Received: 17 August 2022 | Revised: 25 October 2022 | Accepted: 29 October 2022

DOI: 10.1002/psp4.12887

PERSPECTIVE

# Learn–confirm in model-informed drug development: Assessing an immunogenicity quantitative systems pharmacology platform

**Linnea C. Franssen**[1] | **Maciej J. Swat**[2] | **Andrzej M. Kierzek**[2,3] | **Rachel H. Rose**[2] | **Piet H. van der Graaf**[2] | **Hans Peter Grimm**[1]

[1]Roche Pharma Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, Basel, Switzerland

[2]Certara, Sheffield, UK

[3]School of Biosciences and Medicine, University of Surrey, Guildford, UK

**Correspondence**
Linnea C. Franssen, Roche Pharma Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, Grenzacherstrasse 124, CH-4070 Basel, Switzerland.
Email: linnea.franssen@roche.com

**Abstract**

Immunogenicity against therapeutic proteins frequently causes attrition owing to its potential impact on pharmacokinetics, pharmacodynamics, efficacy, and safety. Predicting immunogenicity is complex because of its multifactorial drivers, including compound properties, subject characteristics, and treatment parameters. To integrate these, the Immunogenicity Simulator was developed using published, predominantly late-stage trial data from 15 therapeutic proteins. This single-blinded evaluation with subject-level data from 10 further monoclonals assesses the Immunogenicity Simulator's credibility for application during the drug development process.

Extending the model of Chen et al.[1,2] published in this journal, the Immunogenicity (IG) Simulator[3] integrates multiple factors driving unwanted IG against therapeutic proteins (TPs) by combining bioinformatics predictions of antigen presentation, a pharmacokinetic (PK) model, as well as a quantitative systems pharmacology (QSP) model of lymphocyte activation and antidrug antibody (ADA) production. It was developed by the Certara IG QSP Consortium as an in silico platform to predict the ADA prevalence and the impact of ADA on PK over time for TP studies.

Scientific rigor, utility, and potential regulatory acceptance of such a model-informed drug development (MIDD) platform demand qualification for each of its specific contexts of use.[4] Here, we assessed the credibility of the IG Simulator QSP model for three contexts of use in the drug development process through an unbiased evaluation. For this, preclinical and clinical data that are typically available at the respective stages during drug development were leveraged.

In particular, the objectives of our evaluation were to

- Provide an unbiased IG Simulator assessment by blinding the modeling operator to experimental outcomes;
- Extend the evaluation of the IG Simulator both in terms of the number of compounds assessed and by using a realistic drug development setting;
- Explore how and whether additional data from preclinical in vitro assays can be used to improve predictions; and

Linnea C. Franssen and Maciej J. Swat contributed equally.

- Adjust the platform's workflow to allow its application to real-world clinical study designs.

## METHODS

The data from 10 Roche monoclonal antibodies (mAbs)—two immunocytokines; three monospecifics; one 1 + 1, three 2 + 1, and one 2 + 2 bispecifics—originated from two phase I single-ascending dose trials, seven combined phase I/II multiple-dose trials, and one multiple-dose phase III trial. Nine mAbs were administered intravenously and one subcutaneously. The left panel of Figure 1 shows further characteristics of these trials.

The external model operator received the data in three consecutive stages representative of potential contexts of use of the IG Simulator in the drug development process. Importantly, the transferred data were limited to the minimum necessary to complete the prediction at the respective stage, ensuring an unbiased assessment. The considered stages were the following:

1. Discovery Stage: The molecule amino-acid sequence is central at this stage. In addition, the clinical trial design and a PK projection are needed. In the context of this evaluation, subject-level dosing history, sampling timepoints, and covariates were provided along with a population PK model in absence of IG.
2. Preclinical Development Stage: In vitro assays can further inform the IG potential. In our case, measurements from two in vitro assays that were obtained during the drug development process were provided. The Epibase in vitro DC:CD4+ restimulation assay[5]—hereinafter referred to as "DC-T-cell assay"—characterizes the re-stimulation of CD4+ T-helper cells by dendritic cells (DCs). The Epibase in vitro MAPPs assay[5]—hereinafter "MAPPs"—identifies naturally processed and presented HLA Class II-binding peptides. The assay results were used to refine the model.
3. Clinical Development Stage: During early clinical development, detailed subject-level observations of PK and ADA become available. Using these early clinical trial data allows further model calibration for predictions of IG in later-stage trials. In the context of this assessment, the predictions were compared with the observations only at this stage. Furthermore, we performed sensitivity analyses.

The workflow for the Discovery Stage involved:

I Generating a virtual population representative of the study cohort using Simcyp Simulator V19;

II Determining the ethnicity-specific HLA Class II distribution for the virtual population with the Allele Frequency Net Database;[6]

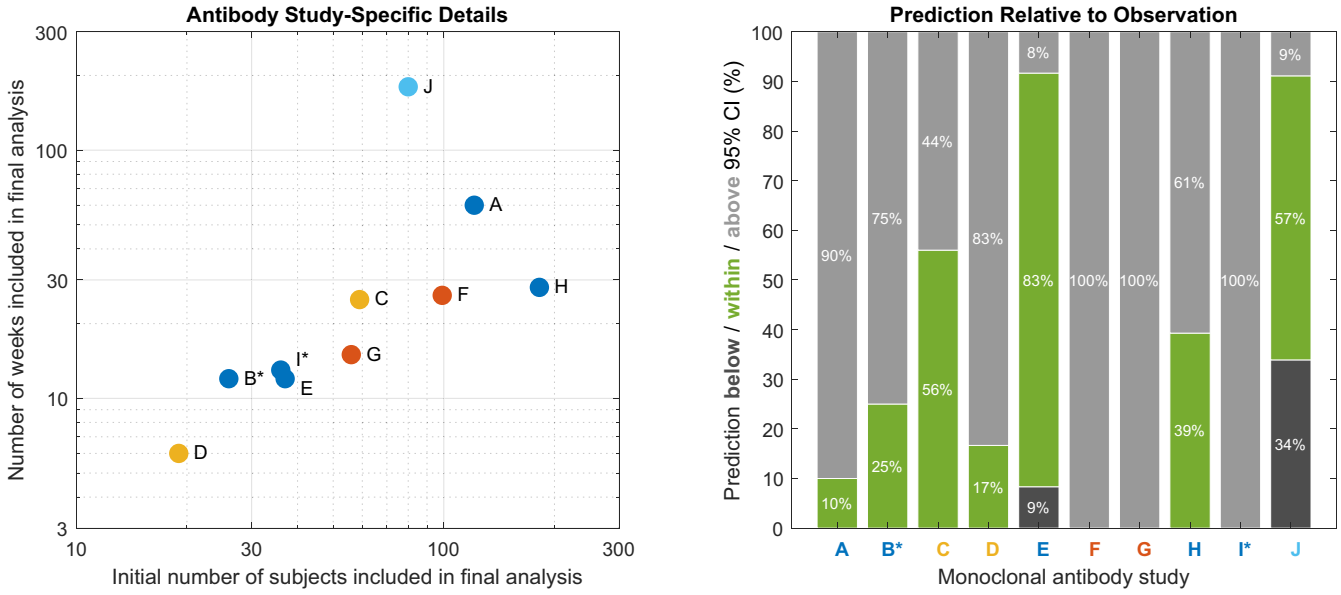


**FIGURE 1** The panels show the characteristics of the molecules and clinical trials included in the evaluation (left) and the high-level comparison of Discovery Stage antidrug antibody (ADA) predictions with clinical observations (right). In monoclonal antibody (mAb) studies marked with an asterisk (*), only a single dose was administered. Left: mAb study-specific details regarding the trial duration accounted for in the analysis and the number of subjects included at the beginning of each trial are illustrated. mAbs are colored by their mode of action effect on IG (cyan: potentially inhibitory, blue: neutral, yellow: potentially stimulatory, red: stimulatory). Right: The stacked bar chart shows the percentages of assessed weeks during which the predicted ADA prevalence was below (dark gray), within (green), and above (light gray) the 95% confidence interval (CI) of the observations for each mAb study. The CI was calculated for a binomial distribution using the function binconf() from R-package Hmisc.

III Predicting the binding affinity of all possible 15-mer peptides along the TP sequence to Major Histocompatibility Complex (MHC) Class II molecules using NetMHCIIpan 4.0;[7]

IV Disregarding any peptides with exact matches in the non-redundant UniProtKB database[8] as identified by blastp;

V Using the binding affinities predicted for the population HLA Class II molecules (cf. Step II) containing the five strongest-binding peptides to any of the population HLA Class II molecules as input to the IG Simulator;

VI Translating population PK models into equivalent minimal physiologically-based PK models[9] within the IG Simulator;

VII Simulating with IG Simulator Version 4.1b accounting for individual schedules for dosing, sampling, and dropout;

VIII For each weekly interval, classifying enrolled subjects as ADA-positive if an ADA concentration of at least 100 ng/mL was predicted.

Ceteris paribus, during the Preclinical Development Stage, we adapted a published model of in vitro T cell proliferation assays[10] to adjust the model parameter representing the initial number of antigen-specific naïve T cells (NT0) based on DC-T-cell assay results for nine mAbs. Separately, we altered Step III for the seven mAbs for which MAPPs data were available to restrict the included NetMHCIIpan-predicted binding affinities to the peptides identified in the assay. Note that the donors for the MAPPs and DC-T-cell assays were distinct from the subjects in the clinical trials.

Keeping other parameters constant, respectively, for the Clinical Development Stage, we performed a sensitivity analysis for all studies on NT0 as well as on the NetMHCIIpan-derived binding affinity parameters describing the antigen presentation. Moreover, we modified system parameters for Study F to account for the mAb's mechanism of action. We also investigated the effect of changing the ADA-positivity thresholds across studies by modulating Step VIII accordingly.

Once clinical observations were unblinded for the Clinical Development Stage, they were retrospectively compared with prediction results from all stages, as the panels of Figure 2 exemplify for mAb C. The population ADA prevalence was summarized weekly by carrying the latest ADA status per individual forward to the end of each week. The predictive performance was summarized per trial by counting the weeks during which the predicted ADA prevalence fell within the 95% confidence interval (CI) of the observed prevalence (see right panel of Figure 1 for Discovery Stage). The CI reflects the varying number of subjects within and across the trials and was calculated for a binomial distribution (using the function binconf() from R-package Hmisc). Throughout, the evaluations were limited to times starting 2 weeks after the first dose and continued as long as at least 12 subjects remained in the trial.

## RESULTS AND DISCUSSION

The performance of the unbiased prediction at the Discovery Stage is summarized in the right panel of Figure 1. Predictions were within the 95% CI of the observed data for >50% of the assessed study duration for three studies, 10%–50% of the duration for four studies, and <10% for three studies. There was a general trend toward overprediction—including for mAbs C, D, F, and G with potentially immuno-stimulatory mode-of-action effects on IG. Because the models did not include mAb-specific mechanisms of action, underpredictions might have been expected in these cases. However, more detailed quantitative information would be needed to improve predictions for these mAbs.

As mAb Study C in the left panel of Figure 2 exemplifies, refinement with preclinical MAPPs and DC-T-cell assay data during the second Preclinical Development Stage failed to improve predictions significantly. This was observed across all studies with available preclinical data.

Robust experimental determination of ADA concentrations is notoriously difficult, and a value for the threshold for ADA-positivity was not available for this analysis. As part of the Clinical Development Stage, we therefore performed a sensitivity analysis showing the significant impact this parameter can have, as illustrated for Study C in the right panel of Figure 2. Moderate improvements could be achieved by selecting the optimal level post hoc for each study. Yet the cutoff is not a free parameter, and the Discovery Stage ADA-positivity threshold of 100 ng/ml was chosen in line with the consortium model default.[3] Only mAb Studies B, D, F, and G were sensitive to changing NT0. For Study F, an overall improvement of the ADA prevalence prediction was obtained by lowering NT0 and increasing the maximum proliferation rate for activated T-helper cells to reflect mAb F's mechanism of action. We identified potential for improvement of the antigen-presentation module because low sensitivity was observed across molecules for changes extending even beyond the range of the binding affinities predicted by NetMHCIIpan 4.0.

We confirmed that the IG Simulator predicts some effect of ADA on PK and hence exposure loss. Yet various attempts of quantifying the comparison between

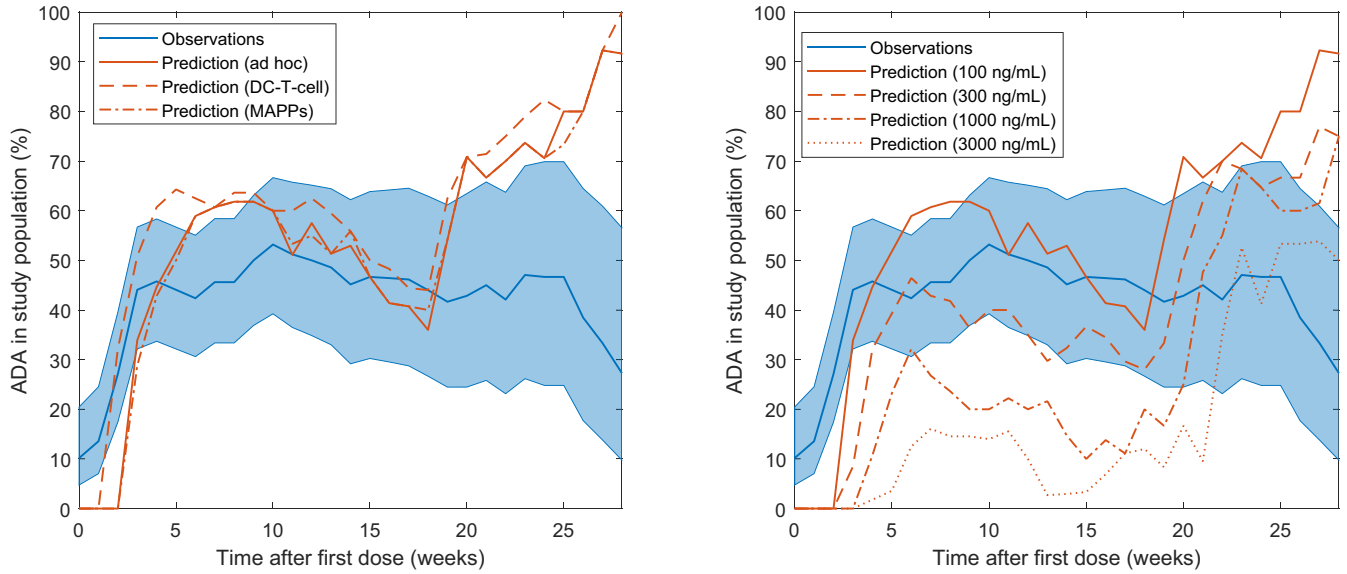


**FIGURE 2** Exemplified by monoclonal antibody (mAb) Study C, the panels illustrate the comparisons between the predicted and observed antidrug antibody (ADA) prevalences that were typically made per mAb study. Each panel compares ADA prevalence observations (blue line with shaded 95% confidence interval (CI) calculated for a binomial distribution to reflect the evolution of the number of subjects enrolled in the trial) with predictions (red lines). Left: Discovery Stage default prediction (solid red line) and refined predictions using in vitro assay data at the Preclinical Development Stage (dashed red lines) are compared to the observations (blue line with shaded CI). Right: Predictions with different ADA detection thresholds illustrated by the Discovery Stage default prediction (100 ng/ml, solid red line) and Clinical Development Stage predictions with higher threshold levels (300–3000 ng/ml, dashed red lines) are compared to the observations (blue line with shaded CI).

observations and predictions of exposure loss highlighted that this is nontrivial.

## CONCLUSIONS

The MIDD platform developer considers this evaluation of a large set of real-world case studies a unique example of a development of a rigorous QSP platform through a sustained, collaborative "learn-and-confirm" approach. Invaluable insights into the model framework, integration of in vitro data, and the associated workflows were gained. The lessons learned will be incorporated into the next version of the IG Simulator model and software. They highlight that the lack of experimental data on absolute ADA concentrations is a major impediment in the field, which emphasizes the critical importance of the parallel evolution of sophisticated QSP platforms and matching experimental and clinical approaches. They further identified opportunities in developing meaningful methods for the quantification of exposure loss, and the IG QSP Consortium is currently working on this.

The sponsor/user concludes that this unbiased evaluation has provided crucial insights into the IG Simulator's performance when applied during the various stages of drug development. The predictive performance was below expectations for the examined stages. This has allowed identifying modules of the IG Simulator necessitating further development. Furthermore, they suggest advancing the integration of in vitro assays; developing methods for model calibration and adaptation; establishing a quantitative comparison method for exposure loss; and providing streamlined, user-friendly workflows for the integration of real-world data. To overcome the practical limitations of comparing simulations to measured ADA, they propose including models of industry-standard bioanalytic ADA assays. The sponsor highlights the unmet need for IG prediction during drug development through qualified MIDD platforms. Therefore, they look forward to advances in performance and credibility of the IG Simulator triggered by this and future performance assessments.

### ACKNOWLEDGMENTS

The authors thank Henry Hofmann, Managing Director at Inovigate Switzerland, for his support as Project Manager of this evaluation, and Timothy Hickling, Investigative Safety and Immunosafety Chapter Lead at Pharmaceutical Sciences, F. Hoffmann-La Roche Ltd., for his scientific advice.

### FUNDING INFORMATION

This work was conducted and funded by F. Hoffmann-La Roche Ltd. and Certara UK Ltd.

## CONFLICT OF INTEREST

L.C.F. and H.P.G. are full-time employees and shareholders of F. Hoffmann-La Roche Ltd. All other authors declared no competing interests for this work.

## ORCID

*Linnea C. Franssen* https://orcid.org/0000-0002-9636-0283

## REFERENCES

1. Chen X, Hickling TP, Vicini P. A mechanistic, multiscale mathematical model of immunogenicity for therapeutic proteins: part 1 – theoretical model. *CPT Pharmacometrics Syst Pharmacol*. 2014;3:1-9.
2. Chen X, Hickling TP, Vicini P. A mechanistic, multiscale mathematical model of immunogenicity for therapeutic proteins: part 2 – model applications. *CPT Pharmacometrics Syst Pharmacol*. 2014;3:1-10.
3. Kierzek AM, Hickling TP, Figueroa I, et al. A quantitative systems pharmacology consortium approach to managing immunogenicity of therapeutic proteins. *CPT Pharmacometrics Syst Pharmacol*. 2019;8:773-776.
4. Musuamba FT, Skottheim Rusten I, Lesage R, et al. Scientific and regulatory evaluation of mechanistic in silico drug and disease models in drug development: building model credibility. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:804-825.
5. Bray-French K, Hartman K, Steiner G, et al. Managing the impact of immunogenicity in an era of immunotherapy: from bench to bedside. *J Pharm Sci*. 2021;110:2575-2584.
6. Gonzalez-Galarza FF, McCabe A, Melo dos Santos EJ, et al. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res*. 2020;48:D783-D788.
7. Reynisson B, Barra C, Kaabinejadian S, Hildebrand WH, Peters B, Nielsen M. Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry MHC eluted ligand data. *J Proteome Res*. 2020;19:2304-2315.
8. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2020;49:D480-D489.
9. Li L, Gardner I, Dostalek M, Jamei M. Simulation of monoclonal antibody pharmacokinetics in humans using a minimal physiologically based model. *AAPS J*. 2014;16:1097-1109.
10. Yogurtcu ON, Sauna ZE, McGill JR, Tegenge MA, Yang H. TCPro: an in silico risk assessment tool for biotherapeutic protein immunogenicity. *AAPS J*. 2019;21:96.

**How to cite this article:** Franssen LC, Swat MJ, Kierzek AM, Rose RH, van der Graaf PH, Grimm HP. Learn–confirm in model-informed drug development: Assessing an immunogenicity quantitative systems pharmacology platform. *CPT Pharmacometrics Syst Pharmacol*. 2023;12:139-143. doi:10.1002/psp4.12887