Macro TDRDAS User's Guide
30 December 2019

Methods

True discovery rate degree of association set (TDRDAS) analysis is a method to identify sets of predictors for which the true degree of association with clinical outcome or state is more than a specified value while controlling the false discovery rate (Benjamini and Hochberg, 1995). The methodology is described in detail by Crager (2010).

The method starts with a set of predictors, each having (1) an estimated degree of association with clinical outcome or state and (2) a standard error of the estimate. These might be, for example, a standardized log hazard ratio and its standard error from a Cox proportional hazards regression analysis, or a standardized log odds ratio estimate and its standard error from a logistic regression analysis. Given the estimates and standard errors, and a specified true discovery rate $1 - \gamma$ (for example, $1 - \gamma = 0.80$ or 0.90) the method identifies set of predictors among which $100(1 - \gamma)\%$ can be expected to have true absolute degree of association higher than a given value θ. Then, letting θ vary over a grid of values, the method calculates for each predictor a maximum lower bound (MLB) absolute degree of association for which the predictor belongs to a TDRDA set with true discovery rate $1 - \gamma$.

We are apt to focus further research on the predictors with the strongest associations. Because of regression to the mean, the absolute degrees of association that these predictors show in subsequent research is likely to be less than the "naïve" estimate of the degree of association in the current study. To get a realistic estimate of the degree of association we could expect for each predictor in subsequent research, the TDRDA set method computes an estimated degree of association corrected for regression to the mean (RM).

The MLB absolute degree of association and the RM-corrected estimate can be displayed together in a concise bar-chart summary of the analysis results. The graph includes all predictors for which the false discovery rate for Wald tests of non-zero degree of

association is γ or less, since all TDRDA sets with true discovery rate $1 - \gamma$ are refinements of this set.

Macro TDRDAS

The macro TDRDAS carries out the calculations for the TDRDA set analysis. The macro takes as input a data set containing the predictor name, the estimated degree of association and its standard error, with one record for each predictor. The macro produces an output data set containing the MLB absolute degree of association and RM-corrected estimate of the degree of association for each predictor. At the user's discretion, the output data set may contain transformed values of the MLB and RM-corrected estimates. For example, if the input estimates are log hazard ratios, the user may with to use the exponential transform to express the MLB and RM-corrected estimates as hazard ratios. The macro also produces a TDRDA set bar chart showing the (transformed) MLB and RM-corrected estimates.

The macro is called as follows:

```
%macro tdrdas(
  /* Input Specification*/   indsn=, predictorname=,
                             estimate=, stderr=, estzero=,
  /* Analysis Parameters */  accuracy=,oneminusq=,lambda=,
  /* Output Specification */ outdsn=, transformx=,
                             MLBvar=, RMCEstvar=,
   /* Graph Options*/        refinterval=, measure=, graphlabel=,
                             goutpath=, graphname=, maxpredictorsppg=,
                             graphcolor1=, graphcolor2=
           );
```

The macro parameters are described in Table 1.

Parameter	Type	Required?	Default Value	Description
			Table 1. Macro TDRDAS Parameters	
indsn	$	Yes	—	Input data set name
predictorname	$	Yes	—	Predictor name
estimate	#	Yes	—	Input data set variable that contains the estimated degrees of association
stderr	#	Yes	—	Input data set variable that contains the standard errors of the estimated degrees of association
estzero	#	Yes	—	Value of the association estimate that corresponds to no association. If the association is measured by log hazard ratio or log odds ratio, then estzero = 0.
accuracy	#	No	0.001	Accuracy to which MLB degree of association is calculated.
oneminusq	#	No	0.9	True discovery rate (equal to 1 minus the acceptable false discovery rate).
lambda	#	No	0.5	The value of the tuning parameter λ in Storey's method for estimating the proportion of true null hypotheses.
outdsn	$	Yes	—	Name of the output data set.
transformx	$	No	exp(x)	Optional transformation to be applied to the MLB absolute degrees of association and RM-corrected estimates. The parameter must include an "(x)" with no spaces between x and each parenthesis. Every time x appears in the function, it must be encased in parentheses.
MLBvar	#	No	MLB	Name of the output data set variable that will contain the (possibly transformed) MLB absolute degree of association for each predictor.
RMCEstvar	#	No	RMCest	Name of the output data set variable that will contain the (possibly transformed) RM-corrected estimates of the degree of association for each predictor.

Parameter	Type	Required?	Default Value	Description
refinterval	#	No	0.05	Spacing of reference lines on the graph. Used in the axis statement as in "axis1 order=(1 to 1.75 by &refinterval.)"
measure	$	No	Degree of Association	The measure of the degree of association. This is used in the legend for the graph and in variable labels. If you change this, you will also need to change the value for graphlabel. The value selected for this parameter is case sensitive.
graphlabel	$	No	Standardized Degree of Association of Predictor with Recurrence	This parameter allows the user to modify the label on the graph.
goutpath	$	No		The goutpath parameter can be used to redirect the graph to a folder other than the one in which the calling SAS program is running.
graphname	$	No	TDRDAS	Name of the file that will contain the TDRDAS bar chart. **Note:** Do not add &timestamp or the file extension; the macro will add them.
maxpredictors ppg	#	No	150	The maximum number of predictors to be shown on each page of the graph. The value selected should be between 1 and 150.
graphcolor1	$	No	black	The color of the bars on the graph representing the predictors with positive association.
graphcolor2	$	No	gray	The color of the bars on the graph representing the predictors with negative association.

Table 1. Macro TDRDAS Parameters

References

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* **57**:289–300.

Crager MR (2010). Gene identification using true discovery rate of association degree of association sets and estimates corrected for regression to the mean. *Statistics in Medicine* **29**:33–45.