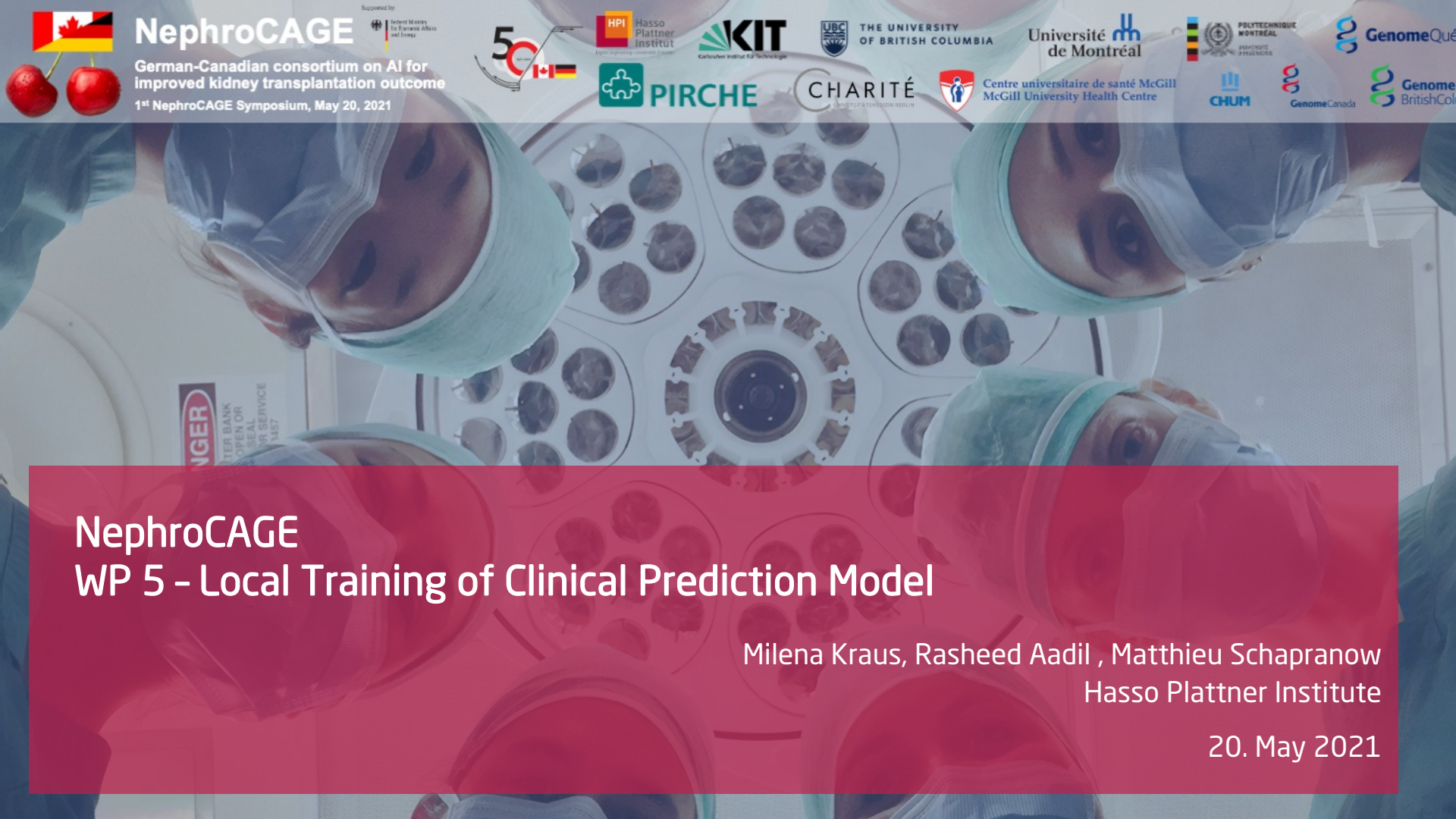


NephroCAGE

German-Canadian consortium on AI for improved kidney transplantation outcome

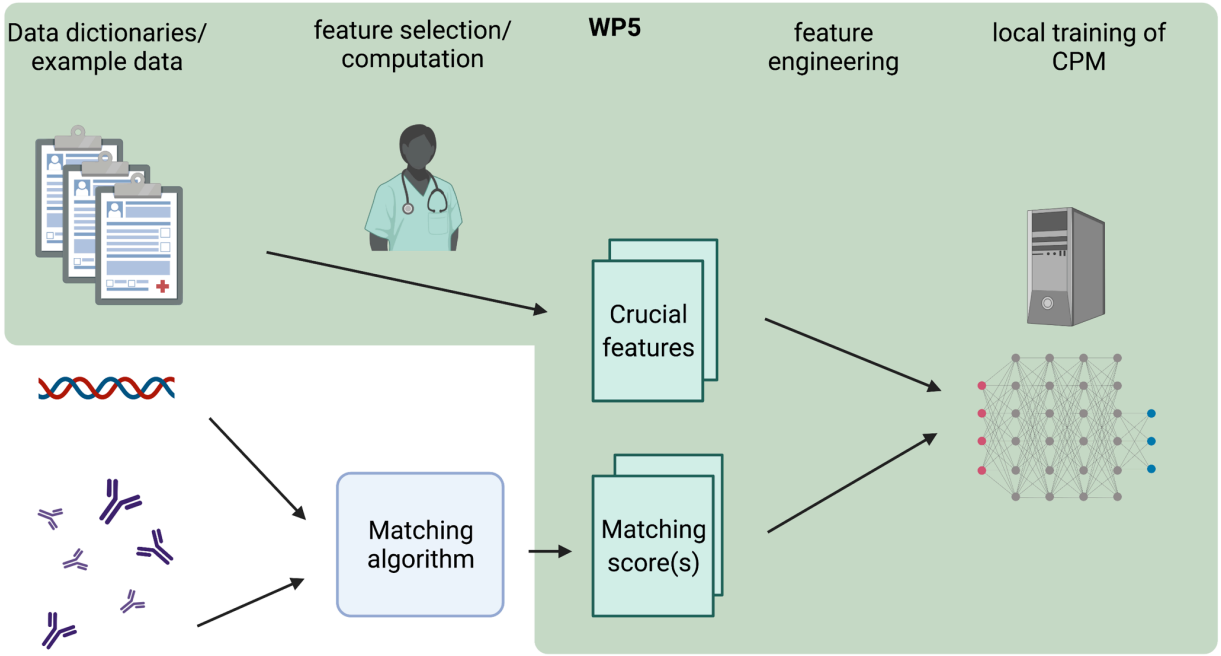
1st NephroCAGE Symposium, May 20, 2021

NephroCAGE
WP 5 – Local Training of Clinical Prediction Model

Milena Kraus, Rasheed Aadil , Matthieu Schapranow
Hasso Plattner Institute

20. May 2021

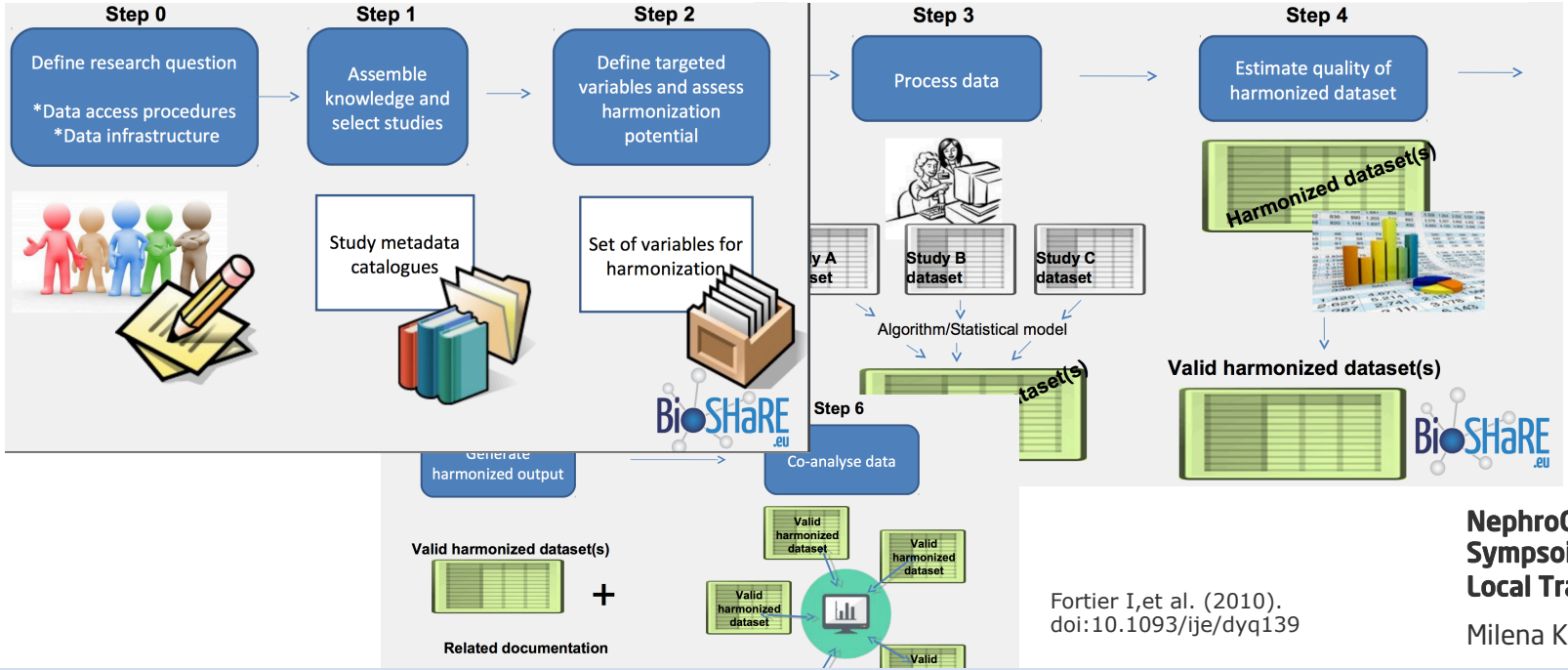
Local Training of Clinical Prediction Model
WP5 - Overview



NephroCAGE Sympsoium, WP5 - Local Training of CPM

Milena Kraus

Retrospective Data Harmonization Steps
The DataShaper Approach



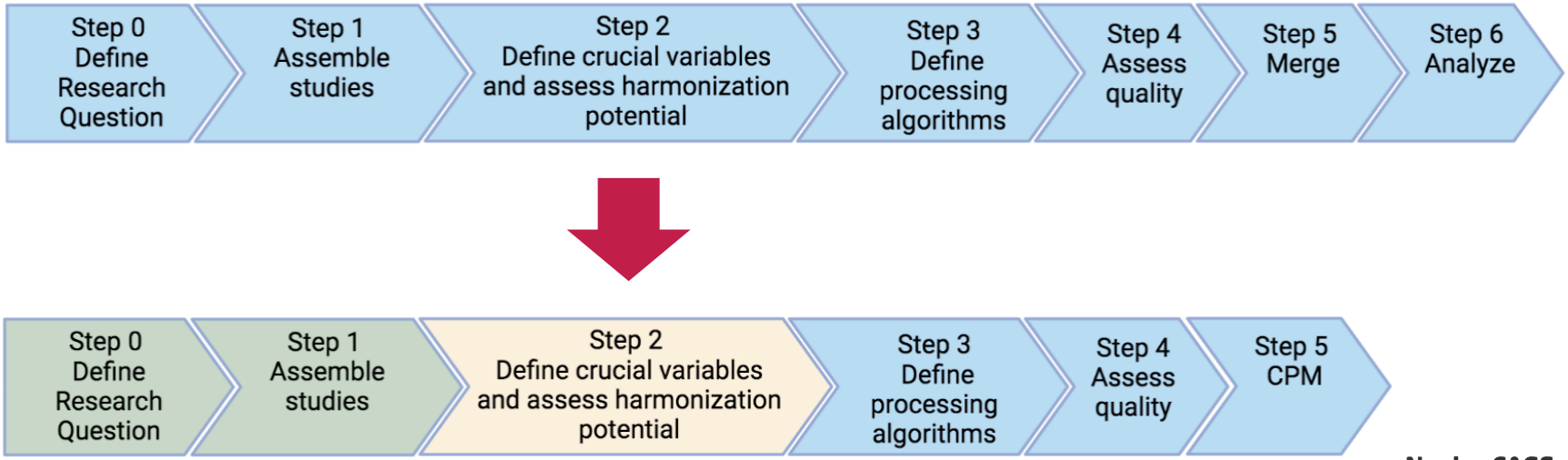
Fortier I,et al. (2010).
doi:10.1093/ije/dyq139

NephroCAGE Sympsoium, WP5 - Local Training of CPM

Milena Kraus

Methodology
Federated learning version of the DataShaper approach

Hasso Plattner Institut



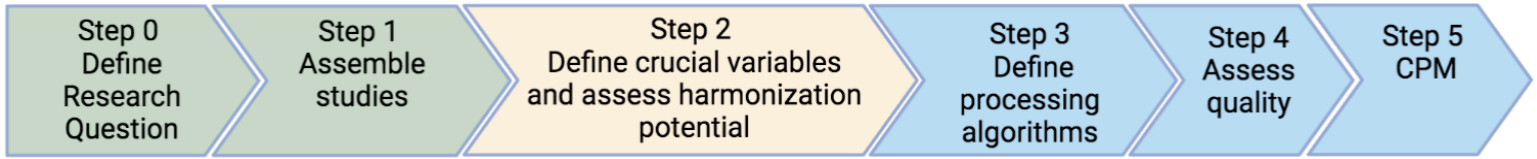
NephroCAGE Sympsoium, WP5 - Local Training of CPM

Milena Kraus

Methodology
Federated learning version of the DataShaper approach

RQ: What is the current risk of graft failure for a given patient based on current and historical data?

1. Tbase/Berlin/Charite, 2. Montreal/CanPrevent, 3. Vancouver/PROMIS

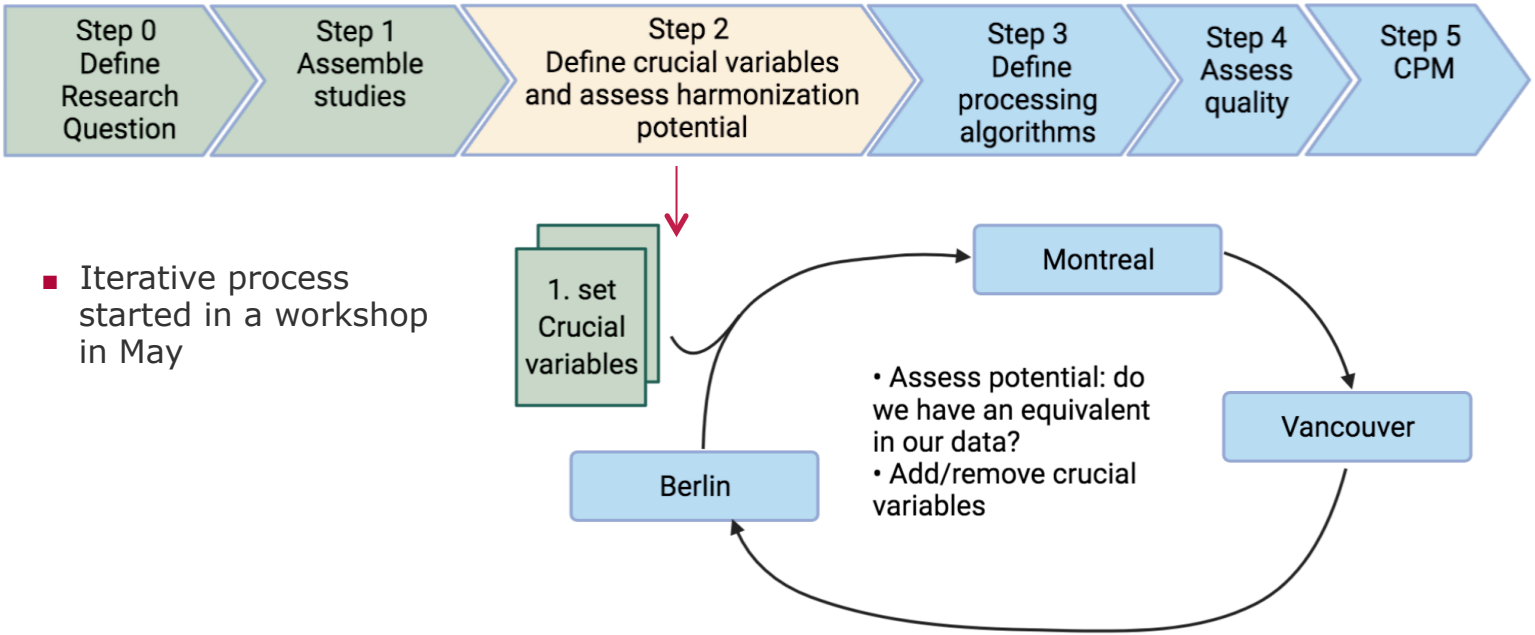


NephroCAGE Sympsoium, WP5 - Local Training of CPM

Milena Kraus

Methodology
Federated learning version of the DataShaper approach

Hasso Plattner Institut



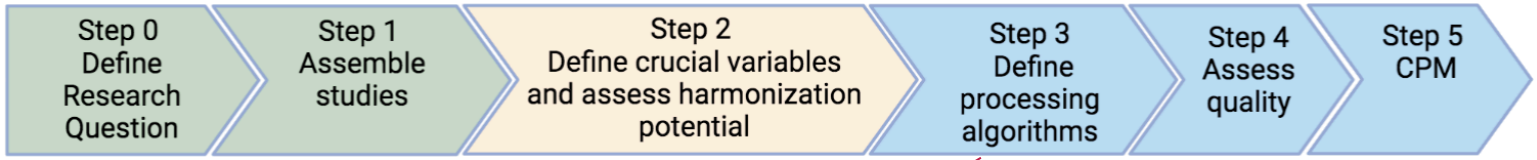
- Iterative process started in a workshop in May

NephroCAGE Sympsoium, WP5 - Local Training of CPM

Milena Kraus

Data Transformation and Quality Control
Step 3 + 4

Hasso Plattner Institut

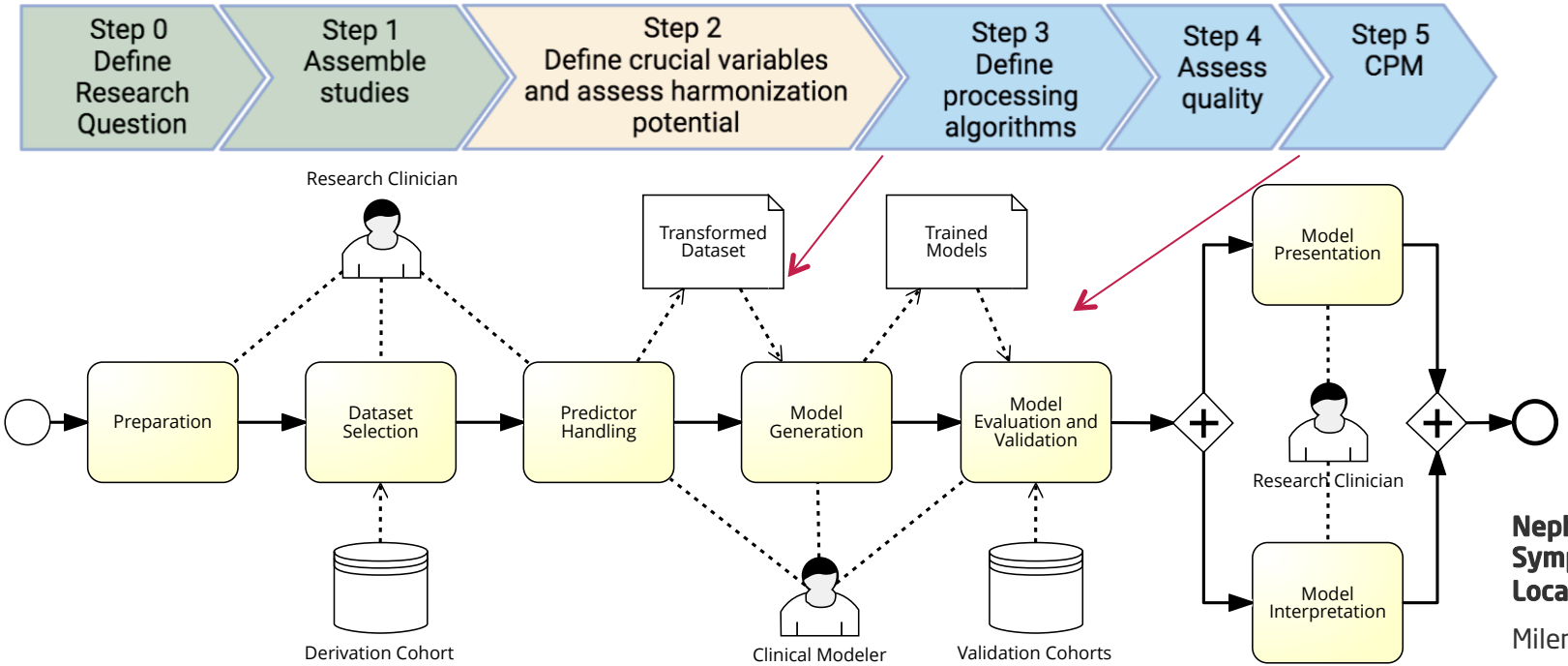


- What is the desired variable format?
  - Is the format capturing the semantics correctly?
  - Is the format optimal for usage in a CPM?
- Assess quality:
  - Prepare visualizations of variables
  - Compare distributions
  - Clarify source of outliers, differences in distributions etc.

NephroCAGE Sympsoium, WP5 - Local Training of CPM

Milena Kraus

Predictive Modeling Process



NephroCAGE Sympsoium, WP5 - Local Training of CPM

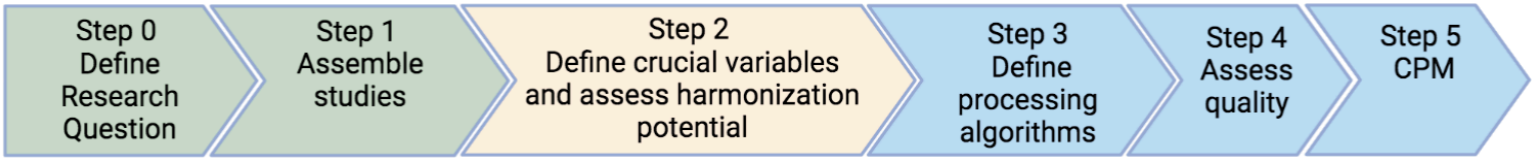
Milena Kraus

Feature Engineering/Predictor Handling

- Selected techniques:

Predictor Handling	Data Imputation [44, 87]	Mean Imputation Soft imputation k-Nearest Neighbors
	Data Sampling [97]	Synthetic Minority Over-sampling Technique Random Over Sampler Random Under Sampler Cluster Centroids
	Data Encoding [44]	Label Binarizer Label Encoding One Hot Encoding Ordinal Encoding
	Data Scaling [44]	Standard Scaler Robust Scaler Normalizer Quantile Transformer

Clinical Prediction Models
Classical Approaches



Probability of outcome = *f*(predictor variables)

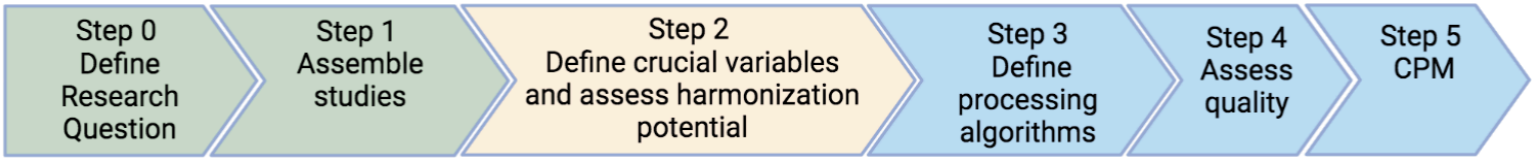
$$Pr(Y = 1) = f(X)$$

NephroCAGE Sympsoium, WP5 - Local Training of CPM

Milena Kraus

Clinical Prediction Models
Simple Example - Regression

Hasso Plattner Institut



The formula applied in the iBox cohort was therefore the following one:

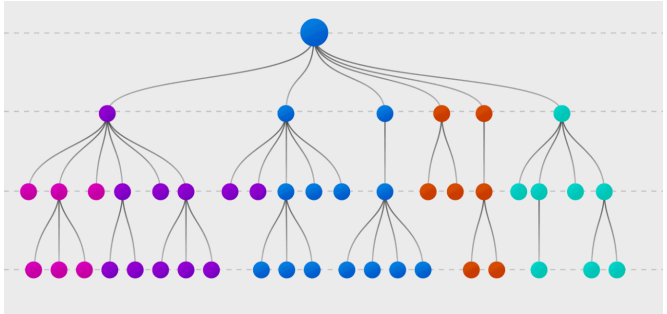
$$\text{NBMM}_{\text{iBox cohort}} = \beta_{\text{UACR}} * (\log_{10}(\text{UACR}-46)) + \beta_{\text{eGFR}} * ((\text{eGFR}-47)/10) + \beta_{\text{eGFR}^2} * ((\text{eGFR}-47)/10)^2 + \beta_{\text{Rejection}} * \text{Rejection} + \beta_{\text{Black ethnicity}} * \text{Black ethnicity} + \beta_{\text{Recipient Age}} * ((\text{Recipient age}-46)/10) + \beta_{\text{UACR with rejection interaction}} * (\log_{10}(\text{UACR}-46) - 0.46 \text{ for rejection}) + \beta_{\text{g score}} * g + \beta_{\text{ci score}} * ci$$

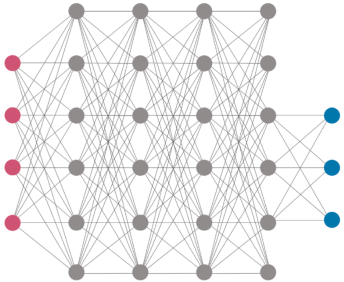
NephroCAGE Sympsoium, WP5 - Local Training of CPM

Milena Kraus

More complex CPMs

- Machine learning
  - Random forests
  - Gradient-boosting decision tree
  - Support Vector Machines
  - ..
- Deep learning
  - Multi-layer perceptron
  - Convolutional neural networks
  - ...

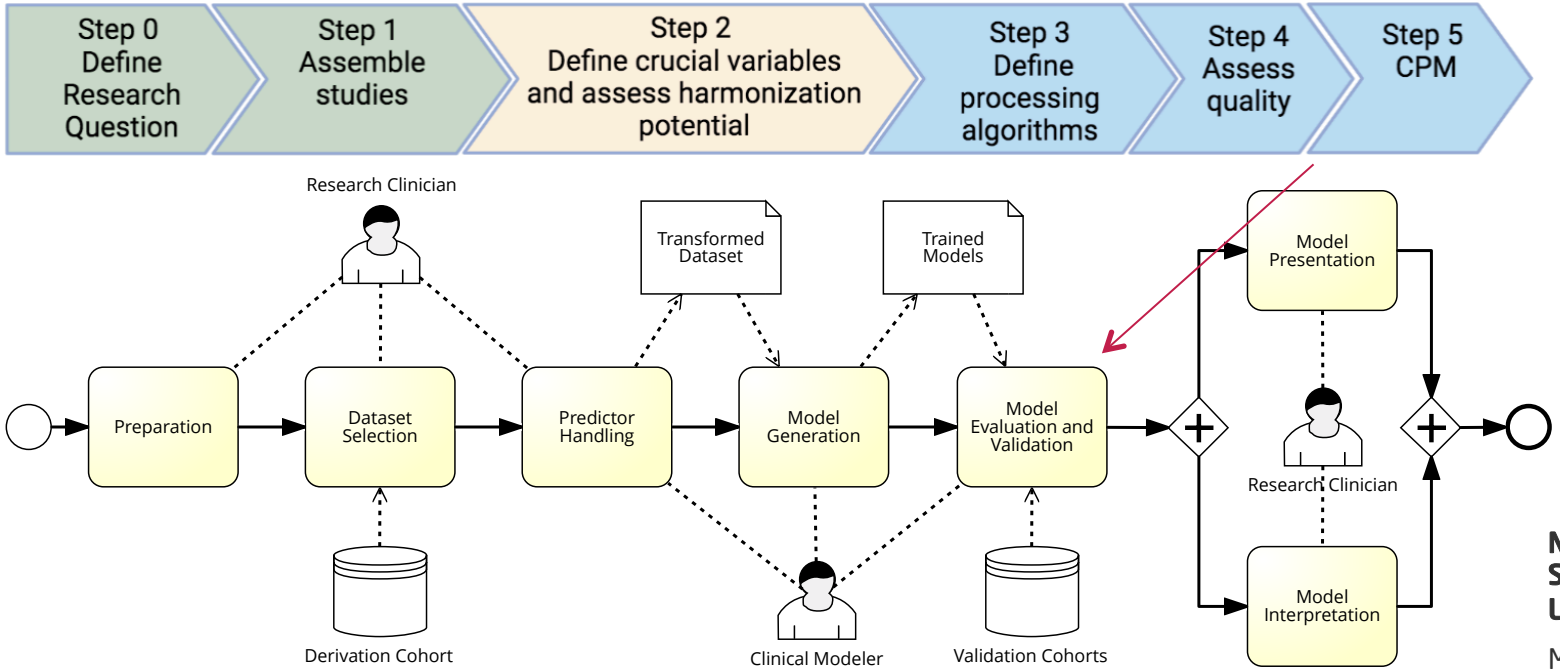




NephroCAGE Sympsoium, WP5 - Local Training of CPM

Milena Kraus

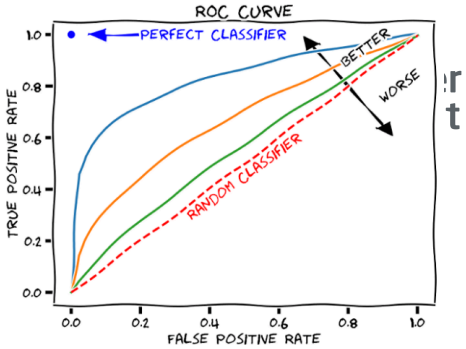
Predictive Modeling Process



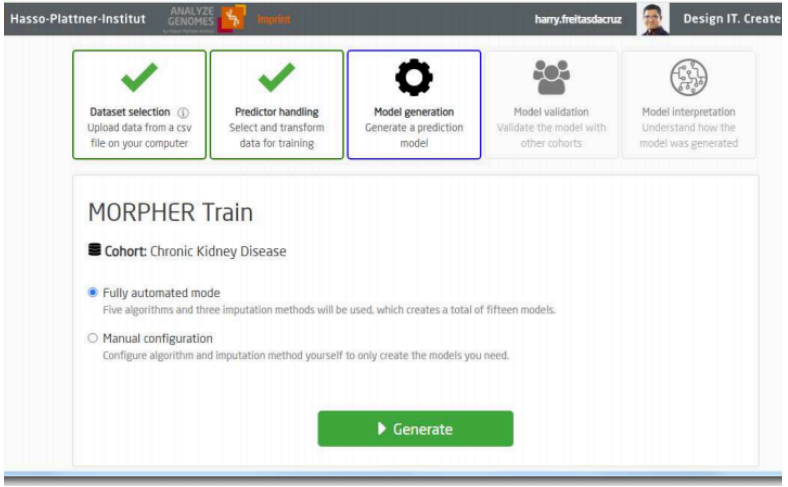
NephroCAGE Sympsoium, WP5 - Local Training of CPM

Milena Kraus

Evaluation of Pre-training



- Model training, internal validation and hyperparameter tuning
- Evaluation criteria for (quasi-)continuous values:
  - Receiver operating curve (ROC)
  - Area under the curve (AUC) with confidence interval
  - Predictiveness curve (Pepe et al. 2008)
  - Calibration metrics
  - Clinical usefulness
- Validation and recalibration at different sites (→ WP6)



Expected Artifacts after Completion of WP5

- Fully defined set of crucial variables encompassing
  - A definition of their data type/format
  - Transformation rules for site-specific formats → CPM expected formats
  - A quality check assuring proper similarity in semantics and distribution
  - Further feature engineering rules, i.e., missing data imputation, normalization
- Pre-trained CPM encompassing
  - The selected model and pre-training parameters
  - Calibration and evaluation metrics
  - Selected features for CPM

NephroCAGE Sympsoium, WP5 - Local Training of CPM

Milena Kraus

Thank you for listening!

- Any questions on work package 5?