_after_subjects	int	No	Yes	Number of subjects whose observation period covers this month.

## 2.5 Table sccs\_censor\_model

Field	Type	Key	Min. count	Description
analysis_id	int	Yes	No	A foreign key referencing the sccs_analysis table.
exposures_outcome_set_id	int	Yes	No	A foreign key referencing the sccs_exposures_outcome_set table.
database_id	varchar	Yes	No	Foreign key referencing the database.
parameter_id	int	Yes	No	The parameter number in the censor model (starting at 1).
parameter_value	float	No	No	The fitted parameter value.
model_type	varchar	No	No	The type of censor model. Can be ‘Weibull-Age’, ‘Weibull-Intensity’, ‘Weibull-Time’.

## 2.6 Table sccs\_covariate

Field	Type	Key	Min. count	Description
analysis_id	int	Yes	No	A foreign key referencing the sccs_analysis table.
exposures_outcome_set_id	int	Yes	No	A foreign key referencing the sccs_exposures_outcome_set table.
covariate_id	int	Yes	No	A unique identifier for a covariate.
covariate_name	varchar	No	No	A description for the covariate.
era_id	int	No	No	A foreign key referencing the sccs_era table.
covariate_analysis_id	int	No	No	A foreign key referencing the sccs_covariate_analysis table.
database_id	varchar	Yes	No	Foreign key referencing the database.

## 2.7 Table sccs\_covariate\_analysis

Field	Type	Key	Min. count	Description
analysis_id	int	Yes	No	A foreign key referencing the sccs_analysis table.
covariate_analysis_id	int	Yes	No	A unique identifier for a covariate analysis.
covariate_analysis_name	varchar	No	No	A name for a covariate analysis, e.g. ‘Pre-exposure’.

Field	Type	Key	Min. count	Description
variable_of_interest	int	No	No	Is the variable of interest (1 = yes, 0 = no).

## 2.8 Table sccs\_covariate\_result

Field	Type	Key	Min. count	Description
analysis_id	int	Yes	No	A foreign key referencing the sccs_analysis table.
exposures_outcome_set_id	int	Yes	No	A foreign key referencing the sccs_exposures_outcome_set table.
database_id	varchar	Yes	No	Foreign key referencing the database.
covariate_id	int	Yes	No	The identifier for the covariate.
rr	float	No	No	The estimated relative risk (i.e. the incidence rate ratio).
ci_95_lb	float	No	No	The lower bound of the 95% confidence interval of the relative risk.
ci_95_ub	float	No	No	The upper bound of the 95% confidence interval of the relative risk.

## 2.9 Table sccs\_diagnostics\_summary

Field	Type	Key	Min. count	Description
analysis_id	int	Yes	No	A foreign key referencing the sccs_analysis table.
exposures_outcome_set_id	int	Yes	No	A foreign key referencing the sccs_exposures_outcome_set table.
covariate_id	int	Yes	No	The identifier for the covariate of interest.
database_id	varchar	Yes	No	Foreign key referencing the database.
mdrr	float	No	No	The minimum detectable relative risk.
ease	float	No	No	The expected absolute systematic error.
time_trend_p	float	No	No	The p for whether the mean monthly ratio between observed and expected rates is significantly different from 1.
pre_exposure_p	float	No	No	One-sided p-value for whether the rate before exposure is significantly different from the rate after exposure.
mdrr_diagnostic	varchar(20)	No	No	Pass / warning / fail / not evaluated classification of minimum detectable relative risk.
ease_diagnostic	varchar(20)	No	No	Pass / warning / fail / not evaluated classification of expected absolute systematic error.
time_trend_diagnostic	varchar(20)	No	No	Pass / warning / fail / not evaluated classification of time trend p-value.
pre_exposure_diagnostic	varchar(20)	No	No	Pass / warning / fail / not evaluated classification of pre-exposure p-value.
unblind	int	No	No	Is unblinding the result recommended? (1 = yes, 0 = no).
unblind_for_evidence_synthesis	int	No	No	Is unblinding the result for inclusion in evidence synthesis?

## 2.10 Table sccs\_era

Field	Type	Key	Min. count	Description
exposures_outcome_set_id	int	Yes	No	A foreign key referencing the sccs_exposures_outcome_set table.
analysis_id	int	Yes	No	A unique identifier for an analysis.
era_type	varchar	Yes	No	The type of era (e.g. 'rx' for drugs).
era_id	int	Yes	No	A unique identifier, corresponding to the ID in the source table.
era_name	varchar	No	No	A name for the era. Is NULL for eras derived from cohorts.
database_id	varchar	Yes	No	Foreign key referencing the database.

## 2.11 Table sccs\_event\_dep\_observation

Field	Type	Key	Min. count	Description
analysis_id	int	Yes	No	A foreign key referencing the sccs_analysis table.

Field	Type	Key	Min. count	Description
exposures_outcome_set_id	int	Yes	No	A foreign key referencing the sccs_exposures_outcome_set table.
database_id	varchar	Yes	No	Foreign key referencing the database.
months_to_end	int	Yes	No	Number of months until observation end.
censored	int	Yes	No	Whether the observation is censored (meaning, not equal to the true effect size).
outcomes	int	No	Yes	Number of outcomes observed during the month.

## 2.12 Table sccs\_exposure

Field	Type	Key	Min. count	Description
exposures_outcome_set_id	int	Yes	No	A foreign key referencing the sccs_exposures_outcome_set table.
era_id	int	Yes	No	A foreign key referencing the sccs_era table.
true_effect_size	float	No	No	If known, the true effect size. For negative controls this equals zero.

## 2.13 Table sccs\_exposures\_outcome\_set

Field	Type	Key	Min. count	Description
exposures_outcome_set_id	int	Yes	No	A unique identifier for a set of exposures and an outcome.
outcome_id	int	No	No	A cohort ID.
nesting_cohort_id	int	No	No	A cohort ID.

## 2.14 Table sccs\_likelihood\_profile

Field	Type	Key	Min. count	Description
log_rr	float	Yes	No	The log of the relative risk where the likelihood is sampled.
log_likelihood	float	No	No	The normalized log likelihood.
covariate_id	int	Yes	No	The identifier for the covariate of interest.
exposures_outcome_set_id	int	Yes	No	A foreign key referencing the sccs_exposures_outcome_set table.
analysis_id	int	Yes	No	A foreign key referencing the sccs_analysis table.
database_id	varchar	Yes	No	Foreign key referencing the database.

## 2.15 Table sccs\_result

Field	Type	Key	Min. count	Description
analysis_id	int	Yes	No	A foreign key referencing the sccs_analysis table.
exposures_outcome_set_id	int	Yes	No	A foreign key referencing the sccs_exposures_outcome_set table.
covariate_id	int	Yes	No	A foreign key referencing the sccs_covariate table. The identifier for the covariate of interest.
rr	float	No	No	The estimated relative risk (i.e. the incidence rate ratio).
ci_95_lb	float	No	No	The lower bound of the 95% confidence interval of the relative risk.
ci_95_ub	float	No	No	The upper bound of the 95% confidence interval of the relative risk.
p	float	No	No	The two-sided p-value considering the null hypothesis of no effect.
one_sided_p	float	No	No	The one-sided p-value considering the null hypothesis of no effect.
outcome_subjects	int	No	Yes	The number of subjects with at least one outcome.
outcome_events	int	No	Yes	The number of outcome events.
outcome_observation_periods	int	No	Yes	The number of observation periods containing at least one outcome.
covariate_subjects	int	No	Yes	The number of subjects having the covariate.

Field	Type	Key	Min. count	Description
covariate_days	int	No	Yes	The total covariate time in days.
covariate_eras	int	No	Yes	The number of continuous eras of the covariate.
covariate_outcomes	int	No	Yes	The number of outcomes observed during the covariate time.
observed_days	bigint	No	Yes	The number of days subjects were observed.
log_rr	float	No	No	The log of the relative risk.
se_log_rr	float	No	No	The standard error of the log of the relative risk.
llr	float	No	No	The log of the likelihood ratio (of the MLE vs the null hypothesis).
calibrated_rr	float	No	No	The calibrated relative risk.
calibrated_ci_95_lb	float	No	No	The lower bound of the calibrated 95% confidence interval of the relative risk.
calibrated_ci_95_ub	float	No	No	The upper bound of the calibrated 95% confidence interval of the relative risk.
calibrated_p	float	No	No	The calibrated two-sided p-value.
calibrated_one_sided_p	float	No	No	The calibrated one-sided p-value considering the null hypothesis.
calibrated_log_rr	float	No	No	The log of the calibrated relative risk.
calibrated_se_log_rr	float	No	No	The standard error of the log of the calibrated relative risk.
database_id	varchar	Yes	No	Foreign key referencing the database.

## 2.16 Table sccs\_spline

Field	Type	Key	Min. count	Description
analysis_id	int	Yes	No	A foreign key referencing the sccs_analysis table.
exposures_outcome_set_id	int	Yes	No	A foreign key referencing the sccs_exposures_outcome_set table.
database_id	varchar	Yes	No	Foreign key referencing the database.
spline_type	varchar	Yes	No	Either 'age', 'season', or 'calendar time'.
knot_month	float	Yes	No	Location of the knot. For age, the month since birth. For season, the month of the year.
rr	float	No	No	The estimated relative risk (i.e. the incidence rate ratio).

## 2.17 Table sccs\_time\_to\_event

Field	Type	Key	Min. count	Description
analysis_id	int	Yes	No	A foreign key referencing the sccs_analysis table.
exposures_outcome_set_id	int	Yes	No	A foreign key referencing the sccs_exposures_outcome_set table.
database_id	varchar	Yes	No	Foreign key referencing the database.
era_id	int	Yes	No	A foreign key referencing the sccs_era table. The identifier for the era.
week	int	Yes	No	The number of the week relative to exposure. Week 0 starts on the day of exposure.
observed_subjects	int	No	Yes	The number of people observed during the week.
outcomes	int	No	Yes	The number of outcomes observed during the week.

## 2.18 Table sccs\_time\_trend

Field	Type	Key	Min. count	Description
analysis_id	int	Yes	No	A foreign key referencing the sccs_analysis table.
exposures_outcome_set_id	int	Yes	No	A foreign key referencing the sccs_exposures_outcome_set table.
database_id	varchar	Yes	No	Foreign key referencing the database.
calendar_year	int	Yes	No	The calendar year (e.g. 2022).
calendar_month	int	Yes	No	The calendar month (e.g. 1 for January).
observed_subjects	int	No	Yes	Number of people observed during the month.
ratio	float	No	No	Observed over expected ratio, where the expected count assumes no effect of the exposure.

Field	Type	Key	Min. count	Description
adjusted_ratio	float	No	No	Observed over expected ratio, where the expected count is adj