# Stata practical for missing data in meta-analysis

### Haloperidol meta-analysis

Data set haloperidol.dta contains the data from the review of trials comparing haloperidol with placebo in patients with schizophrenia. Variables rh, fh and mh contain the numbers of observed successes, observed failures and missing values in the haloperidol arms of 17 studies, and variables rp, fp and mp contain the corresponding numbers for the placebo arms.

1. Start by drawing a forest plot and doing a simple meta-analysis (ACA) using

metan rh fh rp fp, rr fixedi label(namevar=author)

We focus on a fixed-effect inverse-variance-weighted meta-analysis (fixedi option) of the risk ratio (rr option) comparing haloperidol against placebo. You can improve the appearance of the graph by using standard options such as xlabel(0.1,1,10,100) and title(metan analysis).

1. Explore the missing data. For example, you can look at the fraction missing by arm using

gen missfrach = mh/(rh+fh+mh)

gen missfracp = mp/(rp+fp+mp)

gen id = substr(author,1,2) // two-letter abbreviation to label the next graph

scatter missfrac\*, mlab(id)

1. To use metamiss to allow for the missing data, you must download it using

net from http://www.mrc-bsu.cam.ac.uk/IW\_Stata/meta/

net install metamiss

1. Replicate the metan analysis using

metamiss rh fh mh rp fp mp, rr fixed id(author) aca

1. Perform an ICA-0 analysis by replacing aca by ica0 above. Similarly you can use the ica1, icapc and icap options (but you must use only one of these options at a time). In each case, inspect the forest plot and see how the pooled estimate has changed and how the weights given to the studies with most missing data have changed. Also explore the icab and icaw options - note that these give much more extreme results.

Now we explore the analysis using reasons.

1. First look at the data using

list author d\*

The variables dfh, dsh, dch and dgh contain the counts in the haloperidol arm to be imputed as failures (ICA-0), as successes (ICA-1), using the control risk (ICA-pC) and using the same-arm risk (ICA-p) respectively, and similarly for the placebo arm. For example, the haloperidol arm in the Arvanitis study has 2 missing values, and the counts dfh, dsh, dch and dgh are 17, 0, 17 and 0. These counts relate to a different outcome and are used to indicate the proportions in which the 2 missing values are shared out: in this case, the analysis takes one value to be imputed as a failure and one to be imputed using the control risk.

1. Perform the meta-analysis under the above assumptions using

metamiss rh fh mh rp fp mp, rr id(author) fixed ica0(dfh dfp) ica1(dsh dsp) icapc(dch dcp) icap(dgh dgp)

Finally we explore the analysis using IMORs.

1. Since outcomes in patients with missing data are likely to be worse than in observed patients, we consider IMORs lower than 1. To take an IMOR of 0.5, use

metamiss rh fh mh rp fp mp, rr id(author) fixed imor(0.5)

You can also specify imor(0) and see that it agrees with ica0.

1. To allow uncertainty about this IMOR, we use the sdlogimor() option. In the text, we assumed that the IMOR lies between 0.5 and 1 with probability 2/3. Making a normality assumption for the log IMOR, this means that -0.7 (=ln(0.5)) is one standard deviation (SD) below the mean and 0 (=ln(1)) is one SD above the mean, so the mean (logimor) is about -0.35 and the SD (sdlogimor) is about 0.35. We run this using

metamiss rh fh mh rp fp mp, rr id(author) fixed logimor(‑0.35) sdlogimor(0.35)

1. We can further allow the IMORs to be correlated across the arms of each study: higher correlations express the belief that the IMPs are closely related, and tends to down-weight studies with more missing data in one arm than in the other. Use a correlation of 0.5 by adding the corr(0.5) option to the previous command.
2. Also try larger levels of uncertainty. For example, observe how changing to sdlogimor(3) dramatically reduces the weight given to the studies with most missing data.

### Mirtazapine meta-analysis

Data set mirtazapine.dta contains summary data from eight studies comparing the effectiveness of mirtazapine and placebo in patients with major depression. Variables yt, sdt, nt and mt contain the mean value, the standard deviation, the number of completers and the number of non-completers in the mirtazapine (treatment) arms, and variables yc, sdc, nc and mc contain the corresponding numbers for the placebo (control) arms. We will use the metamiss2 command to apply the IMDOM model. The command metamiss2 can be downloaded by typing within Stata

net **from http://www.mtm.uoi.gr**

net install metamiss2

1. Start by drawing a forest plot and doing a simple meta-analysis (ACA) using

metan nt yt sdt nc yc sdc, nostandard lcols(study) randomi

Note that we focus on a random-effect inverse-variance-weighted meta-analysis (randomi option) of mean differences (nostandard option). You can improve the appearance of the graph by using standard options such as

* xlab(-10,-5,0,5,10)
* favours(favors mirtazapine # favors placebo)
* texts(180)
1. Use metamiss2 to replicate the results from the metan command (conduct an ACA):

metamiss2 nt mt yt sdt nc mc yc sdc, lcols(study) md imptype(imdom) impmean(0 0) impsd(0 0)

Similar to the metan command, if eight variables are declared, data are assumed to be continuous and to be the sample size, number of dropouts, mean value and standard deviation in the experimental group (nt mt yt sdt) followed by the corresponding numbers in the control group (nc mc yc sdc).

Note that the metamiss2 command can handle different effect measures. Here the md option gives the mean difference (the default); alternatives are smd and rom for the standardized mean difference and ratio of means respectively. The command imptype()defines the type of relation between the outcome in the missing data and in the observed data. There are two options for the imptype command. Here, we have assumed its default value (imdom). If, instead of IMDOM, we would like to employ IMROM, we should write imptype(imrom).

The informative missingness parameter is assumed to follow a normal distribution and there are two parameters that need to be defined in both arms (mean value and standard deviation). The option impmean($μ\_{t}$ $μ\_{c}$) defines the mean value in the experimental group ($μ\_{t}$) and the mean value in the control group ($μ\_{c}$). Similarly, impsd($σ\_{t}$ $σ\_{c}$)defines the corresponding standard deviations. The default is impmean(0 0) and impsd(0 0) which correspond to an ACA.

Options md, imptype(imdom), impmean(0 0) and impsd(0 0)are redundant in this example because we have defined the command to assume its default values (ACA). Here and below, you can add other options of the metan command, such as those suggested in question 12.

1. Now we explore departures from the MAR assumption using the IMDOM. As in the chapter, we assume a small departure from the MAR assumption in which the IMDOM in each arm lies between ‑3 and +3 with probability 95%. Under a normality assumption, this implies a mean of zero with a standard deviation of 1.5. We specify this by impmean(0 0) impsd(1.5 1.5):

metamiss2 nt mt yt sdt nc mc yc sdc, lcols(study) md impmean(0 0) impsd(1.5 1.5)

We can graphically compare this analysis with the ACA analysis, as in Figure 2 of the paper, using option compare():

metamiss2 nt mt yt sdt nc mc yc sdc, lcols(study) md impmean(0 0) impsd(0 0) compare(impmean(0 0) impsd(1.5 1.5))

1. Now we assume a more systematic departure from the MAR assumption, in which the IMDOM has different distributions in the two arms. In the mirtazapine arm, we assume the IMDOM has mean 0.5 (expressing an expectation that missing values are likely to be higher than observed values) with 95% range from -1.5 to 2.5 (implying a standard deviation of 1). In the placebo group, we assume the IMDOM has mean -1 (missing values are likely to be lower than observed values) with 95% range from -4 to 2 (and a standard deviation of 1.5):

metamiss2 nt mt yt sdt nc mc yc sdc, lcols(study) md impmean(0.5 -1) impsd(1 1.5)

1. Repeat the analysis in step 15 assuming that the IMDOMs in the mirtazapine and placebo groups are correlated (option corr() ) with correlation 0.5. The default correlation is zero; that is that IMPs follow independent normal distribtuions:

metamiss2 nt mt yt sdt nc mc yc sdc, lcols(study) impmean(0.5 ‑1) impsd(1 1.5) corr(0.5)

1. Consider departures from the MAR assumption by assuming $μ\_{c}=μ\_{t}=0$ but increasing the uncertainty in the IMDOMs:

metamiss2 nt mt yt sdt nc mc yc sdc, lcols(study) md impmean(0 0) impsd(2 2)

metamiss2 nt mt yt sdt nc mc yc sdc, lcols(study) md impmean(0 0) impsd(3 3)

As uncertainty in the IMDOM increases, we observe that the weight of studies with a large missing rate (e.g. study 5 MIR 003-021) reduces.

1. Conduct a sensitivity analysis showing how the uncertainty in IMDOM impacts the summary effect:

metamiss2 nt mt yt sdt nc mc yc sdc, sensitivity

The option sensitivity conducts a sensitivity analysis starting with IMDOM=0 without uncertainty (MAR assumption) and gradually increasing the uncertainty of IMDOM (departing from the MAR assumption).