

# CDD user guide

PsN 4.6.0

Revised 2016-04-13

## 1 Introduction

The Case Deletions Diagnostics (CDD) algorithm is a tool primarily used to identify influential components of the dataset, usually individuals. The CDD works by identifying groups in the data set and creating one new data set for each member of the group, where that member has been removed. The model is run once with each new data set. The PsN implementation of the CDD can take any column as base for the grouping and all rows with the same value in that column will be considered a group as long as no individual contain multiple values in that column. One should take care that the grouping creates sensible individual records. PsN will renumber the ID column so that two individuals with the same ID will not end up next to each other. Examples

```
cdd moxonidine.mod -case_column=1
cdd pheno.mod -case_column=AGE
```

## 2 Input and options

### 2.1 Required input

A model file is required on the command-line.

**-case\_column** = *name/number*

The column on which the case-deletion is done. You can either

use the name of the column as specified in the \$INPUT record in the model file or you can use the column number in the \$INPUT record. Numbering starts with 1.

## 2.2 Optional input

**-bins** =  $N$

Sets the number of databins, or cdd datasets, to use. If the number of unique values, or factors, in the case column is higher than  $N$  then one or more factors will be deleted in each cdd dataset. Specifying  $N$  as higher than the number of factors will have no effect.  $N$  is then reset to the number of factors. Default value = Number of unique values in the case column.

**-xv**

Default true. Run the cross-validation step (-xv) or not (-no-xv).

**-selection\_method** = *random* / *consecutive*

Default consecutive. Specifies whether the factors selected for exclusion should be drawn randomly or consecutively from the datafile.

**-outside\_n\_sd\_check** =  $X$

Default 2. Mark the runs with CS-CR outside  $X$  standard deviations of the PCA.

## 2.3 Some important common PsN options

There are many options that govern how PsN manages NONMEM runs, and those options are common to all PsN programs that run NONMEM. For a complete list see `common_options.pdf`, or `psn_options -h` on the command-line.

**-h or -?**

Print the list of available options and exit.

**-help**

With -help all programs will print a longer help message. If an

option name is given as argument, help will be printed for this option. If no option is specified, help text for all options will be printed.

**-directory** = *'string'*

Default `cdd_dirN`, where N will start at 1 and be increased by one each time you run the script. The directory option sets the directory in which PsN will run NONMEM and where PsN-generated output files will be stored. You do not have to create the directory, it will be done for you. If you set `-directory` to a the name of a directory that already exists, PsN will run in the existing directory.

**-seed** = *'string'*

You can set your own random seed to make PsN runs reproducible. The random seed is a string, so both `-seed=12345` and `-seed=JustinBieber` are valid. It is important to know that because of the way the Perl pseudo-random number generator works, for two similar string seeds the random sequences may be identical. This is the case e.g. with the two different seeds 123 and 122. Setting the same seed guarantees the same sequence, but setting two slightly different seeds does not guarantee two different random sequences, that must be verified.

**-clean** = *'integer'*

Default 1. The clean option can take four different values:

**0** Nothing is removed

**1** NONMEM binary and intermediate files except INTER are removed, and files specified with option `-extra_files`.

**2** model and output files generated by PsN restarts are removed, and data files in the `NM_run` directory, and (if option `-nmqual` is used) the xml-formatted NONMEM output.

**3** All `NM_run` directories are completely removed. If the PsN tool has created `modelfit_dir:s` inside the main run directory, these will also be removed.

**-nm\_version** = *'string'*

Default is 'default'. If you have more than one NONMEM version installed you can use option `-nm_version` to choose which one to use, as long as it is defined in the `[nm_versions]` section in `psn.conf`, see `psn_configuration.pdf` for details. You can check which versions are defined, without opening `psn.conf`, using the command

```
psn -nm_versions
```

**-threads** = *'integer'*

Default 5 (if default PsN config file is used). Use the `threads` option to enable parallel execution of multiple models. This option decides how many models PsN will run at the same time, and it is completely independent of whether the individual models are run with serial NONMEM or parallel NONMEM. If you want to run a single model in parallel you must use options `-parafile` and `-nodes`. On a desktop computer it is recommended to not set `-threads` higher the number of CPUs in the system plus one. You can specify more threads, but it will probably not increase the performance. If you are running on a computer cluster, you should consult your system administrator to find out how many threads you can specify.

**-version**

Prints the PsN version number of the tool, and then exit.

**-last\_est\_complete**

is optional and only applies with NONMEM7 and `cdd` option `-xv`. See `common_options.pdf` for details.

## 2.4 cdd rplots

PsN can automatically generate R plots to visualize results for `cdd`, using a default template found in the `R-scripts` subdirectory of the installation directory. The user can also create a custom template, see more details in section “Auto-generated R-plots from PsN” in `common_options.pdf`.

If option `-rplots` is set  $\geq 1$ , a plot with Covariance ratios vs Cook scores for each case, e.g. ID, will be generated. The default `cdd rplots` template

requires no special R libraries. If no pdf is generated, see the .Rout file in the main run directory for error messages.

**-rplots** = *level*

-rplots<0 means R script is not generated

-rplots=0 (default) means R script is generated but not run

-rplots=1 means basic plots are generated

-rplots=2 means basic and extended plots are generated

### **Troubleshooting**

If no pdf was generated even if a template file is available and the appropriate options were set, check the .Rout-file in the main run directory for error messages. If no .Rout-file exists then check that R is properly installed, and that either command 'R' is available or that R is configured in psn.conf.

## **3 Known bugs/issues**

If NONMEM6 is used with the cdd and the S matrix is algorithmically singular (message in lst-file, checked also by sumo script) the Cook scores cannot be trusted. The cdd does not check for this error.