



Current practices for QSP model assessment: an IQ consortium survey

Jason R. Chan¹ · Richard Allen² · Britton Boras³ · Antonio Cabal⁴ · Valeriu Damian⁵ · Francis D. Gibbons⁶ · Abhishek Gulati⁷ · Iraj Hosseini⁸ · Jeffrey D. Kearns⁹ · Ryuta Saito¹⁰ · Lourdes Cucurull-Sanchez¹¹ · Jangir Selimkhanov¹² · Andrew M. Stein¹³ · Kenichi Umehara¹⁴ · Guanyu Wang¹⁵ · Weirong Wang¹⁶ · Susana Neves-Zaph¹⁷

Received: 28 April 2022 / Accepted: 3 June 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Quantitative Systems Pharmacology (QSP) modeling is increasingly applied in the pharmaceutical industry to influence decision making across a wide range of stages from early discovery to clinical development to post-marketing activities. Development of standards for how these models are constructed, assessed, and communicated is of active interest to the modeling community and regulators but is complicated by the wide variability in the structures and intended uses of the underlying models and the diverse expertise of QSP modelers. With this in mind, the IQ Consortium conducted a survey across the pharmaceutical/biotech industry to understand current practices for QSP modeling. This article presents the survey results and provides insights into current practices and methods used by QSP practitioners based on model type and the intended use at various stages of drug development. The survey also highlights key areas for future development including better integration with statistical methods, standardization of approaches towards virtual populations, and increased use of QSP models for late-stage clinical development and regulatory submissions.

Keywords Network Pharmacology · Drug Industry · Decision Making · Modeling and Simulation

✉ Jason R. Chan
jrchan@lilly.com

¹ Global PKPD and Pharmacometrics, Eli Lilly and Company, Indianapolis, IN 46285, USA

² Worldwide Research, Development and Medical, Pfizer Inc. Kendall Square, Cambridge, MA 02139, USA

³ Worldwide Research, Development and Medical, Pfizer Inc., La Jolla, CA 92121, USA

⁴ Eisai Inc., NJ, Nutley, USA

⁵ GlaxoSmithKline, Collegeville, PA 19426, USA

⁶ Takeda, Cambridge, MA 02139, USA

⁷ Merck & Co. Inc., Rahway, NJ, USA

⁸ Genentech Inc., South San Francisco, CA 94080, USA

⁹ Novartis Institutes for BioMedical Research, Cambridge, MA 02139, USA

¹⁰ Innovative Research Division, Mitsubishi Tanabe Pharma Corporation, Yokohama, Japan

¹¹ GlaxoSmithKline, Stevenage, Hertfordshire, UK

¹² Takeda Development Center Americas Inc., San Diego, CA 92121, USA

¹³ Pharmacometrics, Novartis Institutes of BioMedical Research, Cambridge, MA, USA

¹⁴ Roche Pharmaceutical Research and Early Development, Basel, Switzerland

¹⁵ Drug Metabolism and Pharmacokinetics, Vertex Pharmaceuticals, Abingdon, Oxfordshire, UK

¹⁶ Clinical Pharmacology and Pharmacometrics, Janssen Research and Development, LLC, Spring House, PA 19477, USA

¹⁷ Data and Data Science, Translational Disease Modeling, Sanofi, USA

Introduction

It has been over a decade since QSP modeling was formally proposed as a new paradigm in model-driven drug discovery and development [1]. The integration of mechanistic systems models with pharmacology modeling promised a unique means of relating target and disease biology with drug pharmacology. In that time, numerous examples of preclinical [2] and clinical applications [3] have been demonstrated and the field continues to grow and mature. Indeed, the FDA has highlighted QSP and mechanism-based models as a component of the Model Informed Drug Development (MIDD) Pilot Program [4] and has reported the increased use of QSP in regulatory submissions [5, 6]. As we look forward to the next decade, we envision that continued maturation of the field and successful applications of QSP modeling will benefit from community consensus on best practices, as has been observed within the field of physiologically based pharmacokinetic (PBPK) modeling [7, 8].

Given the myriad of potential uses as well as the different types of models that can be applied (e.g. ordinary differential equations, agent based, etc.), understanding how these models are evaluated and assessed is of critical importance [9]. While the general process of fitting the model to a set of data (calibration) and checking against a dataset not used for fitting (qualification/validation) is typical [3], the details and metrics by which to conclude a model is suitable for its intended purpose can vary considerably, especially when considering factors such as data availability, data quality, model's intended use, knowledge of relevant biology, and model complexity. For example, the approach to model qualification may differ depending on how much data is available at the clinical and/or pre-clinical level. The quality of the available data may also necessitate different standards for assessment as data considered to be of poor quality in some situations may be considered gold standard in others. Perhaps most importantly, greater rigor may be necessary for higher risk applications (such as pipeline or regulatory decisions) than for lower risk applications (such as predicting a result that will be verified clinically).

Model assessment is critical for stakeholders to correctly interpret simulation results. These stakeholders typically come from multi-disciplinary fields, both technical and non-technical, which can result in differing expectations for what may be considered a qualified model. At one extreme, some stakeholders may be satisfied with being shown model fits to many datasets. More technical stakeholders may require rigorous quantitative analyses such as cross-validation, cross-species translation, sensitivity analysis (local or global), virtual populations, or confidence

interval criteria for estimated parameters. Consequently, there is a need for a standardized framework to effectively communicate factors that may impact the use and interpretation of QSP modeling so that all stakeholders can have confidence that the model is suitable for its intended purpose.

Methods

To gain an understanding of current approaches towards assessment of QSP models, an industry survey was conducted with the support of the International Consortium for Innovation and Quality in Pharmaceutical Development (IQ, www.iqconsortium.org), a not-for-profit organization of pharmaceutical and biotechnology companies with a mission of advancing science and technology to augment the capability of member companies to develop transformational solutions that benefit patients, regulators and the broader research and development community. The survey was conceived and constructed by the authors, all of whom are experienced QSP modelers in the pharmaceutical industry. The aim of the survey was to collect and analyze responses from individual QSP practitioners, to understand current practices, and to identify gaps for future directions. The survey results presented here are accompanied by insights and interpretations by the authors.

The survey reached out to individual modelers at small and large biotech/pharma companies. Questions were grouped into the four categories proposed by Ramanujan et al. to accommodate a flexible approach to QSP model evaluation: Biology, Implementation, Simulations, and Robustness [10]. To limit confusion, the definition for QSP shared with survey responders was a mathematical model that represents a biological system relevant to pathophysiology and treatment. This encompasses a set of mathematical equations that represent biological processes relevant to mechanisms of a disease and pharmacological intervention that connect molecular and cellular level processes to measurable biomarker and clinical outcomes. QSP models can vary significantly in how they represent key pathophysiology and clinical outcomes, from simpler mechanistic models to establish PK/PD relationships, to mechanistic models that explore complex drug behaviors, to full-scale platform models of disease. For completeness, all types are considered in this survey. The terms calibration, training, validation, testing, qualification, and verification are often used interchangeably by modelers and can cause confusion. For the survey, calibration/training was defined to be the process of reproducing required behaviors. By contrast validation/testing refers to the prediction of behaviors. Data used in the objective function to arrive at model parameterization is therefore calibration data.

Data that the model can match without being explicated calibrated to is validation data. Qualification is synonymous with validation/testing and verification is related to reproducibility.

No information was disclosed to member companies that could identify the survey taker, partners (e.g. companies, contractors, vendors, universities, or other collaborators), or other proprietary information. Due to its novelty as a field, QSP practitioners may work in a range of departments. Because of this, surveys were distributed to a representative of each member company who was instructed to give the survey to individual QSP modelers within their respective company, regardless of departmental affiliation. The goal was to secure the broadest possible coverage of modelers within companies known to be practicing QSP. The survey, consisting of 59 questions in total (see Supplemental Material for summarized survey results to individual questions), was taken by 88 QSP modelers from 23 companies. No attempt was made to consolidate answers within a company to limit any possible company-specific bias. Nonetheless, individual responses, while providing more data points from which to make inferences, may be inherently biased if groups of QSP modelers within a company did not submit responses on time or were inadvertently overlooked.

Survey results

Demographics

To provide the opportunity for dissecting context in the survey results, demographic questions were asked to understand the background and experience of the survey taker, their role in their organization, and the ways in which they apply QSP modeling. The latter assessed the phases of development and therapeutic areas where QSP models are used, how QSP work is done, the type of models used, and the forums where simulation results are communicated.

The educational background of the survey respondents was mostly in engineering, computational biology, and mathematics (Fig. 1a). Notably, there were few respondents with backgrounds in conventional pharmacometrics modeling or statistics which was consistent with a previous survey focused on preclinical QSP [2]. Most respondents (75 out of 88) had work experience in large pharmaceutical companies while 14 out of 88 had experience working with modeling CROs. With respect to experience with the implementation and/or oversight of QSP modeling, the bulk of modeler's experience lies within large pharma and small CROs (Fig. 1b). Most QSP modelers are individual contributors (Supplemental Material, Q4) and have used or managed 2–10 models over the last two years of their work

(Supplemental Material, Q9). With respect to QSP model development activities, the majority are conducted with internal resources, though there is a clear role for CROs, with some who outsource all QSP work (Supplemental Material, Q6). Most QSP work is presented to project teams (87% of respondents answered “usually” or “always”) and within modeling departments (75% answered “usually” or “always”). Less than 30% of survey responders routinely (answered “usually” or “always”) present their QSP work at external conferences (Fig. 1c or Supplemental Material Q8), which likely presents challenges for the community to coalesce around standard procedures. This may, in part, reflect standard business practices to protect the intellectual property and competitive insights of early discovery and preclinical programs. While 28 respondents indicated that their QSP work does not make it to regulatory discussions, 44 did report at least some interactions with regulators.

Most respondents primarily work on QSP models, but many have some experience with other modeling activities, particularly semi-mechanistic PK/PD models (Fig. 1d or Supplemental Material Q11). Survey takers were asked to characterize the QSP models most frequently used and reference this choice to answer the remaining questions in the survey, anticipating that model type would be informative to understand current model assessment practices: 30 respondents predominantly used mechanistic models to establish PK/PD relationships, 26 respondents predominantly used mechanistic models to explore complex drug behaviors and 32 respondents predominantly used platform models of disease. It is recognized that QSP models exist on a continuum between these three discrete categories and that there is a degree of subjectivity in how each respondent interpreted the definitions, especially those of complex mechanistic models and disease models. For the purposes of the interpretation provided here-in, complex mechanistic models are generally considered to include extended biology or pharmacology beyond PK/PD relationships while platform models of disease are a superset of that definition that additionally include tissue and/or disease processes that are often drug-independent. Examples of the latter include models for asthma [11], inflammatory bowel disease [12], nonalcoholic fatty liver disease [13], and calcium and bone homeostasis [14].

The therapeutic areas where QSP models are applied show consistency with a previous survey [2] except for a noticeable increase in models applied towards infectious disease (Supplemental Material, Q7). This may reflect the more recent efforts to address challenges associated with COVID-19 [15]. Most QSP projects are done in the preclinical-Phase 1-Phase 2 space (Supplemental Material, Q7). Indeed, a role for QSP models has been proposed for designing FIH trials [16] and encouraged by the EMA for

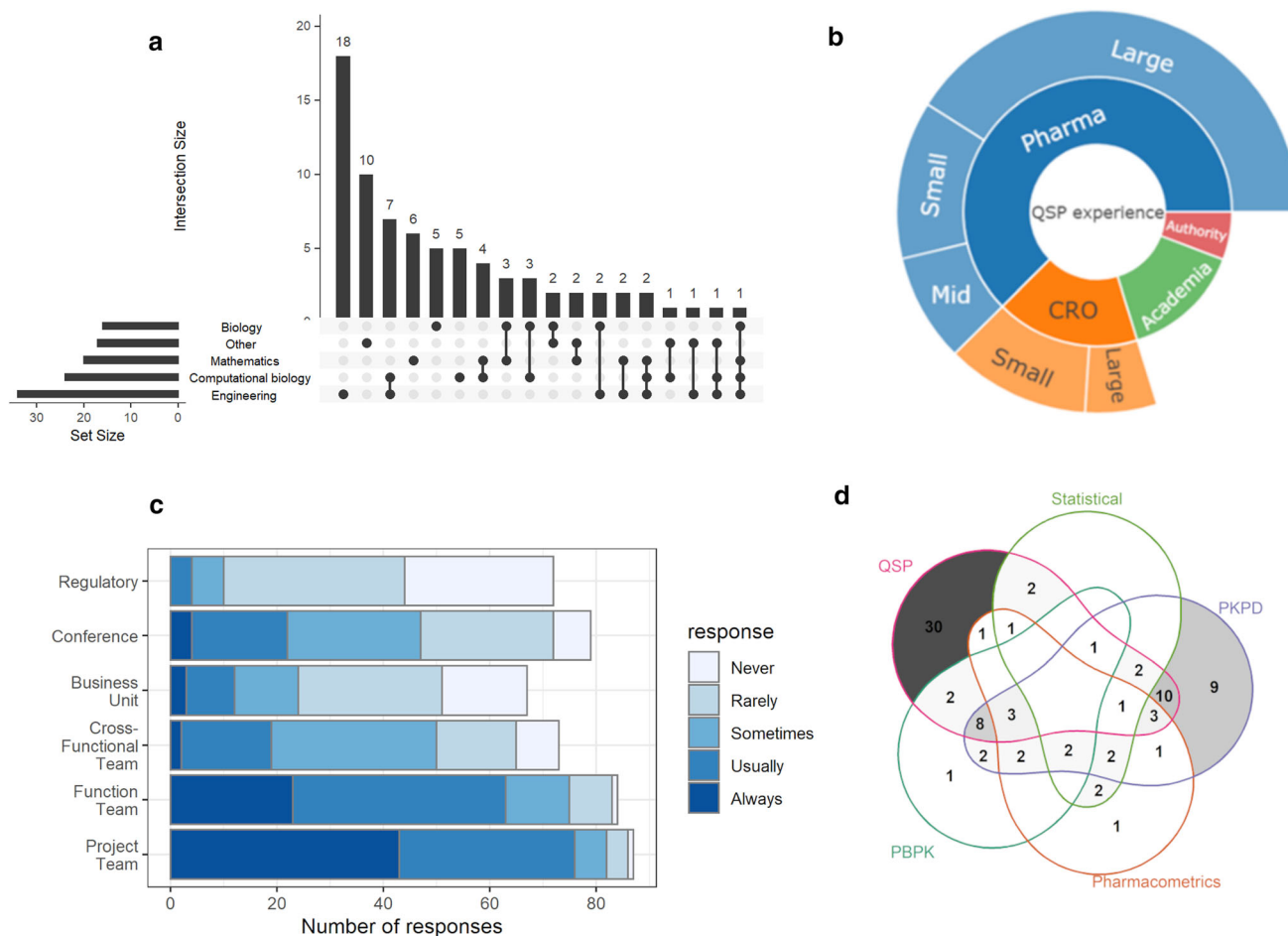


Fig. 1 Responses to selected survey questions in the Demographics section. **(a)** UpSet [31, 32] plot showing the five largest categories of respondents' educational background (set size) and their combinations (interaction size). **(b)** Circle graph showing work experience of

survey takers. **(c)** Stacked bar graph illustrating where QSP simulation results are presented by survey responders. **(d)** Venn diagram illustrating experience of survey takers with different types of models

translational research [17]. Late phase applications, on the other hand, are somewhat sparse. Acceptance of QSP models by stakeholders in early development may be facilitated by an understanding that there is much uncertainty and less patient data available, if at all, for more traditional empirical methods. QSP model simulations in this setting provide an objective approach to consolidate preclinical data and current understanding of the disease to inform decision-making. By contrast, for typical late phase applications, QSP models may be perceived as having little advantage over conventional pharmacometrics models and statistical analyses when the modeling purpose is data characterization or statistical hypothesis testing. However, applications in late phase do exist where QSP models, due to their mechanistic nature, may be relevant including: evaluating combination strategies, straight-to-Phase 3 scenarios in a new indication, exploring efficacy in patient sub-populations or a population not studied in Phase 2, or design of post-marketing studies [18].

Dose selection and competitive differentiation are currently the most notable uses for QSP models, consistent with the prevalence of applications in the preclinical-Phase 1-Phase 2 space (Supplemental Material, Q10). This supports the notion that informing dose selection in the first patient study may be a primary application for QSP, similar to how drug-drug interaction (DDI) assessment is an influential and widely used application of PBPK [19]. Indeed, many of the examples presented at a recent ISOP/FDA scientific exchange were focused on dose decisions [3]. Evaluating combination strategy is another natural utilization for QSP models as indicated by the survey (Supplemental Material, Q10) since a well-developed QSP model can account for biology-based "combination effects" [20]. When there are a large number of potential combinations, QSP models can be particularly useful to prioritize candidate combinations for rational study design and to reach patients with the best therapy as soon as possible.

Biology

The biology section of the survey consisted of thirteen questions covering a range of topics about the collaborative integration of QSP modeling in a matrixed project team, the source data used to build and inform models, and how the results are documented and communicated to stakeholders. The distribution of responses across four of these questions (Fig. 2) exemplify some of the key themes that are apparent in the results (Supplemental Material, Q13–25).

Most respondents identified pharmacometricians, therapeutic area experts, and project lead/team as the key stakeholders defining the research question to be answered by QSP modeling (Fig. 2a). This close collaboration is further supported by frequent interactions with project teams, modeling colleagues, and biology experts (Supplemental Material, Q18–20). These answers may reflect the roots of the QSP modeling community in the fields of standard pharmacology (including receptor theory), physiology modeling, and systems biology.

Interestingly, most respondents also noted little interaction with business/commercial, lifecycle management, and statistics stakeholders. This may suggest that QSP modeling remains less integrated into late-stage development and is not regularly applied as a tool to inform

business/commercial activities. Additionally, responders do not regularly apply QSP models for health authority interactions (Fig. 1c and Supplemental Material, Q16). These survey results suggest there are opportunities for the QSP modeling field to extend beyond its current strengths in drug discovery and early development.

Responses to the types of data incorporated into QSP models revealed several insights (Fig. 2b). First, PK/PD data are the most prevalent data type and suggests that QSP modeling is still widely practiced as an extension to standard pharmacology modeling. This observation is consistent with a previous IQ survey for preclinical QSP where it was noted that modelers primarily work within DMPK and translational PKPD departments [2]. Second, the reported use of clinical data (biomarkers and endpoints) and pre-clinical data (in vivo and in vitro) are more evenly split and may indicate the application of QSP modeling across discovery & early clinical development. Third, few responders reported using data from real-world evidence (RWE) or genomics studies. We note the consistency between low utilization of RWE data and the lack of engagement with late stage or commercial stakeholders in Question 14. The reported low use of genomics data lacks the further qualifier of whether these data are generated in preclinical or clinical studies, but we speculate that there remain

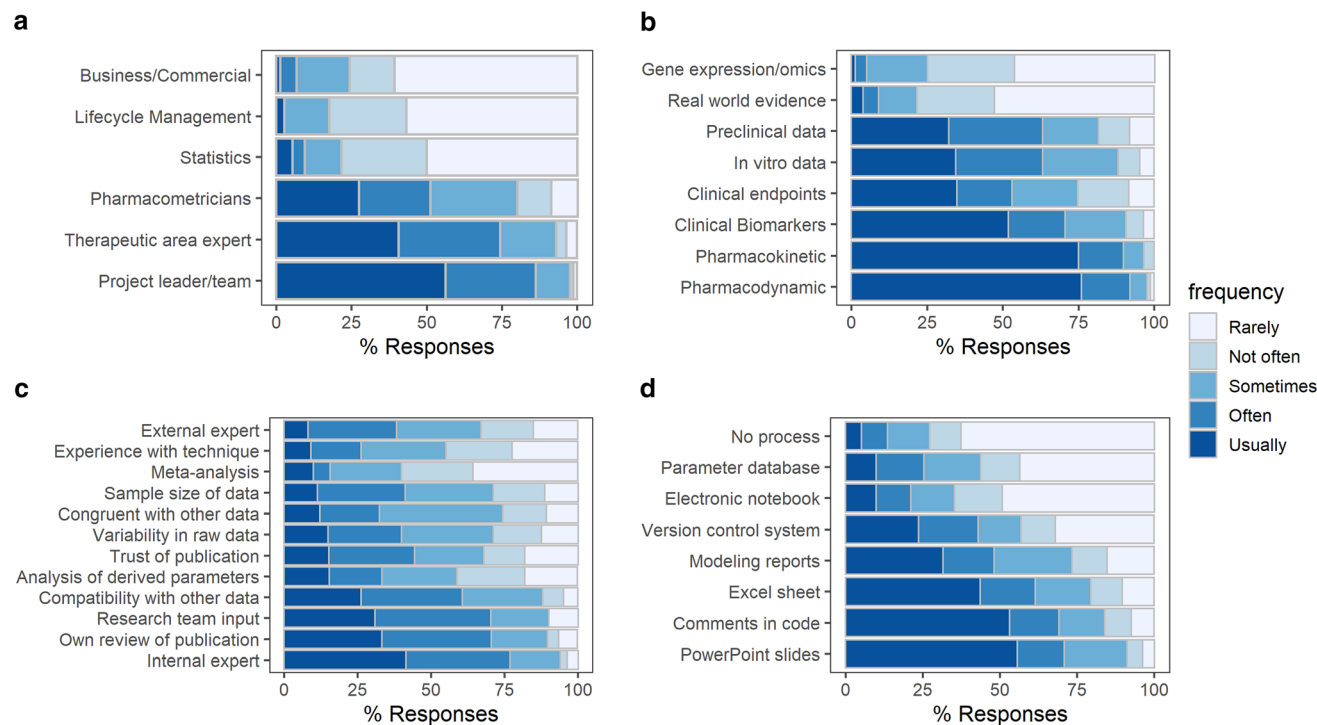


Fig. 2 Responses to selected survey questions in the Biology section. Each plot represents the distribution of answers with respect to the indicated frequencies (rarely, not often, sometimes, often, usually) and is normalized to the total number of responses. **(a)** Q14:

Who are the stakeholders in defining your QSP question? **(b)** Q15: What types of data are incorporated into your QSP models? **(c)** Q17: What criteria do you use to include/exclude data? **(d)** Q21: What is the process for keeping documentation during model development?

opportunities to bridge mechanistic QSP modeling with data science in drug discovery programs [21].

The quantitative and qualitative means by which data are incorporated into a QSP model were addressed by asking what criteria modelers use to include/exclude data (Fig. 2c). We note that subjective criteria are generally favored over more statistically robust methods. For example, the top four answers reflect a preference to rely upon feedback from a subject area expert, the project team, and personal assessments of literature and related data. In contrast, meta-analysis and confidence in the parameters score are less favored. We speculate this survey result may reflect the difficulty in applying rigorous statistical methods to many of the diverse types of data used for QSP model development. This apparent preference for subjective criteria may also be consistent with QSP models being applied in discovery and early development, where assessment standards may not be as rigorous as in late development. Alternatively, this may imply more confidence in internally generated data, internal subject area experts, and internal project team viewpoints, rather than confidence in parameters and meta-analysis that must rely on external data.

Documentation of modeling activities was addressed by two survey questions and the contrast between them is revealing. Survey takers were asked about the frequency at which documentation is performed (Supplementary Information, Q22) and nearly 90% of respondents noted that they do so regularly or at key milestones. The responders' selection of processes used for keeping documentation during model development is shown in Fig. 2d. Here, there is a clear preference for less formal documentation practices such as PowerPoint slides, comments in modeling code or scripts, and Excel sheets for parameter values and associated sources. More formal practices such as Electronic Laboratory Notebooks (ELN), version control systems (e.g. BitBucket), and parameter databases are reported by a plurality (or majority) of respondents as sometimes to rarely used. This may have implications for re-use of QSP models (including in-licensed models; Supplementary Material, Q24) or their application in regulated environments where documentation and traceability are required.

Finally, the skills necessary to be effective as a biology-focused QSP modeler were surveyed (Supplementary Material, Q25). More than 90% of respondents list disease biology understanding, technical/mathematical understanding, and communication skills as very important or essential. Underlying mastery of skills in discovery (i.e. laboratory biology understanding) or clinical development (i.e. clinical understanding, statistics) are not favored to the same degree. This may reflect the collaborative nature of QSP model development in which skilled modelers communicate frequently with stakeholders to build and apply

quantitative understanding and hence the value placed on communication skills.

Notably, the responses to all questions in this section of the survey were consistent regardless of the type of QSP model used or the stage at which the QSP model was applied. This observation suggests that there are no special model type or stage-dependent considerations with respect to the collaborative integration of QSP modeling in multi-disciplinary teams, the criteria used to select source data to build and inform the models, and how the results are documented and communicated.

Implementation

The implementation section of the survey consisted of thirteen questions (Supplementary Material, Q27–39) focusing on various aspects of model development and assessment. Survey takers were asked to consider externally and internally developed models. Visualization of responses to individual questions can be found in the Supplementary Information, with the following results summary focusing on key takeaways from this section of the survey.

A majority (> 60%) of respondents have worked with externally developed models, but were unwilling to accept them as a black box (81.8%), often (> 60%) verifying them before use and even re-implementing them (48.3%) in the case of models coming from publications or databases (Supplementary Material, Q27–29). This may highlight the need for a common approach to QSP model assessment so that modelers can readily understand how applicable published models may be for their particular needs. Furthermore, we found that usage of externally developed models was greater in the development phase versus discovery (Fig. 3a). We speculate that this could reflect available resources and timelines to invest in model qualification and verification, which are potentially requested by regulatory agencies.

To verify a model, respondents almost always required the underlying mathematical equations (85.7%) with many wanting to have access to the code as well (64.3%) (Supplementary Material, Q29). Responders with lowest experience level were more likely to desire the markup language (i.e. the model code) than those with more experience (Fig. 3b). This may be indicative of recent trends in coding and expectations from new modelers in the field and may be related to a lack of clear standards for software and best practices across the QSP field [22].

When a model was not developed by the respondents, they found it important to assess quality of model implementation (87%) (Supplementary Material, Q33). Most responders found it insufficient to only assess the model diagram (55%). Interestingly, while most responders often

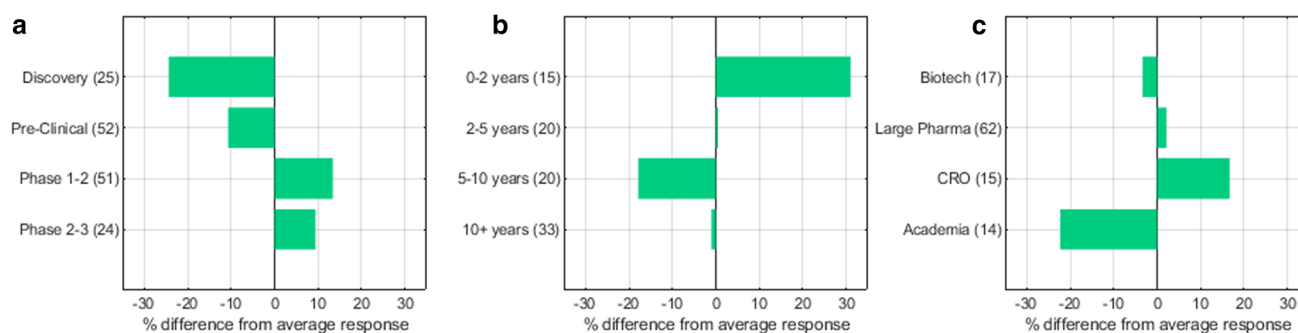


Fig. 3 Deviations from selected responses in Implementation section based on subgroup demographic differences. Ticks on the left side of the plot define responder subgroups with number of total responders belonging to the subgroup in parentheses. Bars represent the difference between average response and subgroup responses. (a) Q28: Do you use QSP models that are developed externally? Answer: Yes (68% of all responders). Subgroup: Phase of drug

(usually or always) considered accuracy (86%), maintainability (61%), and organization (75%) of the code, the speed of simulation was not assessed as often (38%). Solver accuracy was also not often (usually or always) assessed by the majority (> 50%) of responders (Supplemental Material, Q34). This suggests that solver accuracy and speed of simulations are likely not limiting factors to model implementation and may imply that current software, though varied, use mature well-established solvers and are sufficient to meet modeler needs. Alternatively, this answer could also reflect the diverse backgrounds and skillsets of QSP modelers and recognize that deep understanding of the mathematical aspects of solver accuracy and speed of simulations may not be fundamental requirements for success as a modeler.

During model assessment, while local sensitivity on specific parameters was often (usually or always) performed by respondents (62%), consideration of all parameters or a global sensitivity analysis (GSA) were implemented selectively with “sometimes” being the most common survey answer. Global sensitivity on all parameters was the least employed approach, with 46.7% of responders from CROs never performing GSA, while 92.3% of responders from academia performed GSA to some degree (Fig. 3c). This response is likely reflective of the high computational demand of running GSA algorithms and the different purpose of model development in academia versus industry (CRO). Academia may be focused on providing a general assessment of the model, while CROs may focus on a specific application of the model where GSA may not be as relevant.

When employed, sensitivity analysis was often (usually or always) leveraged by the respondents during model development and calibration to understand model simulations (67.1%), identify parameters for model calibrations

development. (b) Q29: In order to use a QSP model developed externally, how often do you require markup language? Answer: Always (25% of all responders). Subgroup: Modeling experience. (c) Q37: In order to assess parameter sensitivity, how often do you use global sensitivity analysis on all parameters? Answer: Never (30% of all responders). Subgroup: Industry

(64.6%), and understand parameter uncertainty (74.1%) (Supplemental Material, Q38). Surprisingly, sensitivity analysis was underutilized to prospectively design follow-on experiments or to identify parameters for virtual populations (30%). This survey result may suggest an opportunity to use sensitivity analyses to further the adoption of QSP models for influencing the design of preclinical experiments. Sensitivity analysis was also underutilized in generation of virtual populations, where biological considerations were the most important parameter-selection criteria (81.7%) (Supplemental Material, Q39).

Simulations

The simulations section of the survey (Supplemental Material, Q40-48) focused on reproduction of behaviors (calibration/training) and prediction of behaviors (validation/testing). Exploration of the inherent uncertainty in QSP models and reproducibility of published simulations were two interesting aspects that emerged which have direct consequences for communication of simulation results to readers and collaborators. The current approaches to communicating uncertainty may depend on the question of interest and context of use. To explore this, we parsed the survey responses into those responders who spent most of their time with a preclinical versus a clinical focus.

For respondents with a preclinical focus, visual check (by overlaying known variability) and parameter driven (i.e. forward simulation by sampling from estimated parameter distributions) approaches were the most used methods for generating a range of likely values for an outcome (Fig. 4a). Notably, a sizeable fraction (~ 16%) of preclinical respondents answered, “I don’t— model used for qualitative learnings”. Perhaps this emphasizes a traditional strength of mathematical biology, deriving

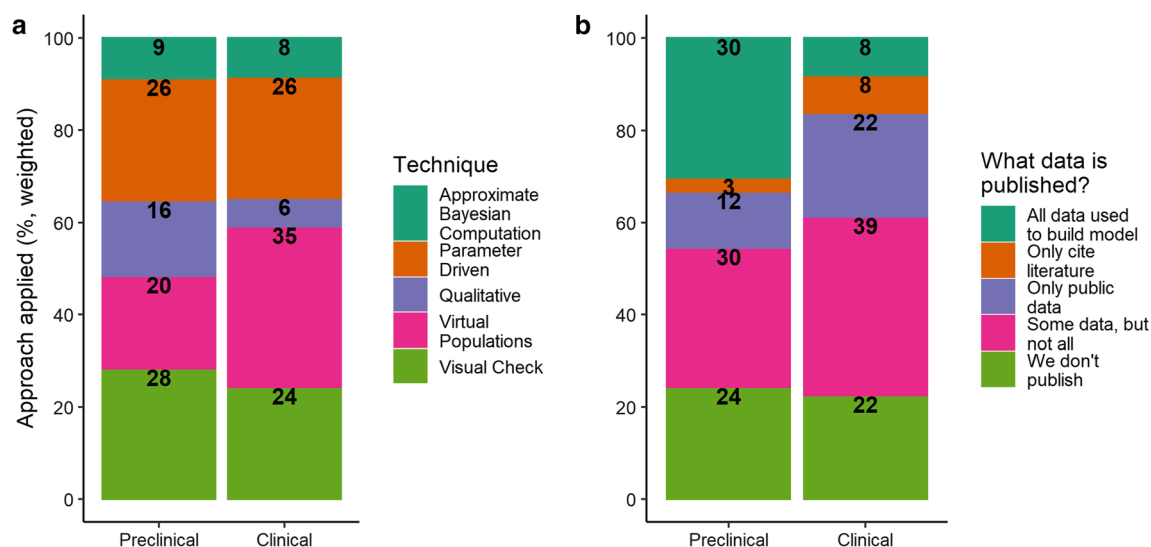


Fig. 4 Responses to selected survey questions in the Simulation section, filtered by respondent preclinical or clinical focus. **(a)** Q42: How do you generate prediction intervals (i.e. range of likely values

for a model outcome)? **(b)** Q47: When you publish a QSP model do you publish all the observations that informed the model (e.g. proprietary pre-clinical observations that informed design decisions)?

mechanistic insights from matching observations qualitatively (for example, inferring the presence of negative feedback with delay from oscillatory data). In contrast, it was clear that the most popular approach applied by clinically focused modelers was virtual populations, perhaps because they can be readily used to capture clinical variability in QSP model simulations (Supplemental Material, Q44) or because they capture outputs measured in the clinic and are thus more relatable to clinical stakeholders. We assume that the difference between approaches to preclinical and clinical simulation is driven by the nature of the modeling problems, the stakes in answering them, and the background of the modelers involved. Nevertheless, we question whether there should be a technical difference and whether preclinical modelers might benefit from considering more quantitative approaches, and conversely whether clinical teams could place more emphasis on qualitative insights (for example, for biological hypothesis generation/refutation).

Depending on the approach to capture uncertainty, and in particular the data used to inform that approach, published models should be more or less “fully reproducible”, which we define as “all published simulations should be reproducible with code provided, and data used to derive or estimate parameters included in a digital format”. In particular, published results may be reproducible but the model is not developable if constraining data is withheld from the publication. For example, if a published model includes a virtual population that matches unshared internal data, a third party might adjust the model without the knowledge that the updated model may now be inconsistent with the internal data.

Across all respondents only 17% said they published all the observations that informed the model. Over 50% said they either do not publish or do not include all of the observations used to inform the model. Interestingly, there was not much difference across type of employer (Supplemental Material, Q47). When the responses are broken down by preclinical versus clinical focus, more modelers with the preclinical focus publish all or most of the data used to inform their model. While those with a clinical focus are more likely to only include partial information (Fig. 4b). This could be due to the proprietary nature of clinical data.

The other questions in this section, which captured how respondents devoted their time, did not yield additional insights when parsed by preclinical/clinical focus. Visual checks were predominantly used compared to sensitivity analysis for validating QSP models (Supplemental Material, Q40), which we speculate may be due to easier interpretation by non-technical stakeholders. Confidence intervals were typically estimated directly from experimental data, rather than from models derived from the experimental data (Supplemental Material, Q43). This practice could have consequences for interpreting simulation results since these means and confidence intervals are often used as calibration targets. With respect to building, fitting, validating, and using QSP models for predictions, modelers indicated that data considerations, rather than computational power, were of primary concern (Supplemental Material, Q45). This is consistent with the findings in the implementation section that speed of simulation was only assessed 38% of the time. Computational limitations if they did become a challenge appear to be addressed mostly

by parallel computing, limiting the scope of simulations, improving code, or simply patience. Interestingly, surrogate models and model reduction were uncommon (Supplemental Material, Q46), though recently mentioned as an avenue towards applying models for regulatory submissions³.

Robustness

For any model intended to inform drug discovery and development, it is important that model robustness be critically evaluated at each stage. Model transparency, including details on expert knowledge and data used for model development, has been identified elsewhere as a major factor for robust model assessment [3]. This enables stakeholders to understand the strengths and limitations of these, often complex, models and to place simulation results in the appropriate context to inform decisions. Due to potentially large differences in the size, scope, and overall complexity of QSP models, the methods used to assess model robustness can be variable. The survey asked respondents a number of questions related to their motivations for characterizing robustness, which aspects of robustness they prioritize, and how they evaluate the robustness of their models. In addition, questions were asked around the robustness of models as a concept to advance scientific understanding, for example, by extending the use of a model from one application domain into another related one. The responses were stratified by stage in the pipeline to provide additional insight. Some questions allowed ‘write-in’ responses, where respondents could add suggestions of their own. The survey questions and results for this section are available in Supplementary Information, Q49–59. The responses to five selected questions are highlighted in Fig. 5 (the write-in responses have been omitted).

Widespread adoption of various practices for evaluation of QSP model robustness was observed (Fig. 5a). The survey responses suggest that all methods: sensitivity analysis, evaluation of predictive ability and exploration of different model structures were used at all stages of drug development. However, there was a trend to rely on predictive ability more in later development stages of development. Of interest, exploration of alternate model structures continued even at the regulatory filing and post-marketing stages, which may reflect the iterative nature of modeling.

Many criteria, which are not mutually exclusive, can be used to evaluate model robustness—data reliability, model assumptions, biology, model predictions (central tendency, variability in response), sources of uncertainty, and clinical variability. We asked which of these are used and how often (Fig. 5b). Respondents reported that they most

frequently assessed robustness by evaluating model assumptions, followed in order of declining frequency by model predictions, biology, data reliability, clinical variability, and finally sources of variability. There did not appear to be a single predominant technique: at least 60% of respondents reported using all criteria ‘Usually’ or ‘Often’. We surmise that these responses reflect context of model use and were consistent whether the QSP application was preclinical or clinical.

Respondents were also asked whether they focus mostly on central tendency or, in addition, on variability around that tendency (Fig. 5c). The survey showed that at the discovery/preclinical phases, respondents cared most about central tendency. By contrast, at the clinical stage respondents cared about both central tendency and variability. There was a clear progression from discovery (71% central tendency only, 29% both) to preclinical (48%/52%) to phase1-2 (14%/86%) to registration (3%/97%). This could reflect the growing importance of characterizing variability as projects move to later stages of development and is consistent with the preferred use of virtual populations in clinical applications of QSP as indicated in Fig. 4a. Equally, it might also reflect a lack of understanding of how variability translates from the preclinical to the clinical domain. To make an oncology-specific analogy, if we do not understand how the variability of rodent xenograft models translates to human clinical response, then there may be limited utility in characterizing it in the first place. Alternatively, modelers and stakeholders at the discovery phase may simply not be concerned with variability and may focus on understanding directionality or where the central tendency is, while inherently accepting that there is a lot of unexplained variability.

Assessment of parameter uncertainty is shown in Fig. 5d. While most modelers did some kind of assessment, this question in particular showed a lot of variability in the methods being used. In addition, it led to several write-in responses from the survey takers, including assessment of how source data was derived, calculation of profile likelihood, sensitivity analysis, use of virtual populations, confidence intervals from fitting, and reported variability between data sources. This survey result suggests a need for standards that are flexible enough to accommodate the myriad of use cases and contexts that QSP modelers encounter.

Some modelers in the QSP community believe that by investing resources in modeling the behavior of both the biological system and a particular drug, a robust model can be developed that is suitable to be repurposed outside its initial area of focus. In part, this justifies the relatively high demands of this modeling approach. Given this investment, we expected that such re-use would be common and were surprised, that 70% of re-used models were used no more

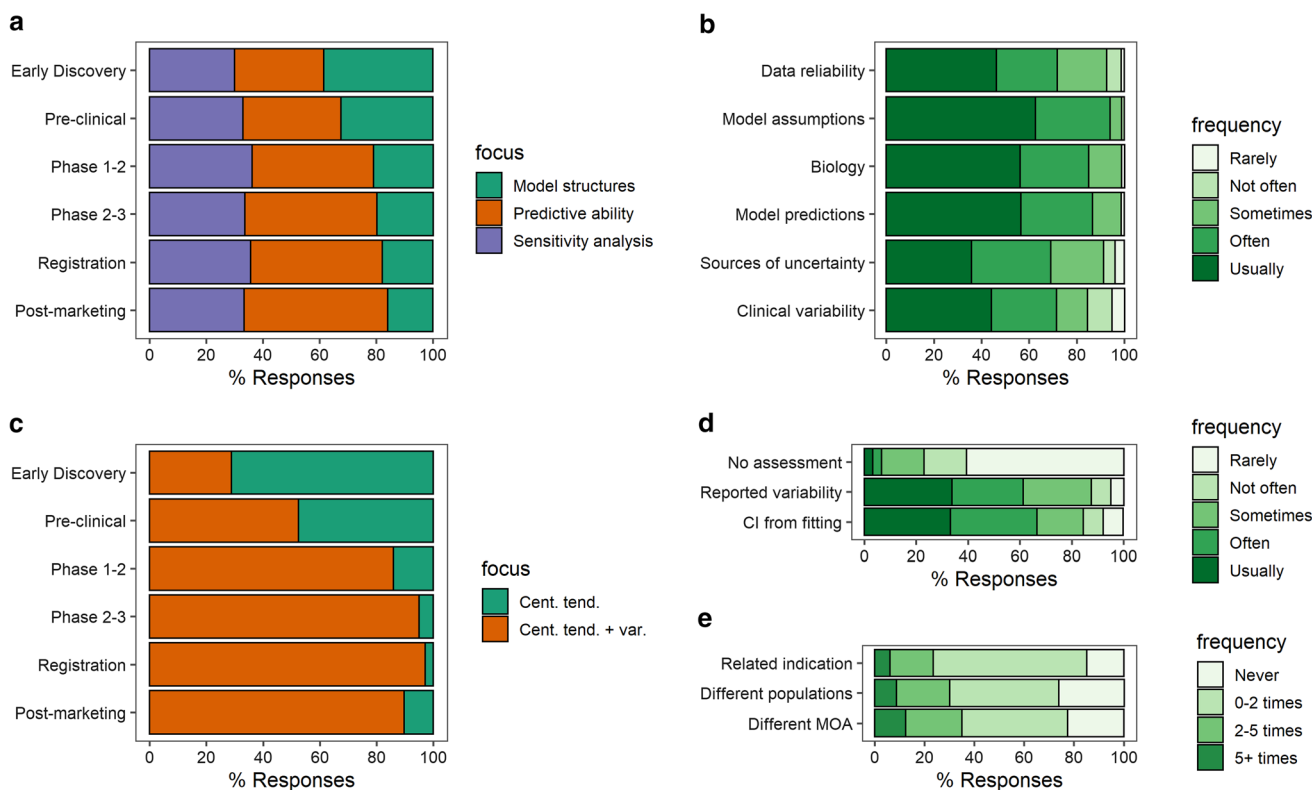


Fig. 5 Selected findings for the Robustness part of the survey. **(a)** Q51: For each stage, what methods do you use to ensure your model is robust? **(b)** Q52: How often do you evaluate model robustness using these criteria? **(c)** Q53: Do you focus on the model being able to describe the central tendency of the dynamics in question or the variability as well? **(d)** Q56: How do you assess

parameter uncertainty? **(e)** Q57: How often have you repurposed a model outside its initial area of focus? Cent. tend. = central tendency, var. = variability, MOA = mechanism of action. ‘% Responses’ indicates the % of responses for that particular question, since not all questions were answered by all respondents

than two times by any individual modeler (Fig. 5e.). This response bears investigation into the causes, and may be related to the interpretation of “repurpose” (e.g. what percentage of the model would need to be re-used to qualify it as “repurposed”) and “initial area of focus” (e.g. whether it refers to one disease, or one therapy). This low figure may also reflect the fact that modelers may support just a few targets within one or two therapeutic areas or that an early-career modeler just has not yet had the opportunity to re-use a model. Alternatively, this survey result may reflect the belief from other QSP modelers that models should be fit for purpose and that for each new purpose, a new model may be needed.

Discussion

The aim of this survey was to evaluate the current state-of-the-art for QSP model assessment. As experienced practitioners of QSP modeling who wish to see continued and broadening adoption of QSP, we offer some interpretation

of the results along with recommendations to advance the discipline.

The survey results emphasize that QSP modeling is currently used predominantly in discovery, preclinical, and early clinical development stages, with relatively fewer interactions with commercial, late stage (Phase 3) clinical development, and regulatory agencies. We surmise that this is because the framework of QSP models is based on knowledge of physiology encoded in the equation forms, rather than empirical considerations. The data used to construct QSP models are often heterogeneous and small in scale, but QSP models provide a means to put the data together in a meaningful and cohesive manner. As such, QSP models are quite relevant in the low-data scenarios early in drug development where hypotheses need to be generated and predictions made by extrapolation; in contrast to empirical models that are better suited for interpolation when data are available later in development. During earlier stages of drug development, QSP modeling can identify knowledge gaps and guide what data is necessary to fill those gaps so that as the drug development programs progress to later stages, the appropriate data are

generated to inform decisions. Thus, while models may not directly inform decisions in later stages of drug development, they may inform what data is generated to enable late-stage decision making.

The above observations are consistent with the survey result that QSP models are still primarily used to facilitate internal decision making in the pharmaceutical industry. The use of QSP modeling to support regulatory interactions appears to be infrequent and is in line with a recent scientific exchange between the FDA and pharmaceutical industry where the presented case studies primarily focused on QSP models used for internal decision making [3]. This lack of industry-reported use of QSP to support regulatory submissions, however, appears to be at odds with a recent FDA report, which showed a dramatically increasing occurrence of QSP in submissions across therapeutic areas [5]. The reasons for this discrepancy are not clear, but could be a question of semantics: no standard exists for how many biological mechanisms a model should contain in order to be labeled QSP; thus, the FDA may have classified more models as QSP than those taking this survey. Alternatively, it could reflect the background of the survey takers, who may use large-scale disease platform models more, which are inherently more difficult to evaluate and thus less likely to be included in regulatory submissions. Regardless of the reasons behind this apparent discrepancy, it is clear that, at the individual level, responders rarely contributed to regulatory submissions even if QSP is growing as a partial component of submissions. This might reflect the high attrition rate of programs (and hence the QSP work supporting them) and/or a skew to investment in QSP applications for preclinical and early clinical development. Regardless, even with the reported increase to 60 instances of QSP in submissions in 2020 [5], an opportunity clearly exists as we estimate that this represents about 4% of annual IND submissions, assuming that there are approximately 1500 IND submissions per year [23].

Model assessment, while performed, appears to be quite variable in implementation. The uneven approach to assess QSP models is perhaps not surprising given the diverse training backgrounds of QSP modelers. While the chosen methods appear to be sufficient for their current use, higher and more uniform standards of model assessment will be necessary to further build confidence in QSP among non-modeler stakeholders and regulators. We emphasize that we are not calling for adoption of such standards as an absolute requirement across the development pipeline, as there will not be a single approach that fits all applications. For example, some models are fit for purpose with a narrow scope to answer a particular question. Platform models, by contrast, are aimed at multiple programs to generate hypotheses that make sense of difficult to explain data.

Given their differing goals and complexities, the methods used for model assessment and corresponding standards should also differ, requiring a flexible approach [10]. As more examples of QSP model assessment arise in the public domain, we expect that context-dependent standards will naturally evolve.

Looking across all segments of model assessment, several themes emerged from the survey results that provide a glimpse for how the QSP community can maximize the uptake and the impact of QSP modeling activities. Relatively few QSP practitioners were identified as having an educational background in statistics as reported in this and in the previous QSP-focused survey [2]. Earlier collaboration with statisticians during QSP model planning, development, and validation could be beneficial for communicating with stakeholders. Project teams typically include statisticians who, as a discipline, are not familiar with QSP. Better communication with statistics colleagues would enable a unified approach towards model-informed drug development, ensuring that appropriate modeling methods are applied and that simulations are consistently interpreted. An opportunity exists for statistical considerations to be applied more robustly as the QSP community moves towards standardized approaches for constructing and evaluating virtual patient populations. As an example, incorporation of placebo response in QSP models has benefitted from statistical approaches [24].

Documentation of QSP models was revealed to be a surprisingly simplistic and informal process by the survey, with many practitioners using separate spreadsheets and PowerPoint presentations, despite recent publications suggesting the need for standardized approaches to model reporting [25]. More-consistent reporting methods would be beneficial to facilitate model assessment and build confidence in the model for internal and regulatory purposes. Documentation of model credibility using a risk-informed credibility assessment framework derived from the ASME V&V40 has been proposed for PBP [26] and the FDA has requested that a model risk assessment be applied towards all submissions to the MIDD pilot program [4], regardless of modeling approach. Such a framework is amenable to QSP models and could enable better communication amongst multidisciplinary stakeholders, both technical and non-technical, by facilitating efficient preparation of documentation and consistent review.

Reproducibility of published work is a crucial consideration within the pharmaceutical industry and could be important towards establishing standard models for various therapeutic areas. While the benefits of an ‘open science’ approach are debated [27], if QSP models are to be used in support of regulatory interactions then published, peer-reviewed, versions could be highly beneficial to facilitating those interactions (allowing independent detailed review of

model components at a level beyond that which is practical in direct interactions). Tools or structures must be developed to allow for more complete sharing of information used to create all QSP models, in order to increase reproducibility and confidence in these published models. At the very least we propose authors explicitly state, generically, what has been left out of the publication (and why) and what components readers would need in order to continue developing the model (for example, ‘calibrate against in vivo model + compound’). If a full model cannot be shared publicly, it could be helpful to explore whether it is possible for reviewers to assess a manuscript confidentially. Simply put, in our view, a QSP publication where reviewers cannot review and run the model code is not a peer-reviewed paper. While we question the general utility of publishing non-reproducible work, at the very least the integrity of the peer review process should be prioritized.

We propose that QSP model assessment must be flexible depending on the nature of the question asked and the context. A risk-based framework for verification and validation, as described previously [26], can be applied although the details need to be carefully considered given that few QSP projects are alike. Other such frameworks do exist in the literature [9, 28–30], and should be considered. While we are hesitant to prescribe a specific framework, modelers are encouraged to adopt one to better communicate model risk and uncertainty, enabling both technical and non-technical stakeholders to interpret simulation results appropriately. Aligning on assessment criteria for a key QSP application like dose selection may provide a starting point from which standard practices may emerge. This could in turn help regulators formulate guidance for application of QSP towards regulatory submissions and ultimately encourage greater use of QSP models in later stage clinical development to inform decisions not easily addressed by empirical methods.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10928-022-09811-1>.

Acknowledgements The authors would like to acknowledge Svetlana Lyapustina, Lee Nagao, Dede Godstey, Fatou Sarr, and James Vergis from the IQ Secretariat for their assistance administering the survey, and collecting and anonymizing responses. We would also like to acknowledge Wen Chyi Shyu from the Translational ADME Leadership Group at the IQ Consortium for her support and guidance and Mark Penney for thoughtful review.

Author contributions All authors contributed to the survey conception, design, and dissemination. Material preparation and data analysis was performed by Richard Allen, Britton Boras, Valeriu Damian, Frank Gibbons, Jeffrey Kearns, Jangir Selimkhanov, Guanyu Wang, and Susana Zaph. The first draft of the manuscript was written by Jason Chan. Antonio Cabal, Lourdes Curcurull-Sanchez, Andrew Stein, Kenichi Umehara, and Weirong Wang, commented on previous

versions of the manuscript. All authors read and approved the final manuscript.

Declarations

Competing interests The authors declare no competing interests.

Conflict of interest The authors declare no competing interests.

References

1. Sorger PK et al (2011) Quantitative and systems pharmacology in the post-genomic era: new approaches to discovering drugs and understanding therapeutic mechanisms. An NIH white paper by the QSP workshop group. NIH, Bethesda, pp 1–48
2. Nijssen MJMA, Wu F, Bansal L, Bradshaw-Pierce E, Chan JR, Liederer BM, Mettetal JT, Schroeder P, Schuck E, Tsai A, Xu C, Chimalakonda A, Le K, Penney M, Topp B, Yamada A, Spilker ME (2018) Preclinical QSP modeling in the pharmaceutical industry: an IQ consortium survey examining the current landscape. *CPT Pharmacometrics Syst Pharmacol* 7(3):135–146
3. Bai JPF, Schmidt BJ, Gadkar KG, Damian V, Earp JC, Friedrich C, van der Graaf PH, Madabushi R, Musante CJ, Naik K, Rogge M, Zhu H (2021) FDA-industry scientific exchange on assessing quantitative systems pharmacology models in clinical drug development: a meeting report, summary of challenges/gaps, and future perspective. *AAPS J* 23(3):60
4. <https://www.fda.gov/drugs/development-resources/model-informed-drug-development-pilot-program> (accessed 8 April 2022)
5. Bai JPF, Earp JC, Florian J, Madabushi R, Strauss DG, Wang Y, Zhu H (2021) Quantitative systems pharmacology: landscape analysis of regulatory submissions to the US Food and Drug Administration. *CPT Pharmacometrics Syst Pharmacol* 10(12):1479–1484
6. Zineh I (2019) Quantitative systems pharmacology: a regulatory perspective on translation. *CPT Pharmacometrics Syst Pharmacol*. 8(6):336–339
7. Shebley M, Sandhu P, Emami Riedmaier A, Jamei M, Narayanan R, Patel A, Peters SA, Reddy VP, Zheng M, de Zwart L, Beneton M, Bouzom F, Chen J, Chen Y, Cleary Y, Collins C, Dickinson GL, Djebli N, Einolf HJ, Gardner I, Huth F, Kazmi F, Khalil F, Lin J, Odinecs A, Patel C, Rong H, Schuck E, Sharma P, Wu SP, Xu Y, Yamazaki S, Yoshida K, Rowland M (2018) Physiologically based pharmacokinetic model qualification and reporting procedures for regulatory submissions: a consortium perspective. *Clin Pharmacol Ther* 104(1):88–110
8. El-Khateeb E, Burkhil S, Murby S, Amirat H, Rostami-Hodjegan A, Ahmad A (2021) Physiological-based pharmacokinetic modeling trends in pharmaceutical drug development over the last 20-years; in-depth analysis of applications, organizations, and platforms. *Biopharm Drug Dispos* 42(4):107–117
9. Braakman S, Pathmanathan P, Moore H (2021) Evaluation framework for systems models. *CPT Pharmacometrics Syst Pharmacol*. <https://doi.org/10.1002/psp4.12755>
10. Ramanujan S, Chan JR, Friedrich CM, Thalhauser CJ (2019) A flexible approach for context-dependent assessment of quantitative systems pharmacology models. *CPT Pharmacometrics Syst Pharmacol* 8(6):340–343
11. Gadkar, K, Feigelman, J, Sukumaran, S, Rodrigo, M, Staton, T, Cai, F, Bauer, R, Choy, DF, Stokes, CL, Scheerens, H, Ramanujan S. Integrated systems modeling of severe asthma:

- exploration of IL-33/ST2 antagonism. Accepted pending minor revisions, *CPT-PSP*.
12. Rogers KV, Martin SW, Bhattacharya I, Singh RSP, Nayak S (2021) A dynamic quantitative systems pharmacology model of inflammatory bowel disease: part 1—model framework. *Clin Transl Sci* 14(1):239–248
 13. DILIsym Services, Inc. <https://www.simulations-plus.com/software/naflidsym/> (accessed 17 May 2022)
 14. Peterson MC, Riggs MM (2012) Predicting nonlinear changes in bone mineral density over time using a multiscale systems pharmacology model. *CPT Pharmacometrics Syst Pharmacol* 1(11):e14
 15. Dai W, Rao R, Sher A, Tania N, Musante CJ, Allen R (2021) A prototype QSP model of the immune response to SARS-CoV-2 for community development. *CPT Pharmacometrics Syst Pharmacol* 10(1):18–29
 16. van der Graaf PH, Benson N (2018) The role of quantitative systems pharmacology in the design of first-in-human trials. *Clin Pharmacol Ther* 104(5):797
 17. Blake K, Bonelli M, Ponzano S, Enzmann H (2018) European medicines agency committee for human medicinal products “First-in-Human Guideline Drafting Group”, response to: the role of quantitative systems pharmacology in the design of first-in-human trials. *Clin Pharmacol Ther*. 104(5):798
 18. Khurana M, Zadezensky I, Lowy N, Roman D, Guettier JM, Li L, Florian J, Sahajwalla CG, Sinha V, Mehrotra N (2019) Use of a systems pharmacology model based approach toward dose optimization of parathyroid hormone therapy in hypoparathyroidism. *Clin Pharmacol Ther* 105(3):710–718
 19. Venkatakrishnan K, van der Graaf PH (2021) Model-informed drug development: connecting the dots with a totality of evidence mindset to advance therapeutics. *Clin Pharmacol Ther* 110(5):1147–1154
 20. Ma H, Wang H, Sové RJ, Wang J, Giragossian C, Popel AS (2020) Combination therapy with T cell engager and PD-L1 blockade enhances the antitumor potency of T cells as predicted by a QSP model. *J Immunother Cancer* 8(2):e001141
 21. Hart T, Xie L (2016) Providing data science support for systems pharmacology and its implications to drug discovery. *Expert Opin Drug Discov* 11(3):241–256
 22. Ermakov S, Schmidt BJ, Musante CJ, Thalhauser CJ (2019) A survey of software tool utilization and capabilities for quantitative systems pharmacology: what we have and what we need. *CPT Pharmacometrics Syst Pharmacol* 8(2):62–76
 23. Lapteva L, Pariser AR (2016) Investigational new drug applications: a 1-year pilot study on rates and reasons for clinical hold. *J Investig Med*. 64(2):376–82
 24. Wang EB, Shen L, Heathman M, Chan JR (2019) Incorporating placebo response in quantitative systems pharmacology models. *CPT Pharmacometrics Syst Pharmacol* 8(6):344–346
 25. Cucurull-Sanchez L, Chappell MJ, Chelliah V, Amy Cheung SY, Derks G, Penney M, Phipps A, Malik-Sheriff RS, Timmis J, Tindall MJ, van der Graaf PH, Vicini P, Yates JWT (2019) Best practices to maximize the use and reuse of quantitative and systems pharmacology models: recommendations from the united kingdom quantitative and systems pharmacology network. *CPT Pharmacometrics Syst Pharmacol* 8(5):259–272
 26. Kuemmel C, Yang Y, Zhang X, Florian J, Zhu H, Tegenge M, Huang SM, Wang Y, Morrison T, Zineh I (2020) Consideration of a credibility assessment framework in model-informed drug development: potential application to physiologically-based pharmacokinetic modeling and simulation. *CPT Pharmacometrics Syst Pharmacol* 9(1):21–28
 27. Rostami-Hodjegan A, Bois FY (2021) Opening a debate on open-source modeling tools: pouring fuel on fire versus extinguishing the flare of a healthy debate. *CPT Pharmacometrics Syst Pharmacol* 10(5):420–427
 28. Skottheim Rusten I, Musuamba FT (2021) Scientific and regulatory evaluation of empirical pharmacometric models: an application of the risk informed credibility assessment framework. *CPT Pharmacometrics Syst Pharmacol* 10(11):1281–1296
 29. Stein AM, Kearns JD, Kim J, Margolskee A (2021) Cheat sheet for model uncertainty assessment. Zenodo. <https://doi.org/10.5281/zenodo.4409236>
 30. Friedrich CM (2016) A model qualification method for mechanistic physiological QSP models to support model-informed drug development. *CPT Pharmacometrics Syst Pharmacol* 5(2):43–53
 31. Lex A, Gehlenborg N, Strobel H, Vuillemot R, Pfister H (2014) UpSet: visualization of intersecting sets. *IEEE Trans Vis Comput Graph* 20(12):1983–1992
 32. Conway JR, Lex A, Gehlenborg N (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33(18):2938–2940

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.