

Maintaining Consistency between a Substance Database and single Study Databases

Dieter Hartz, Boehringer Ingelheim Pharma GmbH & Co. KG ¹

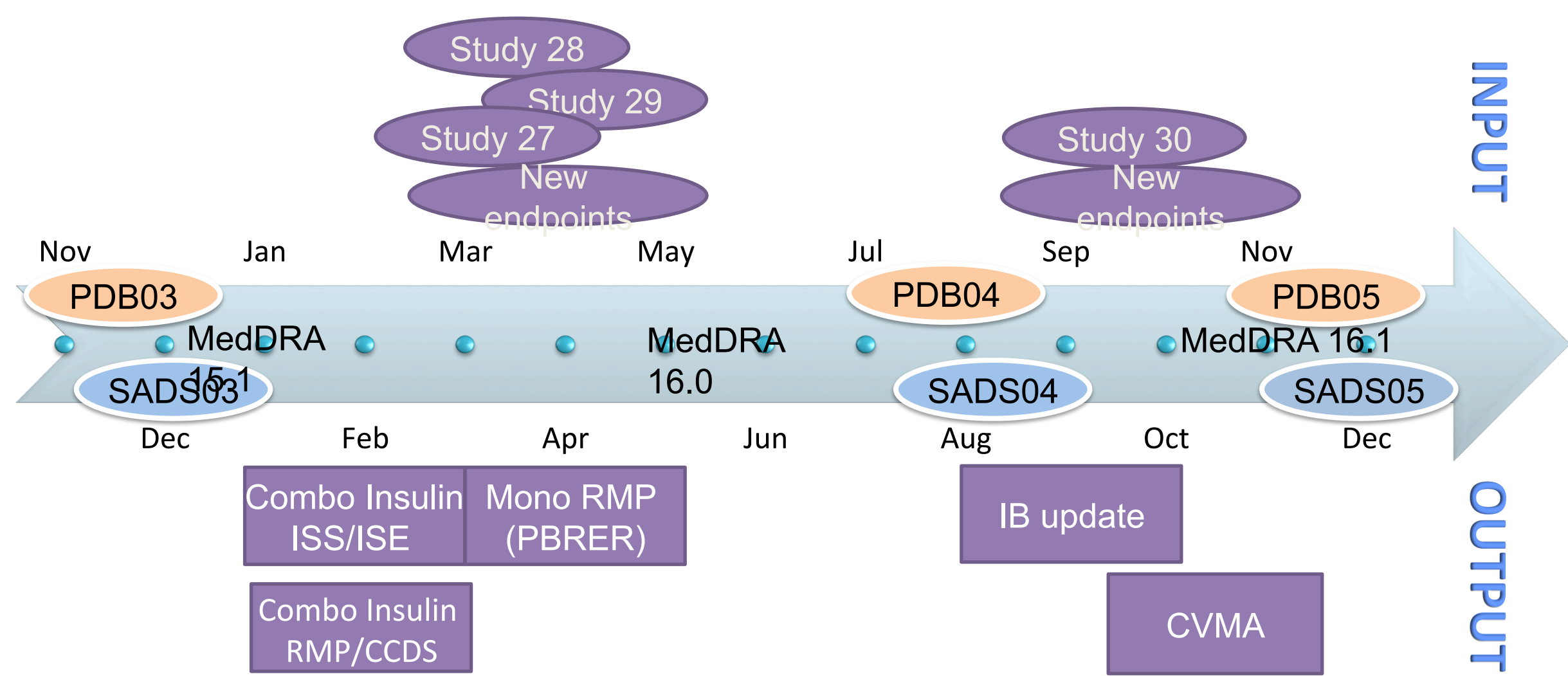


Boehringer
Ingelheim

1. Introduction and Objectives

- A substance database for a DPP-4 inhibitor against type II diabetes has been established for pooled analyses.
- A snapshot of this database consists of substance analysis datasets (SADS) and source datasets at the appropriate data lock points in a project database (PDB).
- About 30 completed studies are contained in the current snapshot.
- As needed, the database is being expanded with data of newly completed studies and stored as a new snapshot.
- Each snapshot is used for a defined set of analysis tasks.
- For reproducibility of previous analyses the consistency between different snapshots as well as between study data and snapshot data should be maintained.

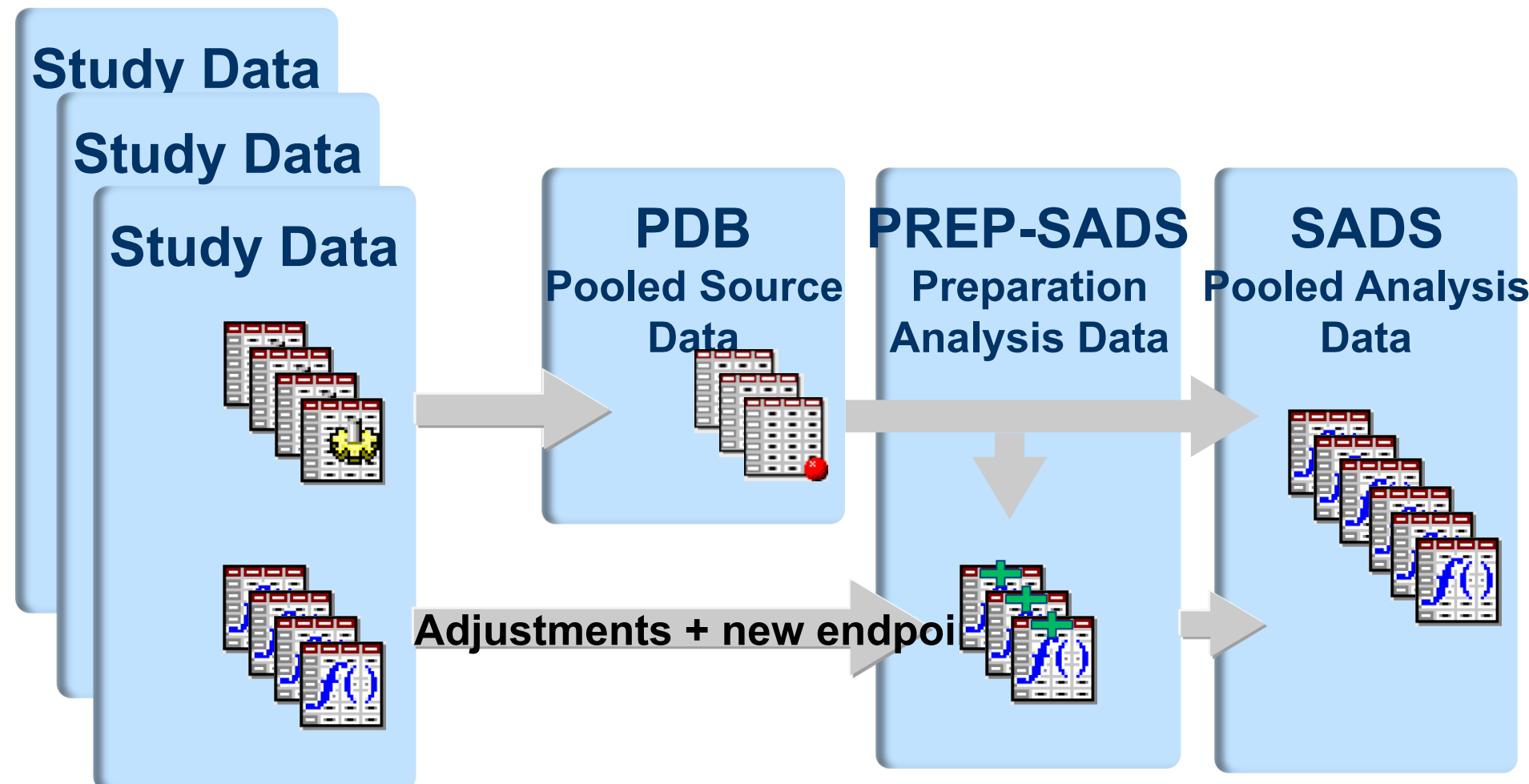
2. Substance Database Snapshots



- As needed, the substance database (SADS + PDB) is being expanded with data of newly completed studies and stored as a new snapshot. If new endpoints are required these are being derived across studies.

3. Main Challenges and Solutions

Relationship of Study Data, PDB, SADS



- The consistency of **safety data** largely depends on the consistency of PDB datasets. Analyses are directly done from PDB source datasets (AE, LAB) using BI internal standard macros or from SADS datasets newly derived from PDB source datasets for each snapshot (SADS containing treatment information, rescue information, analysis populations).
- The consistency of **efficacy data** largely depends on study analysis datasets because efficacy endpoints are directly taken from study efficacy datasets without re-derivation of study endpoints. Only new endpoints will be added or new imputation methods will be implemented in each snapshot.

Managing Hardcodings

- Elimination of double screened patients by hardcoding.

EXCEL file hardcoding.xls	Macro call %HC generates SAS® dataset HCODE with hardcodings	Macro call %HC in program for SADS POPU invokes the code associated with ID												
<table><thead><tr><th>ID</th><th>Code</th><th>Approval</th></tr></thead><tbody><tr><td>POPU0001</td><td>IF study EQ '1' AND ptno EQ 2 THEN ynschr=.</td><td>YES</td></tr><tr><td>POPU0002</td><td>IF study EQ '1' AND ptno EQ 5 THEN ynschr=.</td><td>YES</td></tr><tr><td>POPU0003</td><td>IF study EQ '1' AND ptno EQ 10 THEN ynschr=.</td><td>NO</td></tr></tbody></table>	ID	Code	Approval	POPU0001	IF study EQ '1' AND ptno EQ 2 THEN ynschr=.	YES	POPU0002	IF study EQ '1' AND ptno EQ 5 THEN ynschr=.	YES	POPU0003	IF study EQ '1' AND ptno EQ 10 THEN ynschr=.	NO	<pre>%HC(in=file.xls);</pre> Export	<pre>Data POPU; %HC(POPU0001); %HC(POPU0002); %HC(POPU0003); ... IF ynschr=. THEN DELETE;</pre> sads-popu.log <pre>ERROR: Approval missing ID=POPU0003</pre>
ID	Code	Approval												
POPU0001	IF study EQ '1' AND ptno EQ 2 THEN ynschr=.	YES												
POPU0002	IF study EQ '1' AND ptno EQ 5 THEN ynschr=.	YES												
POPU0003	IF study EQ '1' AND ptno EQ 10 THEN ynschr=.	NO												

- The employment of the hardcoding tool allows for easy documentation and traceability of the hardcodings.

Which Formula is used for eGFR?

- For pooled analyses, eGFR (= estimated Glomerular Filtration Rate) calculation may employ various formulas depending on the terms of creatinine determination and whether a patient is Japanese or not, while in many studies the original MDRD (= Modification of Diet in Renal Disease) formula was employed.

Original MDRD formula: EGFR

$$\text{eGFR (mL/min/1.73m}^2\text{)} = 186 \times \text{creatinine (mg/dL)}^{-1.154} \times \text{age}^{-0.203} \times (0.742 \text{ if female}) \times (1.212 \text{ if black})$$

Creatinine measurement not calibrated to be IDMS traceable

New formula: EGFRJ

$$\text{eGFR (mL/min/1.73m}^2\text{)} = 175 \times \text{creatinine (mg/dL)}^{-1.154} \times \text{age}^{-0.203} \times (0.742 \text{ if female}) \times (1.212 \text{ if black})$$

Creatinine measurement calibrated to be IDMS traceable

Japanese formula: EGFRJ

$$\text{eGFR (mL/min/1.73m}^2\text{)} = 194 \times \text{creatinine (mg/dL)}^{-1.094} \times \text{age}^{-0.287} \times (0.739 \text{ if female})$$

used for Japanese patients only

eGFR
(estimated
Glomerular
Filtration
Rate)

Pooled Analysis: EGFRJ

- Selection per patient of
- EGFRJ (Japanese)
 - EGFRJ (creatinine IDMS traceable)
 - EGFR (creatinine non-IDMS traceable)

- To maintain reproducibility of study results for eGFR the different formulas were assigned to different parameter names (EGFRJ, EGFRJ, EGFR). For studies that had originally used only EGFR also EGFRJ was derived because for the majority of them the new formula was to be applied. For Japanese patients also EGFRJ was derived. The selected eGFR per patient in the pooled analysis was then captured in a new parameter (EGFRJ).

4. Further Challenges and Solutions

Challenge	Solution
The periodic update of dictionaries leads to slight inconsistencies between PDB source datasets for each snapshot with respect to coded information.	PDB safety datasets also contain LLT coding information from previous and originally used MedDRA/WHO-DD versions so that original outputs could be reproduced.
Changes in eCRFs might lead to new variables in PDB source dataset while other variables might become obsolete.	Variables becoming obsolete in PDB source datasets were renamed and kept as such. Contents were re-mapped to the new variables.
Introduction of new standard units for laboratory parameters leads to change of values for parameters concerned.	In PDB the SI standard units from new study data were converted back into conventional units so that previous outputs could be reproduced.
The request of additional imputation methods for efficacy endpoints might also demand for the re-derivation of already existing efficacy endpoints.	When re-derivation of existing imputations was necessary baseline selection rules employed in the studies were also employed in SADS.
Derivation of new efficacy endpoints from parameters also reported in safety analyses.	Data were crosschecked with data derived for safety analyses, especially for consistency in selected baseline values .

5. Conclusions

The maintenance of consistency between a pooled substance database and study data supports the long term goal of reproducibility of previously reported analyses from newly pooled data. This was mainly achieved by maintaining the endpoints from the single studies while in addition deriving new endpoints for analyses across studies that employ different algorithms. The concept of building defined snapshots of the substance database for reporting also demands consistency checking between snapshots. While at first this appears to be extra work, it allows the easy detection of unexpected changes so that erroneous reporting is avoided. Furthermore it is possible to prepare analyses from one snapshot while working at the same time on the generation of the next snapshot.

Acknowledgements:
Michael Knoessl (BI), Felix Mader (BI), André Haugg(BI), Ralf Minkenber(BI), Barbara Doll (BI)

1 Contact Information:
Dieter Hartz, Boehringer Ingelheim Pharma GmbH & Co. KG (BI), Binger Strasse 173, 55216 Ingelheim am Rhein, Phone: +49-6132-77-142500, E-Mail: dieter.hartz@boehringer-ingelheim.com