Machine learning models for predicting human *in vivo* PK parameters using chemical structure and dose

Filip Miljković

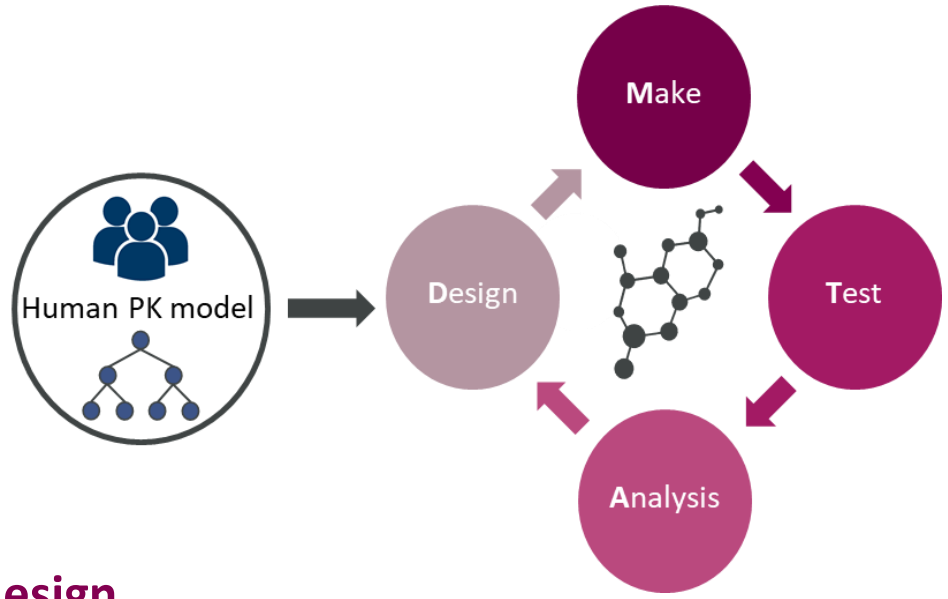
Imaging and Data Analytics, Clinical Pharmacology & Safety Sciences, R&D, AstraZeneca, Gothenburg, Sweden

3rd *In Silico* Toxicology Conference 2022

September 29, 2022

Mission: Machine learning tools to guide early discovery

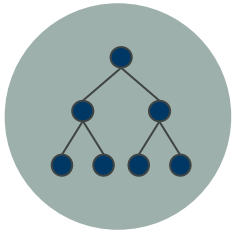
- **Develop *in silico* models for prediction of *in vivo* PK from chemical structure**
- Drive prioritization of compounds for *in vivo* assays
- Design compounds with better safety and PK properties
- Improve speed and efficiency in DMTA cycle



Ultimate goal - enable human PK prediction at the point of design

Content

Background & Data Curation



Modelling & Results

Future Applications & Conclusion

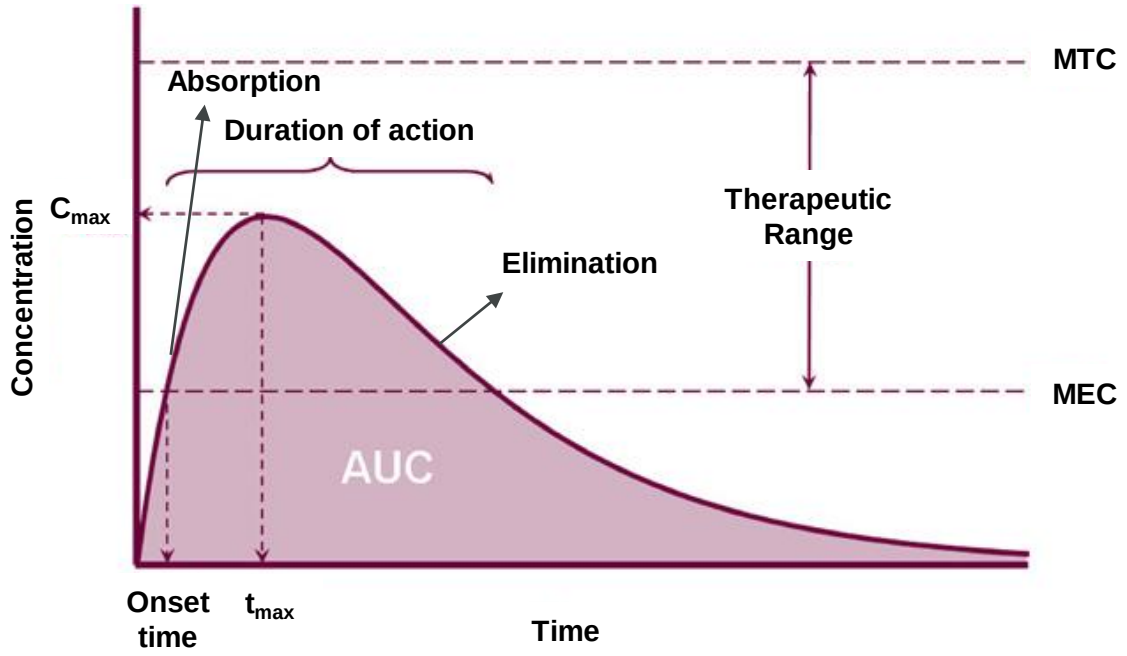
PK is important to achieve the “Right tissue” and “Right safety”

Absorption: Oral drug delivery projects strive for high oral absorption and bioavailability (F) to achieve optimal *in vivo* exposure

Distribution: Volume (V_d IV) is a measure of the extent of distribution and binding of a compound in organs and tissues

Metabolism & Elimination: Projects strive for low clearance (CL) to achieve acceptable duration of target engagement

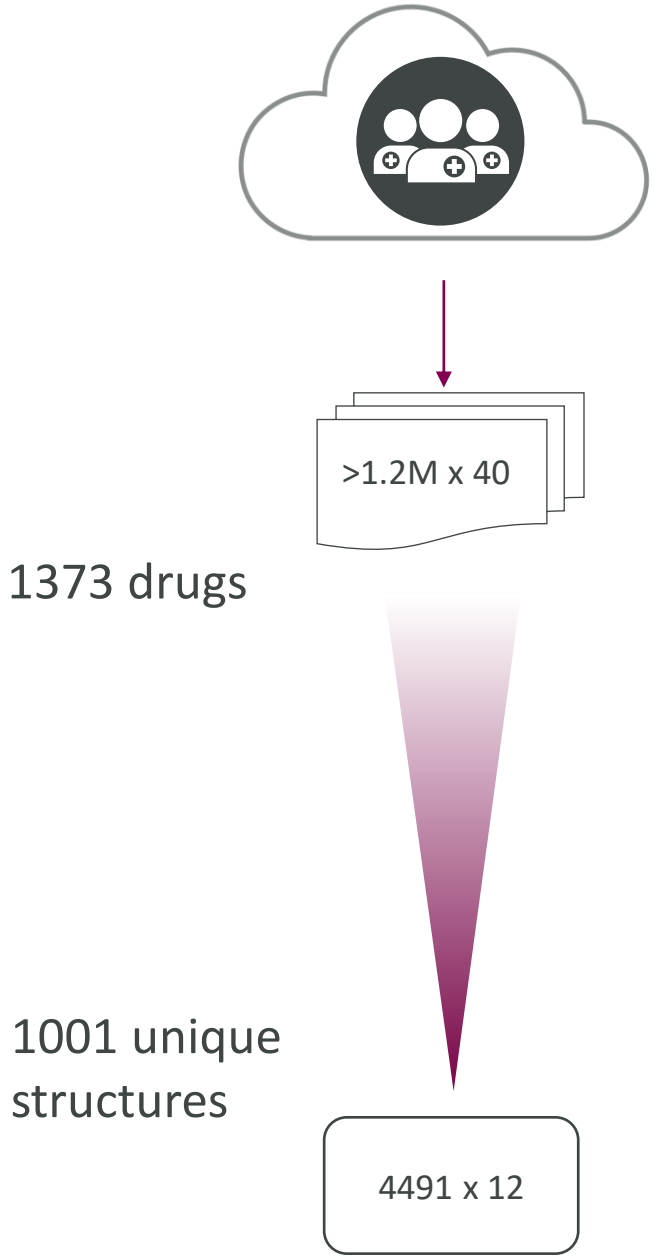
C_{max} within the therapeutic range to achieve efficacy and the right safety



Data extraction and curation

Aggregated human clinical data was collected and cleaned based on expert opinion

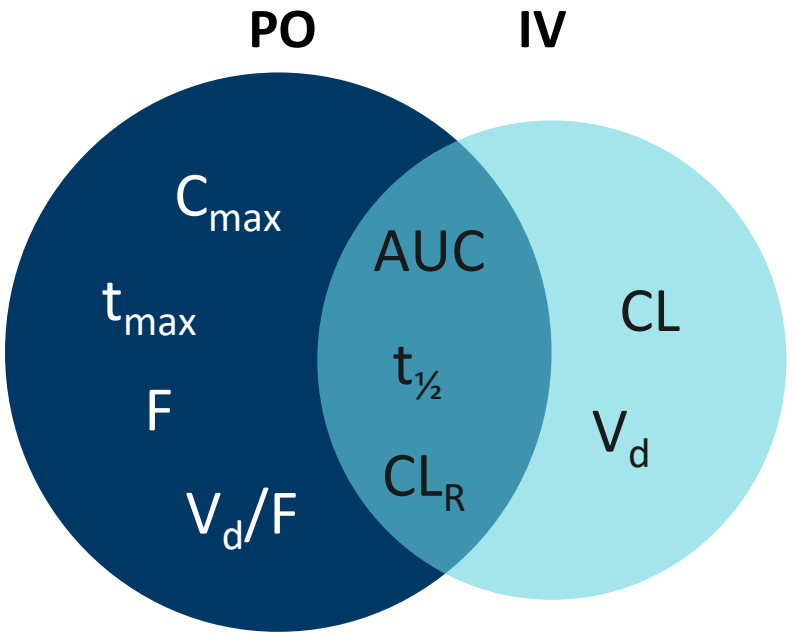
- **Included:**
  - Adult healthy volunteers and patient groups (oncology)
  - Single dose PK for compounds with Mw ≤750 Da
  - Non-tissue specific measurements
- **Excluded:**
  - Inadequate patient groups with signs of decreased PK functionality
  - Incomplete data points
  - Radiolabeled compounds
  - Metabolites and enantiomers
  - Incompatible assay technologies
  - Concomitants with potential impact on PK



Parameter selection and data aggregation

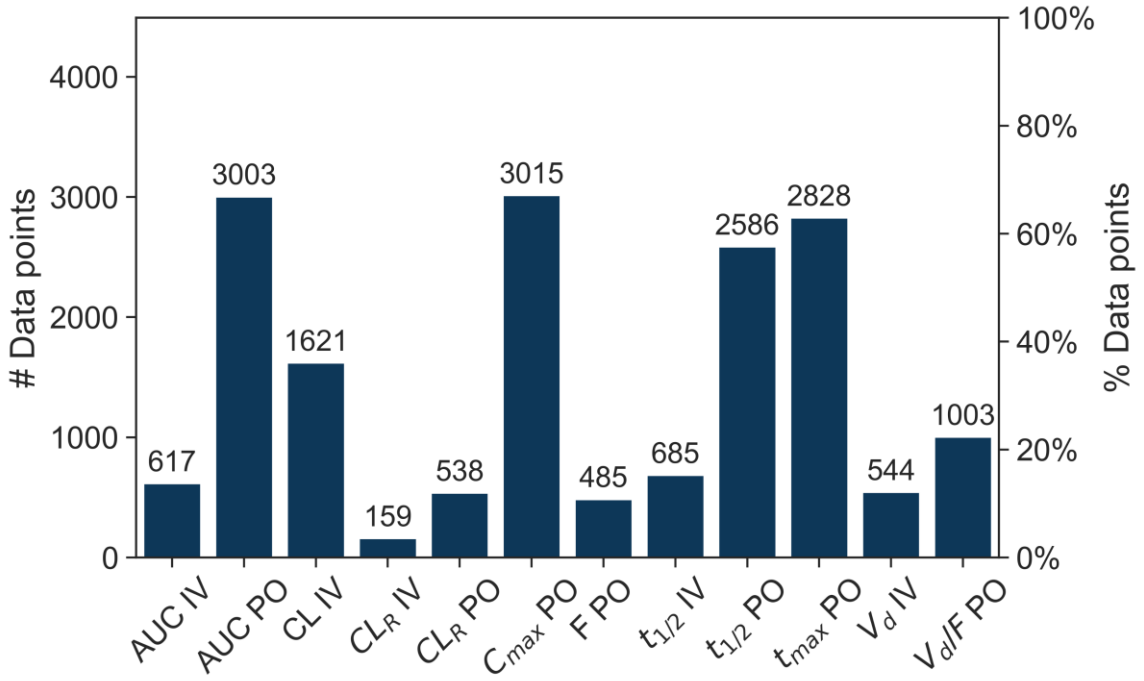
- Nine PK parameters - 12 PO/IV variants
- Data aggregation: median value for all compound-dose combinations
- Doses rounded to two decimals (below 10mg) or zero decimals (above 10mg)

Final data set:
4491 compound-dose combinations
for 12 PK parameters (1001 SMILES)



Data availability differs depending on parameter

- Different levels of data completeness
  - CL_R IV (3.5%) - C_{max} PO (67.0%)
- PO parameters have more data points than IV parameters
- The data is biased to optimized compounds with good PK profiles
  - E.g. high F and low CL IV
- Limited data for model building for certain PK parameters



Four feature designs used for model building and comparison

Dose

Used as a single descriptor to investigate its predictiveness compared to other, complex features

ADME/PK

In silico predictions from common ADME and *in vivo* rat PK assay models

2D descriptors

Around hundred most common 1D and 2D descriptors (OESelma)

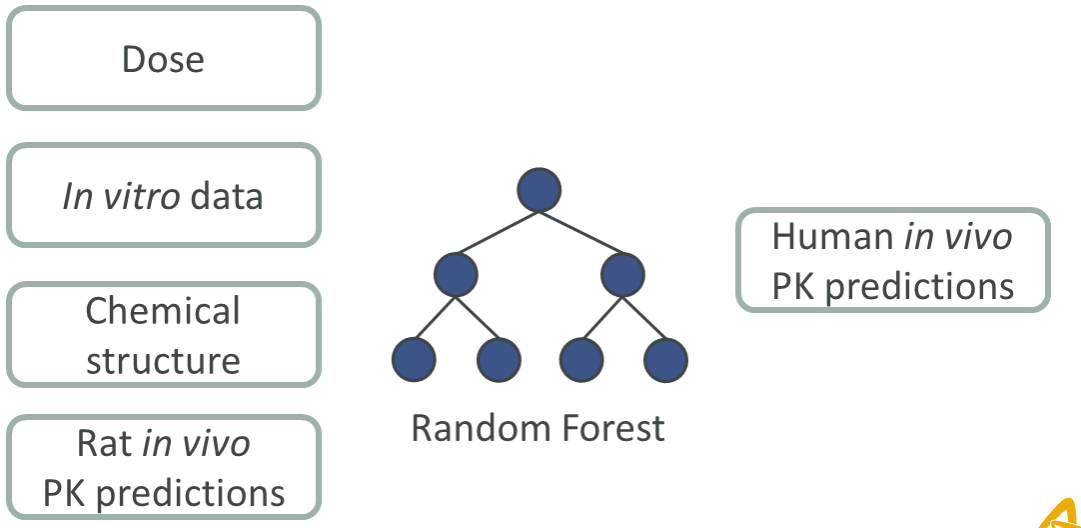
ComboFP

Concatenated version of all three designs (*Dose* + *ADME/PK* + *2D descriptors*)

Random Forest was chosen as the modelling method

- Random Forest was used as the modelling technique
- Five-fold cross-validation & hyperparameter optimization
- Custom data split for validation and test sets to ensure no overlapping compounds between the sets (20% hold-out)

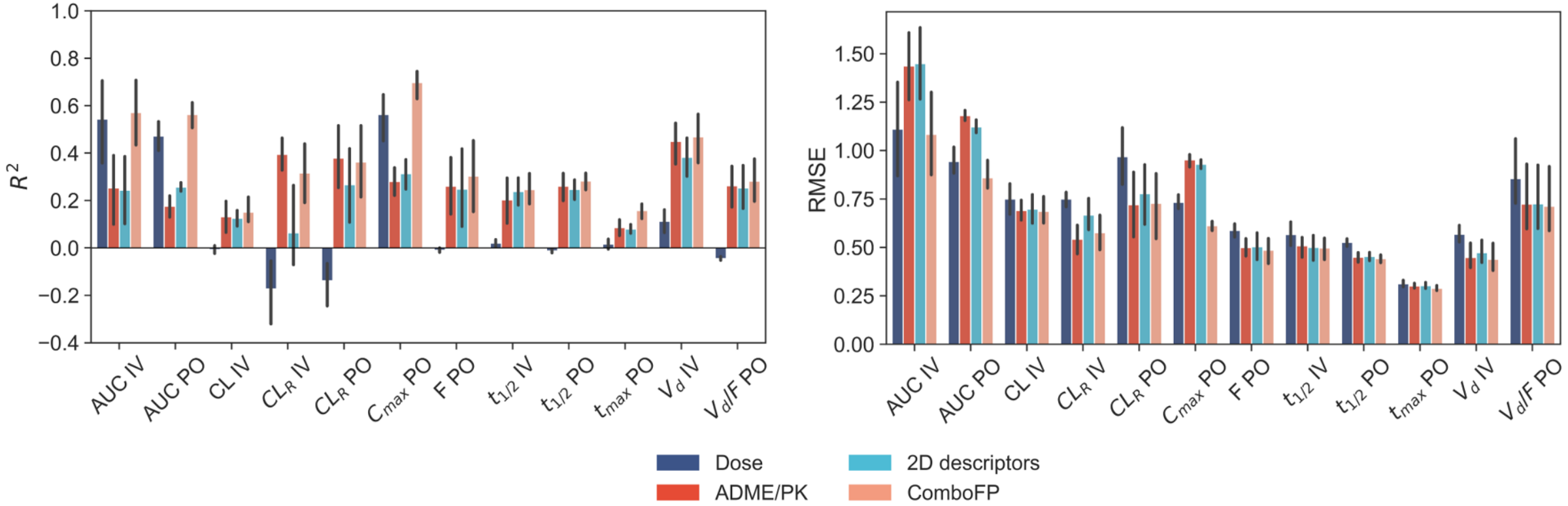
***In vitro* ADME and PhysChem properties, *in vivo* rat PK parameters**	
LogD	Human hepatocyte intrinsic clearance
Solubility (dried DMSO)	Rat plasma protein binding
Caco-2 intrinsic permeability	Human plasma protein binding
Caco-2 efflux ratio	Fraction unbound in rat hepatocytes
Human liver microsome intrinsic clearance	*In vivo* rat PK parameters: F, CL, C_{max} PO, $t_{½}$ IV, V_{ss} IV



Cross-validation results for all descriptor sets

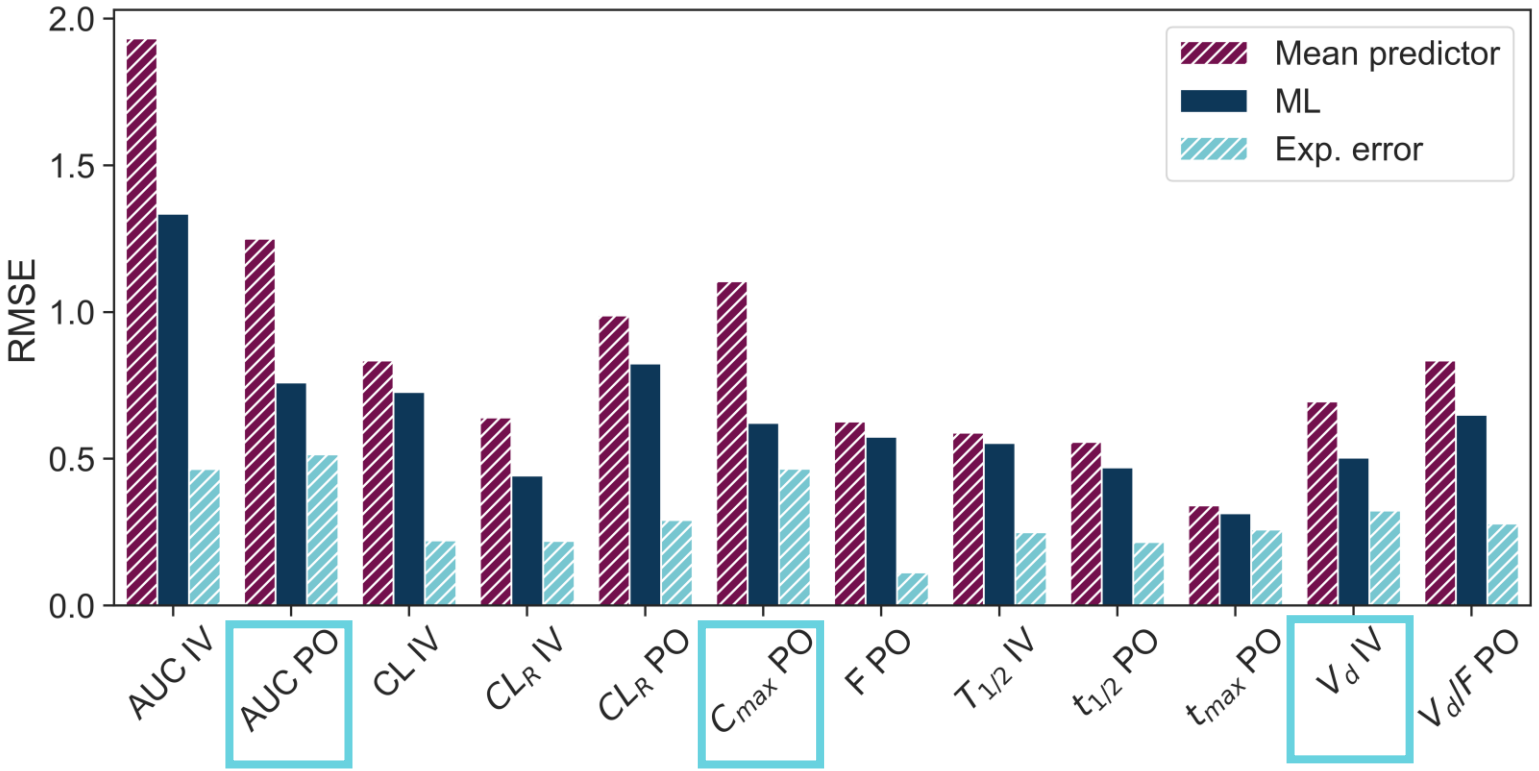
ComboFP

Best performing descriptor set



Hold-out test set modelling results for ComboFP

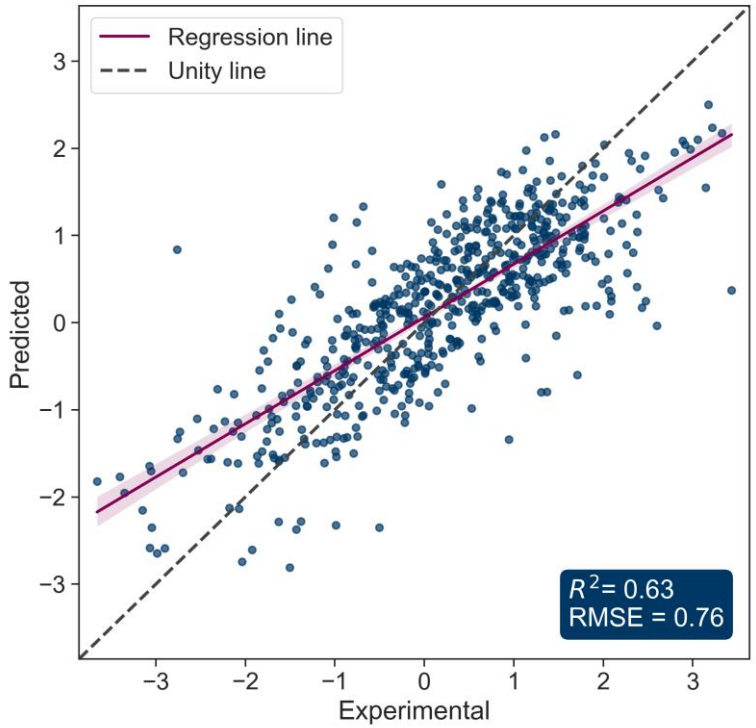
- Varying data distributions and data availability impact the ability to model endpoints
- Benchmarking against mean predictor and experimental error
- **Three endpoints have satisfactory models: AUC PO, C_{max} PO, and V_d IV**



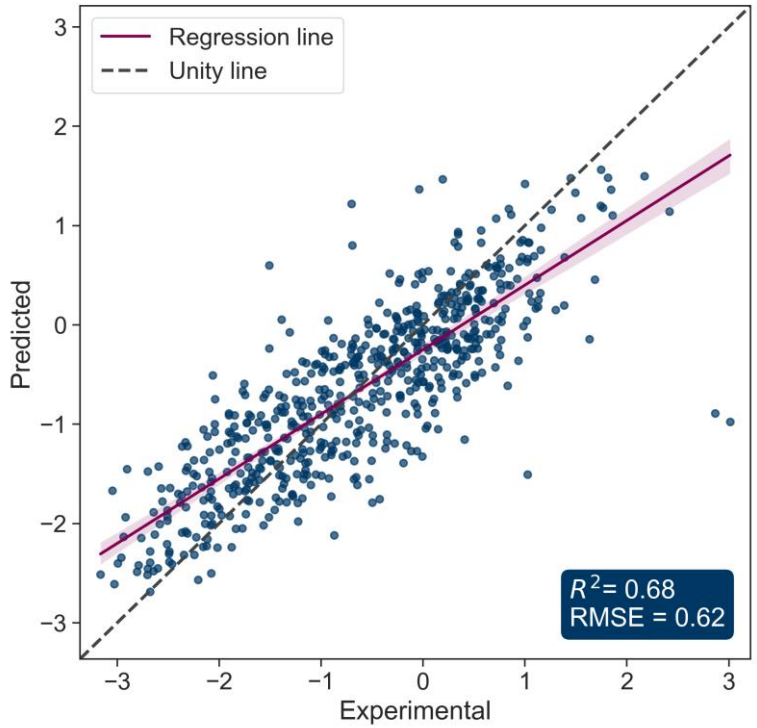
Correlation viewpoint of model performance

- Wide range of experimental values for all parameters – representative of distribution for each parameter
- Good spread of predictions where line of best fit projects proximally to the identity line

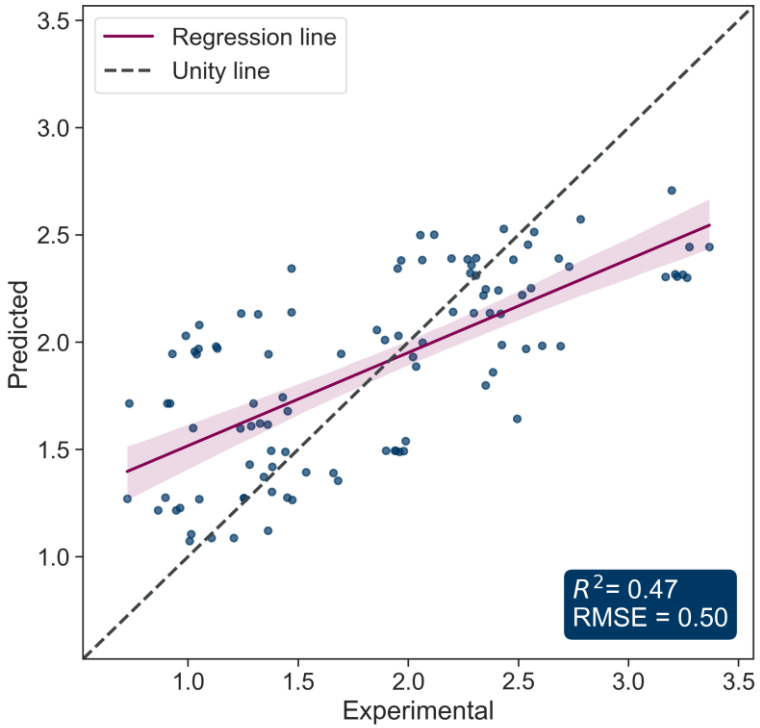
AUC PO



C_{max} PO



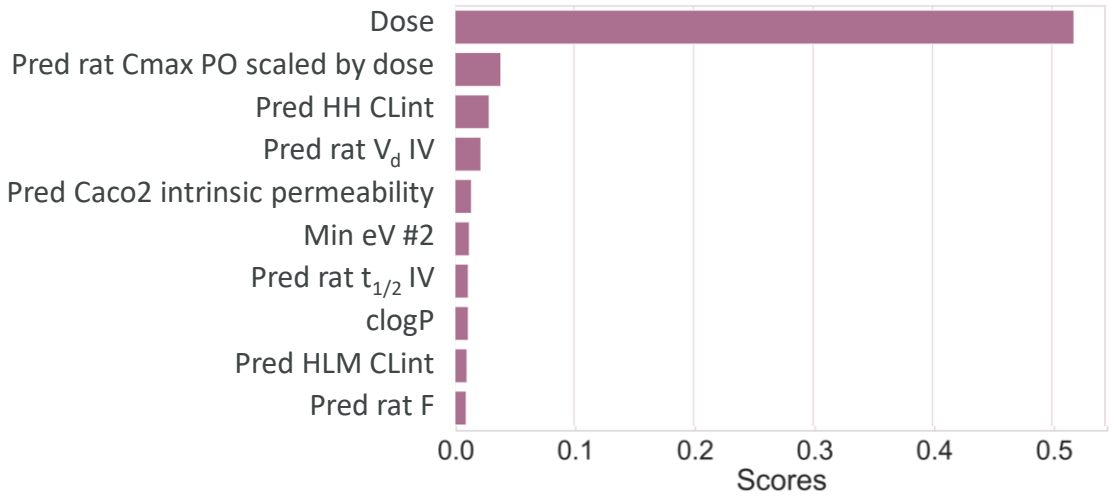
V_d IV



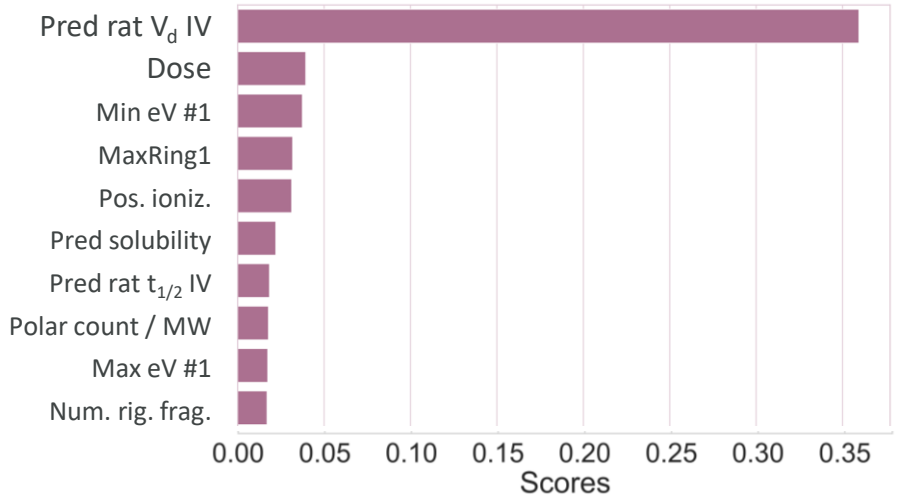
Dose, *in vivo*, and *in vitro* features are most important

- Dose is the most important feature to model AUC PO and Cmax PO
- Predictions of *in vivo* rat PK parameters and *in vitro* ADME properties are also important

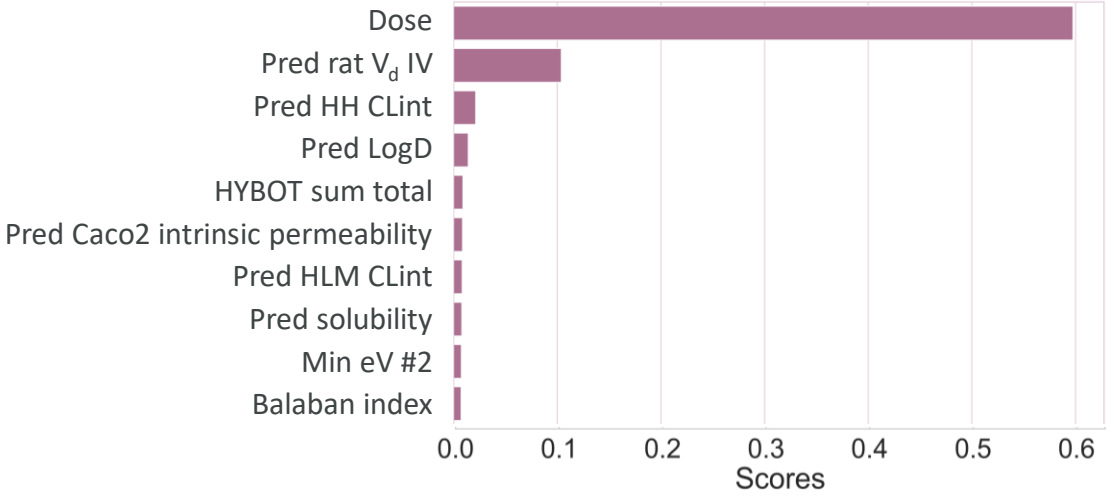
AUC PO



V_d IV

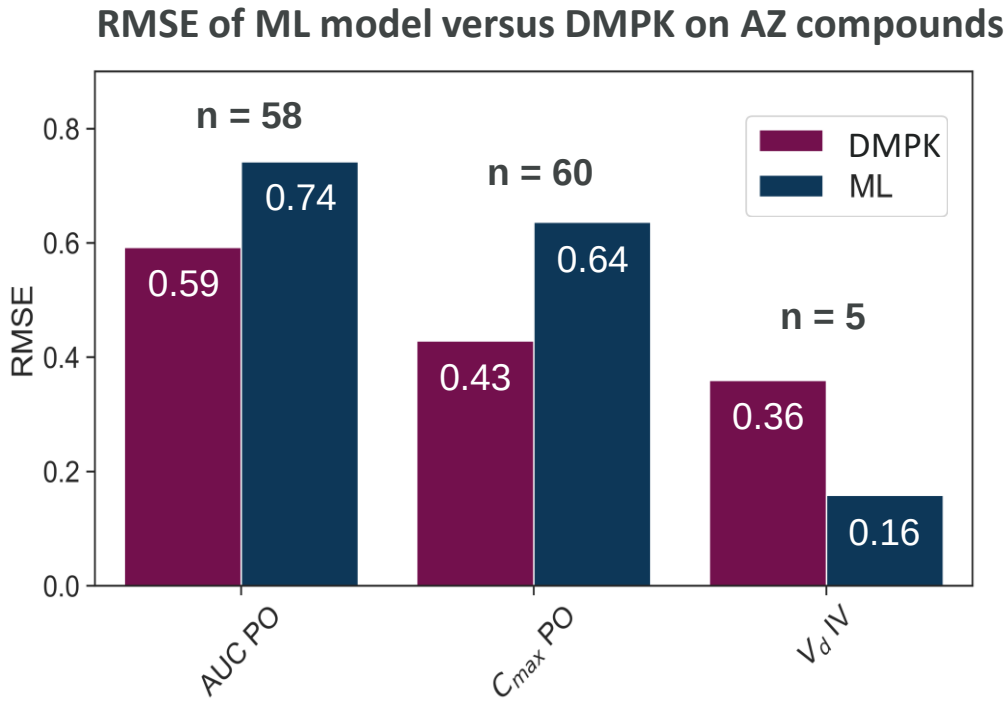


C_{max} PO



Model validation on internal AstraZeneca data

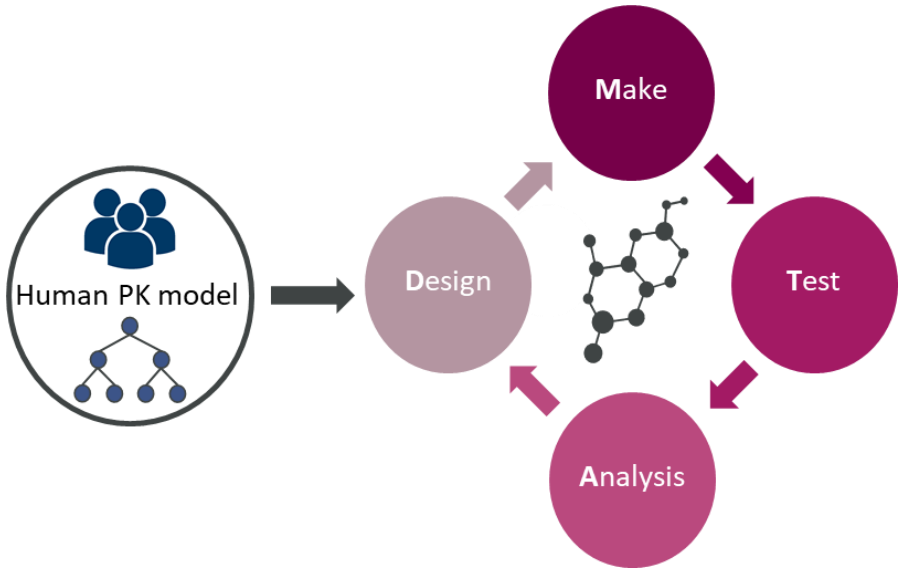
- Benchmarking models on AstraZeneca compounds from Davies *et al. Trends in Pharmacological Sciences,* 2020
  - 126 compounds with clinical data, dose and predicted pre-clinical PK parameters
- Accuracy of ML models is lower than that of pre-clinical prediction (DMPK)
- Enabling prediction from chemical structure – value at the point of design
- **Models are fit-for-purpose to be used in early drug discovery**
- **Machine learning models are complementary to current pre-clinical predictions**



Potential model applications

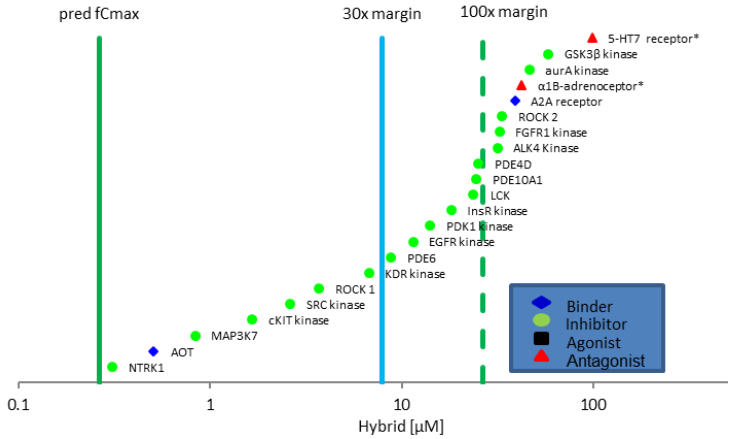
Increase efficiency of DMTA cycle

- Influence decision making and compound prioritization on projects in early drug discovery



Safety-related models and visualisations

- Utilize predicted C_{max} to use following tools at earlier stages
- Account for exposure in organ toxicity models



Conclusion

- Comprehensive protocol for the curation of human PK data which investigates machine learning capabilities to predict PK parameters using only chemical structural information and dose
- Three fit-for-purpose models for AUC PO, C_{max} PO, and V_d IV successfully validated using AstraZeneca internal clinical data
- DMPK modelling showed better performance - their development and validation is often cumbersome which plays into strength of *in silico* alternatives
- Paper published in *Molecular Pharmaceutics, https://doi.org/10.1021/acs.molpharmaceut.1c00718*

Our models show great potential to assist candidate prioritization towards the clinic, aid individual project decisions, or guide *de novo* generative models to design compounds with improved PK properties

Thank you.

Anton Martinsson
Olga Obrezanova
Andreas Bender
Ioana Oprisiu
Nigel Greene
Graham Smith
Beth Williamson
Martin Johnson
Andy Sykes