



UPPSALA
UNIVERSITET

Visual Predictive Checks for Censored and Categorical data

Martin Bergstrand, Andrew C Hooker, Mats O Karlsson
Department of Pharmaceutical Biosciences, Uppsala University, Sweden

Objective

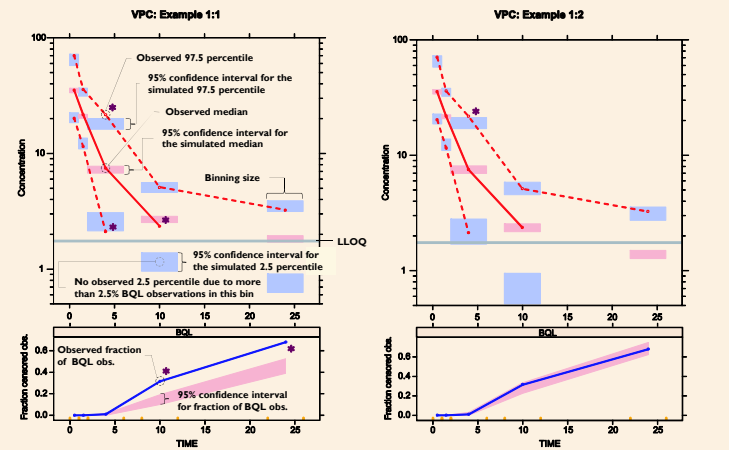
To illustrate a new approach for Visual Predictive Checks (VPCs) in the presence of categorized data.

Background

- Non random censoring of data such as in the presence of observations below the quantification limit (BQL) can harm the interpretation of diagnostic plots such as VPCs^[1].
- The visual interpretation of a VPC is its strength, however it can be difficult to distinguish if lack of agreement is due to random chance or model misspecification.

Methods

- 1000 simulated datasets for each VPC.
- To ensure correctly calculated percentiles all censored observations were retained in the dataset. Percentiles for observed data can only be adequately calculated and presented for percentiles where the censored observations constitute a smaller fraction than the percentile in question.
- Non parametric 95% confidence intervals (CI) for percentiles of continuous data and proportion of total for categorical data.



Example 1. Principle components of the VPC illustrated for Example 1. VPCs constructed with parameter estimates obtained with BQL observations omitted (Example 1:1, left) and with the M3 method applied for handling of BQL data (Example 1:2, right). Principal components of the VPC pointed out in the left hand figure. Indications of discrepancy between model simulations and observations pointed out with (*).

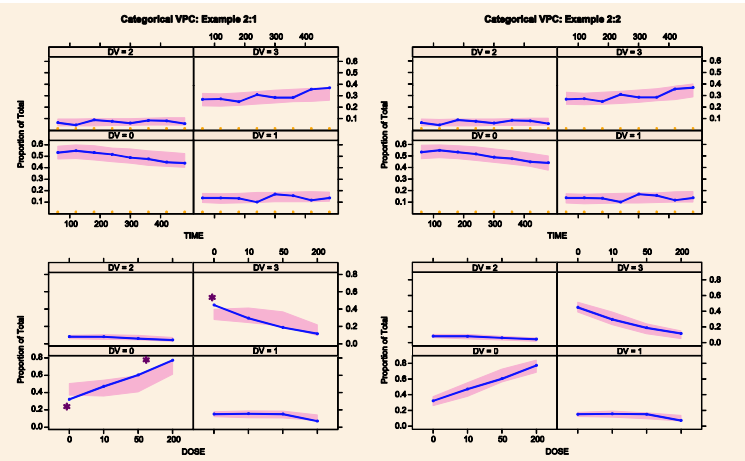
Conclusions

- Treating censored data as a combination of continuous data and categorical data (e.g. BQL) can facilitate unbiased interpretation.
- VPCs can be adopted for any type of categorical data by plotting the observed and the simulated fraction of observations of each category versus an independent variable.
- One good looking VPC is not a guarantee of an unbiased model.

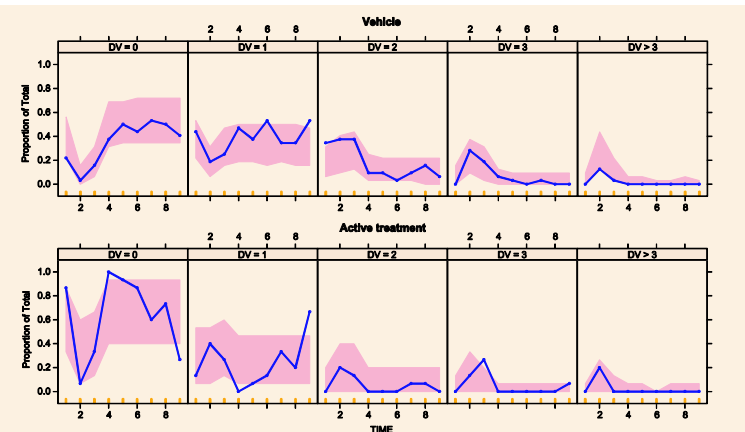
Application scope for suggested approach:

- Censored continuous data**
 - BQL censoring (**Example 1**)
 - Adaptive censoring of observations (**See poster 1605^[2]**)
- Categorical data**
 - Non-ordered / Ordered categorical data (**Example 2**)
 - Count data (**Example 3 and poster 1630^[3]**)
 - Time-to-event data
- Illustrative categorization of continuous data**
 - Neutropenia: Grade 0-4 (**Example 4 on hand-out paper**)
 - ALAT: >ULN (Upper Limit of Normal), >3xULN, >10xULN

Results



Example 2. VPCs for order categorical adverse event score. VPCs created for a model with a linear drug effect (Example 2:1, left) and a model with an E_{max} effect (Example 2:2, right). The upper plots were constructed with TIME as the independent variable and did not indicate a misspecification for either models. In the lower plots dose was used as the independent variable and misspecification was indicated for the linear drug effect model (*).



Example 3. Categorical VPC applied to acid reflux count data^[4]. The upper panel shows the observed fraction of each category (count(s)) versus time (blue line) and the corresponding model based 95% CI (red field) for vehicle treated dogs. The lower panel shows similar plots for active treatment.



Xpose and PsN are freely available at
<http://xpose.sf.net> and <http://psn.sf.net>

[1] Bergstrand M, Karlsson MO. Handling Data Below the Limit of Quantification in Mixed Effect Models. AAPS 2009 Jun;11(2):371-80.
[2] ÅM Johansson et al. Pharmacokinetics of high-dose methotrexate in adults and children. PAGE 18 (2009) Abstr 1605 [www.page-meeting.org/abstract=1605]
[3] Plan E, Karlsson MO. New models for handling correlated underdispersed Likert pain scores. PAGE 18 (2009) Abstr 1630 [www.page-meeting.org/abstract=1630]
[4] G. Ma, et al. PK/PD modeling for transient lower esophageal sphincter relaxations: count method and repeated time to event method. ASCPT meeting 2009, PII-53.

Basic PsN functionalities for VPC

This is a small excerpt from the NPC/VPC user-guide and process description (<http://psn.sourceforge.net/docs.php>).

Basic example:

```
vpc run1.mod -samples=1000 -seed=123 -bin_by_count=1 -no_of_bins=5
```

Selected input options:

-samples=<number> (required)

the number of simulated datasets to generate.

-msfo=<file> (optional, required if the model has \$NONP record)

name of msfo-file used for simulations.

-lst=<file> (optional, but forbidden together with -msfo)

name of lst-file from which parameter estimates should be obtained.

-stratify_on=<variable> (optional).

either PRED, STRT, or a variable in the \$INPUT record used for stratification of the VPC plots.

-no_of_strata=N (optional, only allowed when -stratify_on is used)

used to define how many strata the defined stratification variable should be divided into.

-dv=<variable> (optional, default is DV)

defines the dependent variable of the VPC.

-mirrors=N (optional)

the number of mirror plot data sets to produce (mirror plot = "how could these plots look like if the model was the true model").

-idv=<variable> (optional, default TIME)

defines the independent variable.

-lloq=<number> (optional)

the lower limit of quantification for left censored data.

-uloq=<number> (optional)

the upper limit of quantification for right censored data.

-levels=<number1,number2,...> (optional)

the boundaries for a categorization of the dependent variable. Category 1 <= number1 < category 2 <= number2, etc.

-nopred,

if this option is chosen "NOPRED" will be added to the \$SIMULATION record of the simulation model in addition to ONLYSIMULATION. This option is recommended with likelihood models for odd type data (see NONMEM manual for additional information).

Binning of independent variable (IDV):

There are a number of options for automatic binning across the IDV in addition to the possibility of a user defined array for the binning. These options are explained in detail in the on-line user-guide.

PsN output:

VPC output is found in npc_dir<number> (or any name assigned by the option "-dir=<name>"). The original data run with MAXEVAL=0 is done in subdirectory NM_run1 and the simulations in NM_run2. A subdirectory m1 contains the PsN generated model files for simulation and the simulated raw data.

The final results VPC results are presented in the file vpc_results.csv.

This file contains for each strata:

- The following percentiles as calculated for each IDV bin of the observed data: 2.5%, 5%, 10%, 30%, 50%, 70%, 90%, 95%, 97.5%.
- Medians and 95% non-parametric confidence intervals (CI) for the above mentioned percentiles as calculated from the model simulations.
- With the "-lloq" and "-uloq" option: the proportion of left and or right censored observations is presented for each IDV bin together with corresponding median and 95% CI based on the simulated data.
- With the "-levels" option: the proportion of observations in each category (defined by array following "-levels") and IDV bin is presented together with corresponding median and 95% CI based on the simulated data.
- Furthermore the output contains a numeric interpretation of the VPC diagnostic and statistics on false positives and false negatives. More about these statistics can be found in the user-guide. These functionalities are yet not adopted for the case when there is censored data ("-lloq" / "-uloq").

The file vpctab contains the original observed data and together with vpc_results.csv it is the standard input recognised by Xpose.

Basic Xpose functionalities for VPCs

Much more detailed information about the Xpose in general and the VPC plots specifically is available in the html help files distributed together with Xpose (http://xpose.sourceforge.net/specific_chm/xpose.VPC.html). The full range of functionalities for VPC plots in Xpose are only available with command line interface and not implemented in the menu system.

Basic examples:

```
## Reading in PsN results and observed data ##
```

```
vpc.file <- "vpc_results.csv" # reading PsN VPC output
```

```
vpctab <- "vpctab" # reading observed data
```

```
## Basic VPC plot for continuous data ##
```

```
xpose.VPC(vpc.info=vpc.file, vpctab=vpctab)
```

```
## Basic VPC plot for categorical data ##
```

```
xpose.VPC.categorical(vpc.info=vpc.file, vpctab=vpctab)
```

Additional options:

Pl.real = <NULL or TRUE> (not applicable for categorical)

Plot the percentiles of the real data in the various bins. Note that for a bin with few actual observations the percentiles will be approximate. For example, the 95th percentile of 4 data points will always be the largest of the 4 data points.

Pl.ci = <"both", "area" or "lines">

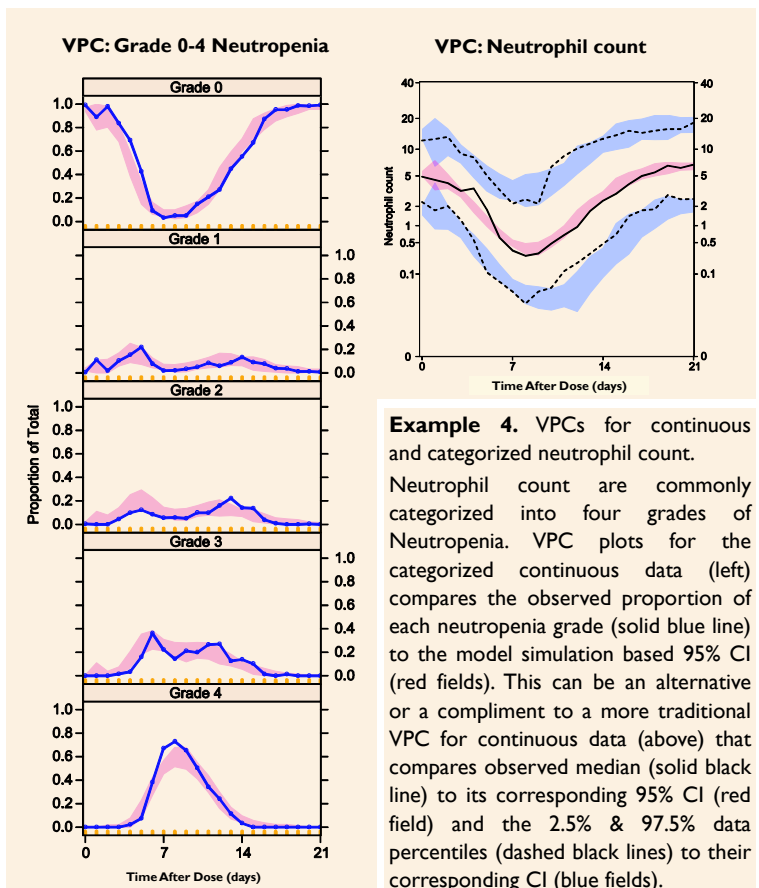
Plot the confidence interval for the simulated data's percentiles for each bin. Use area option to achieve colored fields as presented in the poster.

Pl.limits = c(0.025, 0.975) (not applicable for categorical)

A vector of two values that describe the limits of the prediction interval that should be displayed. These limits should be found in the 'vpc_results.csv' file.

Further more there is a full range of options for customising design aspects of the VPC plots.

Results – Example 4 “Categorized Continuous”



Example 4. VPCs for continuous and categorized neutrophil count.

Neutrophil count are commonly categorized into four grades of Neutropenia. VPC plots for the categorized continuous data (left) compares the observed proportion of each neutropenia grade (solid blue line) to the model simulation based 95% CI (red field). This can be an alternative or a complement to a more traditional VPC for continuous data (above) that compares observed median (solid black line) to its corresponding 95% CI (red field) and the 2.5% & 97.5% data percentiles (dashed black lines) to their corresponding CI (blue fields).