Predictive Performance of Internal and External Validation Procedures in Various Simulation Settings for a Simple PK Model

UPPSALA UNIVERSITET

Paul G. Baverel, Kristin K. Karlsson and Mats O. Karlsson
Dept of Pharmaceutical Biosciences, Uppsala University, Sweden

Background

- Internal (IV) and external (EV) validation procedures both provide information on how well a model will perform in predicting into the population from which the original study data were sampled from.

- Considerations when selecting a model-based diagnostic tool include:
  - Optimism is a potential problem of predictive models when internally validated and prediction errors may be underestimated.
  - For EV, splitting the dataset results in a smaller basis for estimating parameters and thus a lower precision when estimating parameters as well as prediction errors. The prediction errors are, however, expected to be unbiased.

Objectives

1. **Compare the predictive performance of internal (IV) and external (EV) procedures using the predictions of a separate large dataset from the same population as reference (population validation: PV):**
   - **for comparable learning data sizes (IV *vs.* EV data splitting)**
   - **when relaxing the normality assumption in the estimation step (NONP vs. FOCE)**
2. **Investigate the behaviour of nonparametric estimates (NONP) when predicting population datasets (PV) simulated from a multivariate normal distribution in comparison with FOCE estimates across a range of various data designs.**

Materials and Methods

A. Learning process (same for IV, EV, and PV)

PK Model:

oral one-compartment: random effect on Ka, CL, V (30% CV), RV (10 % CV)

Simulated Data:

- 55 simulated datasets for each data size (NM 6.2)
- Data sizes studied: 3,6,12,24,48,96,192, and 384 individuals
- 3 optimal sampling points at 0.2, 6, and 10 hrs post-dose (PopED[1] 2.0)

Re-estimation with FOCE and NONP

B. Validating process: Numerical Predictive Checks (NPCs)

Each set of model parameter estimates obtained in A. were used for predictions in 3 different automated NPC procedures (500 newly simulated datasets each) based on the following scheme:

- **Internal (IV):** predictions of the learning data statistics
- **External (EV):** predictions of an independent data statistics originating from the same population and of same size as learning (simulated another "learning" data with a different seed)
- **Population (PV):** predictions of an independent data statistics from 1000 individuals (large size)

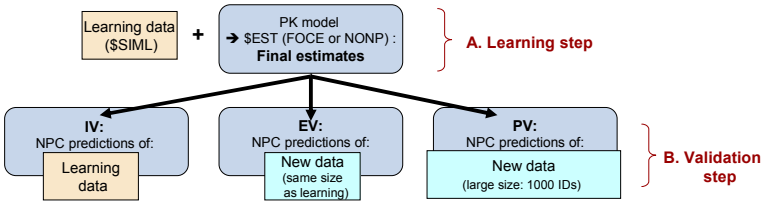


Figure 1. *Scheme of the two consecutive steps of internal (IV), external (EV), and population (PV) validation procedures. A Perl script, calling PsN[2], automates the entire procedure.*

C. Assessment of prediction errors[3] (bias and imprecision)

- between IV and EV (PV outcomes being used as reference)
- between PV outcomes obtained with FOCE and NONP estimates

Results

1. Predictive Performance of Internal (IV) & External (EV) Validations

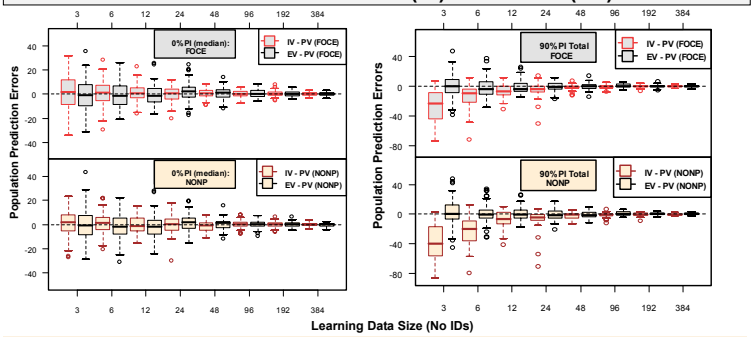


Figure 2. *Distributions of discrepancies between IV and PV (red), and EV and PV (black) in predicting the median (0% PI: Left) and the 90% PI (right) using FOCE (grey boxes) and NONP (light brown boxes) for various data designs.*

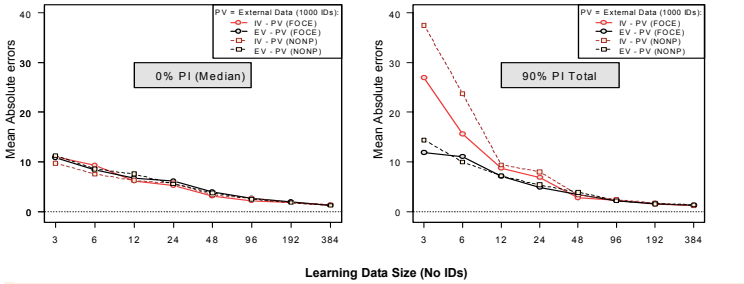
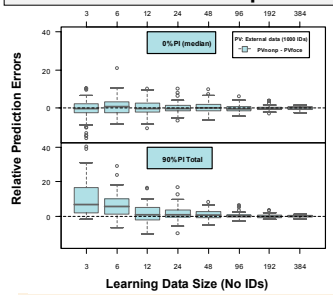


Figure 3. *MAEs of internal (IV) and external (EV) validation methods vs. learning data size.*

- At the median of observations/predictions:
  - No pronounced bias in prediction errors (Fig. 2)
  - Similar performance for IV and EV (Fig.3)
- At the extremes (90% PI) of the observations/predictions:
  - For small datasets, IV gives biased (optimistic) and imprecise estimates of the prediction errors, especially for NONP (Fig. 3)
  - Similar performance for IV and EV when ≥ 48 IDs

2. Discrepancies between NONP & FOCE :



- Similar FOCE and NONP predictions at the median of observations regardless of data informativeness
- FOCE predictions of extreme observations (90% PI) superior to NONP predictions when datasets are small (and underlying random effects distribution is normal)

Figure 4. *Relative prediction errors of NONP vs. FOCE estimates when predicting the population data statistics (median (top) and tails of distribution (90% PI, bottom)).*

Conclusions

- Here IV and EV were compared for similar sized learning datasets. In reality, the choice is often to perform IV on the full dataset or split the dataset and do model building and parameters estimation on one part, and EV on the other part. If the aim is to achieve precise predictions in the target population, the results here point to a preference for data splitting followed by EV when the dataset is small, and preference for IV when the dataset is large.

- Nonparametric models had higher prediction errors for the tails of the distribution in small datasets, but otherwise similar to parametric models.

References:

1. PopED version 2.08 (2008). http://poped.sf.net
2. Perl-speaks-NONMEM (PsN software): http://psn.sourceforge.net
3. Sheiner LB, Beal SL. Some suggestions for measuring predictive performance. (1981)