Complementing QSP with Machine Learning to Predict IGA-Based Population Response in an Atopic Dermatitis Model

Anjali Mittal[a], Nicole Parish[a], Destiny A. Simpson[a], Ryan Suderman[a]

[a]Quantitative Systems Pharmacology Solutions, Simulations Plus, Inc.

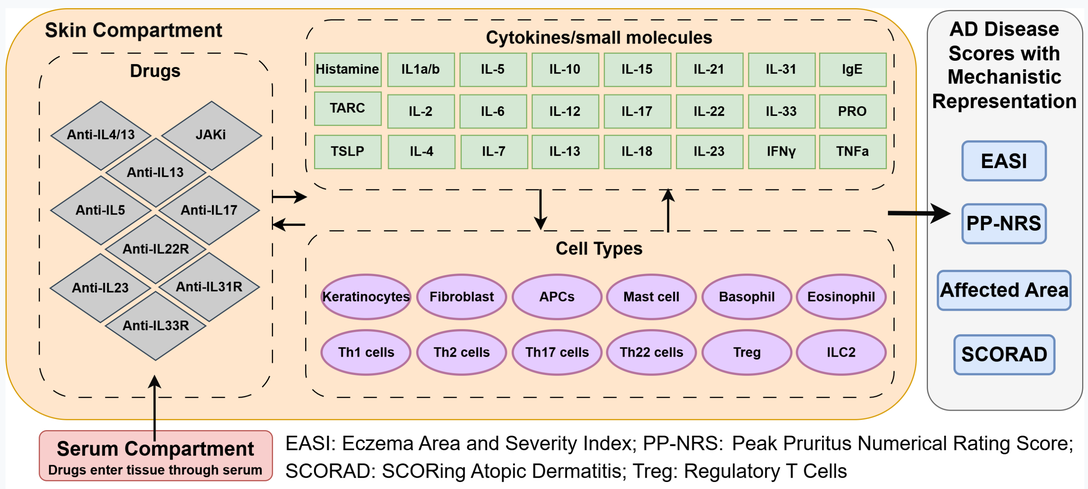
Contact: ***anjali.mittal@simulations-plus.com***

OBJECTIVE

- Atopic Dermatitis (AD) is a chronic skin condition characterized by inflammation, itching, and redness. AD is an active area of drug development, with 120+ drug candidates in various stages of pre-clinical and clinical development[1].
- Quantitative Systems Pharmacology (QSP) combines mechanistic disease understanding with drug pharmacodynamics and can guide dose selection and trial design in AD.
- AD clinical trials differ in their choice of endpoints, which complicates trial design and limits comparability across studies[2]. A QSP model capable of predicting multiple endpoints could enable cross-therapeutic comparison.
- The Investigator's Global Assessment (IGA) is a clinician-administered composite score (typical range 0-4) designed to quantify disease severity. The percentage of patients achieving IGA 0 or 1 serves as a key endpoint in trials enrolling moderate-to-severe AD patients. IGA is inherently subjective and the lack of published data directly linking IGA to disease biomarkers presents challenges for mechanistic modeling.
- We implement a machine learning strategy to overcome these limitations and integrate a population-level IGA response endpoint into an established AD QSP model.

ATOPIC DERMATITIS QSP MODEL

- The AD-QSP model is developed in the Thales™ platform and represents major disease pathways and endpoints. The AD SimPops® (virtual population) is calibrated and validated to data from 32 clinical trials enrolling moderate-to-severe AD patients across 12 drugs and placebo (**Figure 1,2**).



EASI: Eczema Area and Severity Index; PP-NRS: Peak Pruritus Numerical Rating Score; SCORAD: SCORing Atopic Dermatitis; Treg: Regulatory T Cells

Figure 1: Overview of AD QSP Model

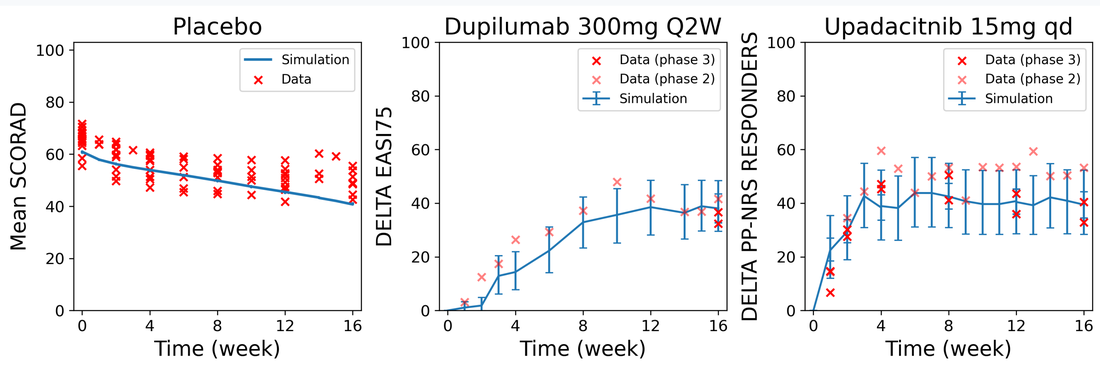


Figure 2: Representative AD SimPops® Simulations. DELTA represents difference from trial placebo. All data and simulation values represent population mean. The error bars represent 90% credible intervals of model simulation. Credible interval calculation is based on sample size of the trial (N). Phase 3 trials with higher N have tighter 90% credible intervals.

METHODS

- Clinical data from 38 moderate-to-severe AD trials was compiled to train a machine learning predictor for population-level IGA response. 14 trials across 5 drugs were reserved for *blind validation*. Data for 7 drugs and placebo across 29 trials was randomly split 80:20 into *train* and *test* sets. The *train* dataset was used to select the model structure after filtering for missing data.
- Candidate input features were selected from existing AD QSP model endpoints based on their correlation with IGA response and data availability (**Figure 3,4**). Ridge regression, support vector regression, and random forest models were trained on different feature combinations. Performance was measured using root mean square error (RMSE) and R^2 scores. The model with the most favorable performance and structure was selected, re-trained, and validated (**Figure 5**). Bootstrap resampling was implemented using 1000 samples to quantify uncertainty in the trained predictor arising from variability in the training data (**Figure 6**).
- The machine learning model was integrated into the existing QSP model and a new AD SimPops® was calibrated (**Figure 3**).

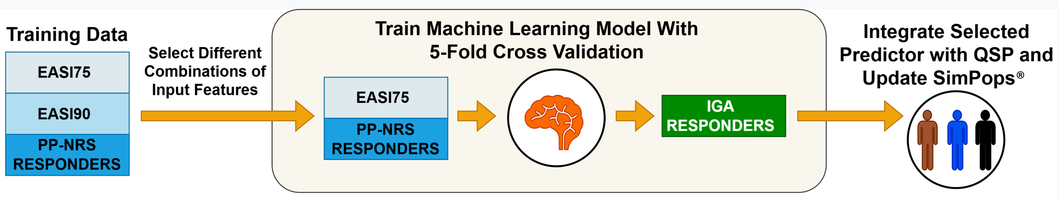


Figure 3: Overview of machine learning and QSP pipeline.

RESULTS

- IGA responder demonstrated strong correlation with other clinical endpoints (**Figure 4**), despite the variable relationships reported between IGA and other AD scores like EASI and SCORAD[3]. This observation supported a machine learning approach to estimate IGA responders.
- Ridge regression with input features [EASI75, EASI90] was selected as the best choice among the evaluated models. The final machine learning model achieved good performance, with R^2 values exceeding 0.9 in training, test and validation datasets (**Figure 5**). Validation against held-out data demonstrated good predictive performance of the machine learning model.
- Bootstrap resampling showed low variability in model predictions, with 95% of bootstrap predictions within 2.5 percentage points of the implemented predictor (**Figure 6**).
- AD SimPops® reproduced clinically observed IGA response across multiple therapeutics (**Figure 7**). Validation to held-out abrocitinib dose response data showed good agreement with clinical values.

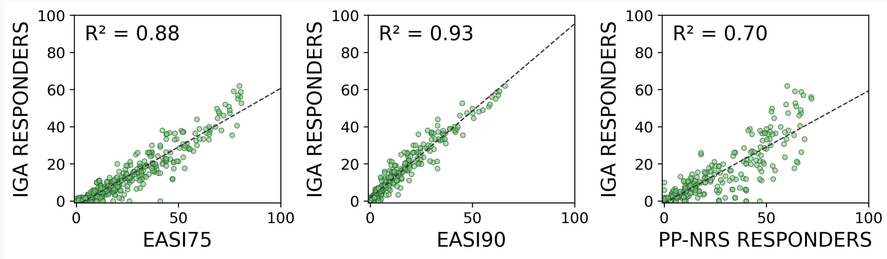


Figure 4: Correlation between IGA response and other clinical endpoints. Each marker represents one trial arm at one reported timepoint. Pearson R^2 values are shown.

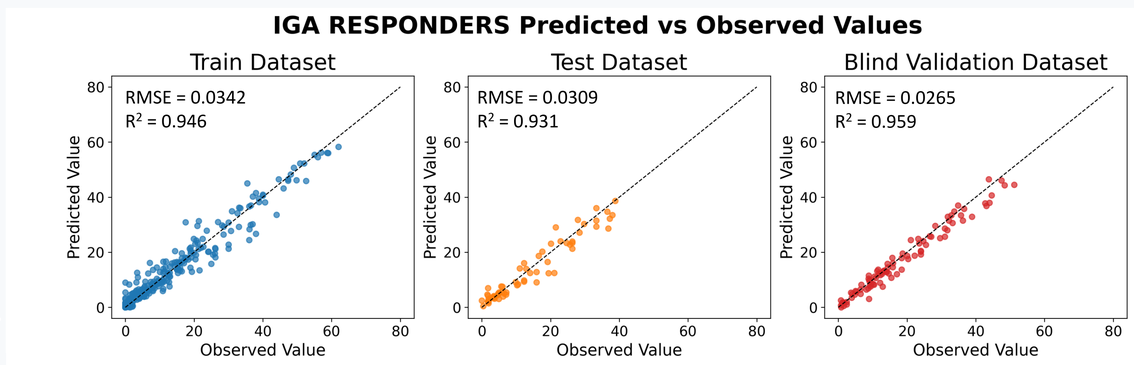


Figure 5: Comparison of predicted vs. observed values for *train*, *test*, and *validation* datasets. Drugs in the *validation* dataset were not used for model selection or training. Each marker represents one trial arm at one reported timepoint. Dashed line represents $x = y$. RMSE: Root Mean Square Error; R^2: coefficient of determination.

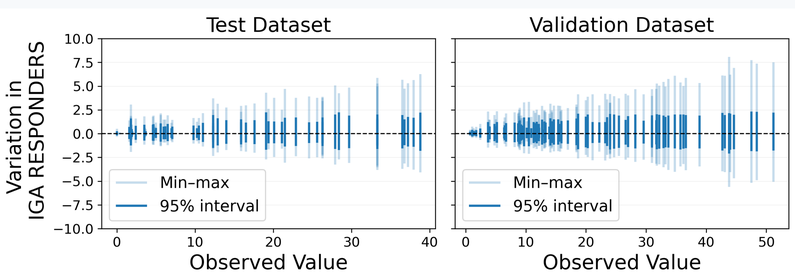


Figure 6: Quantification of predictor uncertainty. Plots show difference between predictions from bootstrap-trained models and the implemented predictor across the range of observed values. Each vertical line represents one observed data point; the darker segment represents the central 95% of bootstrap predictions.

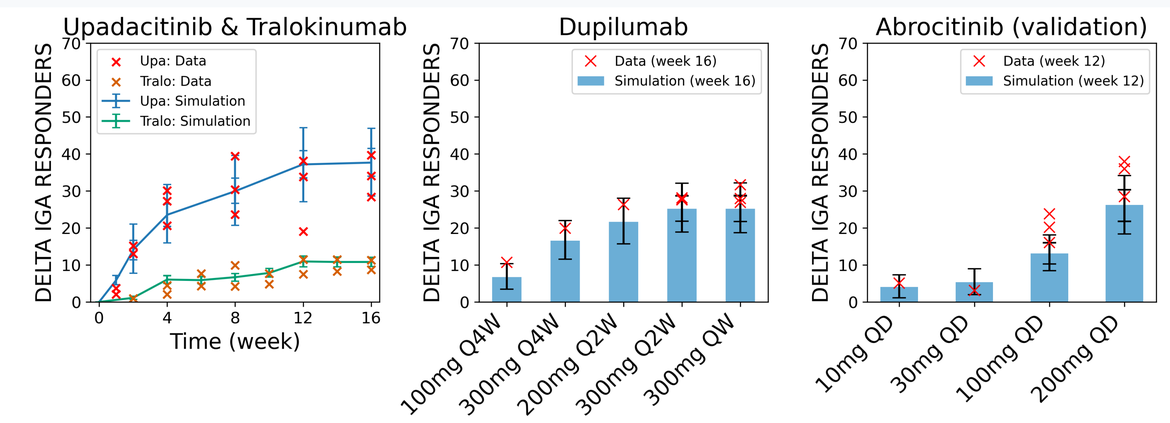


Figure 7: Representative AD SimPops® Calibration and Validation Results for IGA responder. DELTA represents difference from trial placebo. All data and simulation values represent population mean. The error bars represent 90% credible intervals of model simulation. Credible interval calculation is based on sample size of the trial (N). Phase 3 trials with higher N have tighter 90% credible intervals. Abrocitinib trial data was held out and not used for SimPops® calibration. Dose regimens: Upadacitinib 15mg QD, Tralokinumab 300mg Q2W.

CONCLUSION

- Machine learning was successfully integrated with mechanistic QSP for modeling a population-level clinical endpoint.
- This methodology enhanced the utility of an existing QSP model and provides a scalable framework that can be adapted to other disease areas. Importantly, this approach enables modeling of clinical endpoints in situations where mechanistic data is insufficient, thereby broadening the applicability of QSP models for clinical trial simulation and outcome prediction.

REFERENCES

1. DelveInsight: Atopic Dermatitis – Pipeline Insight, 2026
2. Williams et al., JACI (2022)
3. Hurault et al., Br J Dermatol (2018)

AI Disclosure: During the preparation of this work, the authors used Microsoft CoPilot and ChatGPT in order to assist with literature review, generate computer code, and refine poster text. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.