



# A comprehensive regulatory and industry review of modeling and simulation practices in oncology clinical drug development

Ana Ruiz-García<sup>1,2</sup> · Paul Baverel<sup>3,4</sup> · Dean Bottino<sup>5</sup> · Michael Dolton<sup>6</sup> · Yan Feng<sup>7</sup> · Ignacio González-García<sup>8</sup> · Jaeyeon Kim<sup>9</sup> · Seth Robey<sup>10</sup> · Indrajeet Singh<sup>11</sup> · David Turner<sup>6</sup> · Shu-Pei Wu<sup>12</sup> · Donghua Yin<sup>13</sup> · Di Zhou<sup>14</sup> · Hao Zhu<sup>15</sup> · Peter Bonate<sup>16</sup>

Received: 19 June 2022 / Accepted: 16 February 2023  
© The Author(s) 2023

## Abstract

Exposure–response (E–R) analyses are an integral component in the development of oncology products. Characterizing the relationship between drug exposure metrics and response allows the sponsor to use modeling and simulation to address both internal and external drug development questions (e.g., optimal dose, frequency of administration, dose adjustments for special populations). This white paper is the output of an industry-government collaboration among scientists with broad experience in E–R modeling as part of regulatory submissions. The goal of this white paper is to provide guidance on what the preferred methods for E–R analysis in oncology clinical drug development are and what metrics of exposure should be considered.

**Keywords** Oncology · Exposure–response · C–QT · Markov · Semi-mechanistic · Disease progression · Logistic regression · Time-to-event · Tumor growth dynamics

---

✉ Ana Ruiz-García  
Ana.Ruiz5@gilead.com

<sup>1</sup> Modeling and Simulation, Metrum Research Group, Tariffville, CT, USA

<sup>2</sup> Present Address: Clinical Pharmacology and Pharmacometrics, Gilead Sciences, Foster City, CA, USA

<sup>3</sup> Clinical Pharmacology, F. Hoffmann-La Roche AG, Basel, Switzerland

<sup>4</sup> Present Address: Clinical Pharmacology and Clinical Biomarker, Molecular Partners, Schlieren, Switzerland

<sup>5</sup> Oncology Modeling and Simulation, Takeda, Newton, MA, USA

<sup>6</sup> Clinical Pharmacology, Genentech, South San Francisco, CA, USA

<sup>7</sup> Clinical Pharmacology, Genmab, Copenhagen, Denmark

<sup>8</sup> Simulations Plus, Cognigen, Alicante, Spain

<sup>9</sup> Early Development Analytics, Novartis Institutes for BioMedical Research, Cambridge, MA, USA

<sup>10</sup> Quantitative Pharmacology and Pharmacometrics, Merck, Rahway, NJ, USA

<sup>11</sup> Clinical Pharmacology, Gilead Sciences, Foster City, CA, USA

<sup>12</sup> Clinical Pharmacology and Pharmacometrics, BioNTech SE, Boston, MA, USA

<sup>13</sup> Clinical Pharmacology, Early Oncology, Pfizer Inc., New York, NY, USA

<sup>14</sup> Clinical Pharmacology, Modeling & Simulation, Amgen Inc., Thousand Oaks, CA, USA

<sup>15</sup> Division of Pharmacometrics, Office of Clinical Pharmacology at FDA, Silver Spring, MD, USA

<sup>16</sup> Clinical Pharmacology and Exploratory Development, Astellas, Northbrook, IL, USA

## Introduction

Exposure–response (E–R) analyses of oncology drugs are an integral component in their clinical development [1, 2], both for internal decision making and externally to support regulatory approval. Understanding the relationship between dose, exposure, and response allows the sponsor to demonstrate that they understand the pharmacokinetic–pharmacodynamic (PKPD) behavior of their drug. Variability in pharmacokinetics and pharmacodynamics is well known and there are certain patient subgroups that may be at increased risk for adverse events, e.g., patients with impaired renal and/or hepatic function, or decreased efficacy, e.g., ultrarapid CYP2D6 metabolizers [3]. E–R analysis allows one to predict and/or confirm response in these patient subgroups and whether dose modification is needed in these subgroups.

The term ‘exposure’ is a broad one that tries to capture how much drug a person is “exposed” to. Exposure may encompass any or all of the following:

- Some aspect related to dose, such as the daily dose or total cumulative dose a patient receives;
- Some measure of drug concentration in the body at some point in time, such as maximal concentration ( $C_{\max}$ ), average concentration ( $C_{\text{avg}}$ ) or trough concentration at steady-state ( $C_{\text{trough}}$ );
- Time above some threshold, such as time above some minimum effective concentration, or
- May include integrated measures of concentration, for instance area under the curve at steady-state ( $AUC_{\text{ss}}$ ), cumulative area under the curve, or dose at the time of some response.

Indeed, there are many different possible choices for quantifying exposure over time and space (e.g., systemic or tumor). Similarly, ‘response’ is also a broad term that may encompass any or all of the following:

- Assessment of drug efficacy: response rate, time to progression, progression free survival (PFS) or overall survival (OS);
- Assessment of drug safety: whether a patient experiences nausea or vomiting, or the severity of a rash a patient may develop during treatment; or
- Measurement of pharmacodynamic biomarkers: degree of phosphorylation of some important protein, or level of receptor occupancy, etc.

Sponsors may submit several E–R analyses in support of a regulatory submission that encompass different measures of exposure and different measures of clinical outcomes. Identifying dose/dosing regimens that provide a favorable benefit/risk profile is critical in drug development [4–6].

Sponsors may also conduct E–R analyses for internal decision making for go/no go decisions or to inform dose selection for Phase 2/3 trials.

In modern oncology, the old paradigm that “more is better” with the maximum tolerated dose (MTD) being used as the dose studied in late phase studies is no longer the norm. The MTD is commonly defined as the highest dose that most patients can tolerate without unacceptable side effects. However, the introduction of targeted therapies, biologics in general, and immuno-oncology is moving drug development away from the MTD concept in search of the optimal dose/dosing regimen. Recently, the Food and Drug Administration has issued a draft guidance that will require to sponsors to study a range of doses in clinical development with the goal to use the optimal dose in registration trials [7]. As such, alternative approaches to clinical development will be required for dose optimization, facilitating the use of E–R analysis by including safety and efficacy information from more than one dose level.

One of the earliest, most comprehensive, E–R analyses was presented by Houk et al. [8] in patients with advanced solid tumors, including patients with gastrointestinal stromal tumor (GIST) and metastatic renal cell carcinoma (mRCC) treated with sunitinib. Their analyses used different measures of exposure: dose, drug systemic concentration,  $AUC_{\text{ss}}$ , and cumulative AUC during 1 cycle of treatment. Using a combination of population PKPD modeling, repeated-measures logistic regression, and correlation analyses, authors demonstrated that increased sunitinib exposure was associated with longer time to progression, longer overall survival, and a greater chance of clinical response. They also showed that increased exposure was associated with increased blood pressure, increased incidence of fatigue, and a greater probability of neutropenia. These analyses were supportive of the recommended dosing regimen in GIST and mRCC patients that was approved at 50 mg once daily for 4 weeks every 6 weeks (4 weeks on/2 weeks off). Recently, a less intense dose regimen has been proposed for sunitinib: 50 mg for 14 days every 3 weeks (traditional “2/1 schedule”). Both regimens have the same dose intensity (4 weeks on) and both have a 2-week drug holiday period every 6 weeks. The model from Houk et al. mathematically explains why the alternate schedule (2/1) presents less toxicity, and therefore, it is better tolerated [9].

Typically, due to lack of established standard methods, there is no one standard approach to conducting an exposure–response analysis. For example, when performing time-to-event analysis, Kaplan–Meier (KM) curves are a useful graphic assessment for the exploration of observed data. Further analysis such as Cox proportional hazards regression model, parametric time-to-event models, or an accelerated

failure time model could also be considered based on the nature of the data. All of these are equally valid and have different assumptions with different pros and cons, and different prediction outputs. Furthermore, it is important to mention that E–R analyses in oncology using survival metrics, may be subject to selection bias and immortal time bias [10]. Landmark analyses and multi-state analyses have been proposed to correct the inherent selection bias resulting in part from the fact that responders must live long enough for response to be observed [11–13]. Therefore, the question arises as to which method should be used and if there is a preferred method of choice. In an attempt to address these concerns, this white paper is the output of an industry-government collaboration between scientists with broad experience in E–R modeling in oncology. The goal of this white paper is to provide guidance on the preferred methods for E–R analysis commonly used in oncology and the measures of exposure to consider in different scenarios. To make things easier for the reader, an executive summary of recommendations and comments from each of the following sections is presented in Table 1.

## Exposure metrics considerations

When performing E–R analyses, one of the first decision points in the analysis is the exposure metric of choice. One needs to consider what type of data are available, the duration of treatment, and dose compliance. It is not uncommon in oncology for patients to experience dose modifications during the course of treatment [14, 15]. Often, these dose adjustments happen because of adverse events (AEs) that require either a dose holiday or dose reduction. The result of the dose adjustments could lead to lower average drug exposure for subjects with long treatment duration. Further, a high percentage of dropouts is expected due to AEs associated with concurrent exposure. Thus, under this situation, an E–R analysis for some of the most common efficacy endpoints (overall survival, OS, or progression-free survival, PFS) may suggest an inverse relationship between exposure and efficacy. To eliminate the bias introduced by the large percentage of dose reductions, an earlier exposure metric prior to any dose modification (Cycle 1) could be considered. However, this early exposure metric will have limited value in establishing any relationship between exposure and response. Under scenarios where AEs are leading to significant dose adjustments, it is important to consider if the right dose has been identified and whether there is any longitudinal model for efficacy endpoints or surrogate endpoints of efficacy that can be evaluated to better understand the E–R relationship. The oncology dose-finding workshop organized by FDA and the American Association for Cancer

Research (AACR) in 2016 presented levatinib as an example of dose adjustment integrated E–R analysis (DAIER). For levatinib, FDA suggested an E–R analysis using dose-altering AEs models to evaluate different dosing regimens and efficacy [16].

Typically in E–R analyses, drug exposure is assumed to be the cause, and response to be the outcome. However, if disease progression or remission influences pharmacokinetic (PK) parameters over time, this interaction between treatment response and PK parameters could result in artificial E–R relationships. Anti-programmed death-1 (anti-PD1) immunotherapies nivolumab and pembrolizumab exhibited time-dependent pharmacokinetics and a correlation between drug clearance changes over time and survival rates [17–19]. In these situations, directly linking drug exposure at steady state to clinical outcomes in a single-dose trial may yield an over-steep E–R relationship, deviating from the true underlying relationship. Interestingly, the nivolumab baseline clearance had a strong association with survival, relative to all evaluated exposure and covariates in a multivariable E–R analysis [20, 21]. Wang et al. showed that baseline nivolumab clearance can be predicted by a composite of cytokine signatures using machine learning approach and the patients with predicted high nivolumab's clearance (CL) is associated with poor survival regardless of treatment (nivolumab or chemotherapy) in patients with advanced melanoma and renal cell carcinoma (RCC) [22, 23]. These results support the hypothesis that nivolumab CL can be used as a prognostic marker for patient disease status. Moreover, this highlighted the importance of including more than one dose level in E–R analysis to reduce the confounding effect between exposure and CL. It has been demonstrated by Liu, et al., through simulation, that using exposure variables observed or derived from the first treatment cycle for an E–R analysis may minimize this bias [18]. Furthermore, confounded relationships between baseline risk factors for survival and drug exposure have also been reported, complicating the choice of exposure metric to use in these circumstances and the interpretation of any observed E–R relationship [20, 24]. An example of confounded baseline risk factors has been reported for trastuzumab, which is indicated for the treatment of HER2 overexpressing breast cancer, metastatic gastric, and gastroesophageal junction adenocarcinoma [25]. Yan et al. performed an exploratory analysis with simulated trough concentration in the first treatment cycle and overall survival. The Kaplan–Meier curves stratified by different exposure quartiles suggested a E–R trend based on the exposure metric. However, the unbalanced distribution of baseline disease burden across different exposure quartiles was responsible of the apparent E–R relationship. To minimize the confounding effect, a propensity matching

**Table 1** Executive summary

Topic	Recommendations and comments
Exposure metric considerations	<ul style="list-style-type: none"> <li>• This is a key component of any analysis and may include dose, concentration, time-averaged concentration, time above a threshold, or area-related metrics</li> <li>• Be careful of dose adjustments and dropouts, and their effect on exposure metrics</li> <li>• In choosing a metric consider whether the relationship may be a direct effect (like nausea/vomiting) or a time delay effect like tumor growth</li> </ul>
Safety and efficacy endpoints: categorical endpoints <ul style="list-style-type: none"> <li>• Examples: presence/absence of nausea/vomiting, presence/absence of grade 3 or higher neutropenia, RECIST</li> </ul>	<ul style="list-style-type: none"> <li>• Primarily logistic modeling (or some modification thereof) is used</li> <li>• 10 Events/predictive variable is recommended for precise estimation of regression coefficients</li> <li>• Consider the confounding effect of drug clearance on the outcome (higher clearance may lead to poorer treatment outcomes); use clearance as a covariate in the model</li> <li>• If time-dependent categorical endpoints are of interest, use Markov models</li> </ul>
Safety endpoints <ul style="list-style-type: none"> <li>• Time course of myelosuppression</li> </ul>	<ul style="list-style-type: none"> <li>• Empirical approaches (e.g., maximum % decrease from baseline vs. AUC)               <ul style="list-style-type: none"> <li>◦ Empirical models have limited values extrapolating outside the doses or dosing regimen tested</li> <li>◦ Need to use some integrated measure of exposure like AUC because of time delay between first dose and peak effect</li> </ul> </li> <li>• Semi-mechanistic models, like a Friberg model, allow assessment of time course of myelosuppression</li> </ul>
<ul style="list-style-type: none"> <li>• QTc interval prolongation</li> </ul>	<ul style="list-style-type: none"> <li>• Recommend following Garnett white paper [63]</li> <li>• Oncology trials may not be able to study 2x-above therapeutic exposure for safety reasons</li> <li>• 20 ms is generally accepted as the upper safety threshold compared to 10 ms in healthy volunteers</li> </ul>
Efficacy endpoints <ul style="list-style-type: none"> <li>• Time to event: survival</li> </ul>	<ul style="list-style-type: none"> <li>• Can use nonparametric (Kaplan–Meier), semiparametric (Cox proportional hazard), or parametric (accelerated lifetime) models</li> <li>• KM curves often assessed by quartile of exposure vs OS or PFS</li> <li>• CPH models should include other prognostic covariates, like baseline tumor size and drug clearance, for controlling these confounding effects</li> <li>• May be subject to inherent selection bias and immortal time effects</li> </ul>
<ul style="list-style-type: none"> <li>• Tumor growth dynamics</li> </ul>	<ul style="list-style-type: none"> <li>• Allows for a better understanding of the entirety of a patient’s tumor burden growth/shrinkage time-course to assess the possible impact of dose or schedule selection on disease response</li> <li>• Many different models to choose from</li> <li>• Secondary parameters may be more intuitively linked with survival outcomes in time-to-event analyses</li> <li>• Informative censoring may affect parameter estimates</li> <li>• Pretreatment tumor growth trajectories may allow better interpretability</li> <li>• Rely on prespecified target lesions which may not be indicative of overall disease burden</li> </ul>
Hematologic malignancies	<ul style="list-style-type: none"> <li>• All recommendations in the above sections can be applied to hematologic malignancies, where total target tumor size is replaced by the appropriate continuous tumor burden metric for that particular malignancy</li> <li>• May require bounded endpoint models, e.g., minimal residual disease</li> </ul>
Tumor biomarker and disease progression	<ul style="list-style-type: none"> <li>• Example: PSA kinetics or circulating tumor cells</li> <li>• May require semi- or mechanistic models to explain</li> </ul>

**Table 1** (continued)

Topic	Recommendations and comments
Immunogenicity	<ul style="list-style-type: none"> <li>• May affect both exposure and response (safety/efficacy)</li> <li>• Need to consider anti-drug antibodies (ADAs) vs neutralizing antibodies</li> <li>• Covariate analysis may include binary grouping (presence or absence of ADAs) or titer in models</li> </ul>
Cell therapies	<ul style="list-style-type: none"> <li>• CAR-T cells display strange kinetics compared to traditional small molecules or biologics</li> <li>• 4 Phases including distribution, expansion, contraction, and persistence</li> <li>• Cannot use allometric principles for scaling of doses</li> </ul>

Each topic in the table is covered in the text. Bulleted recommendations and comments are excerpts from the text. Read text for further details

strategy for adjusting measured confounders, which are defined by a stepwise Cox regression model, was applied. After appropriate matching, patients in the first exposure quartile of trastuzumab show no survival benefit over control [26].

Although dose could be used as the exposure metric of choice, systemic exposure (i.e., plasma/serum/blood drug concentration) often is a more precise metric as it accounts for nonlinearities and inter-individual variability in the pharmacokinetics of the drug. Depending on the type of E–R analyses, metrics of early exposure (and its correlations with surrogates of efficacy), exposure at steady state, or the time course of drug concentration (time-varying concentration) could be used.

Another important consideration when choosing the exposure metric is potential dose regimen comparisons. Prediction of response in other schedules of administration based on just one dose schedule often will lead to inaccurate outcomes. However, when information is available from more than one dose schedule, evaluating the most sensitive metric of exposure (e.g.,  $C_{\text{trough}}$ ,  $C_{\text{max}}$ ,  $C_{\text{avg}}$ ) for clinical outcomes may help to appropriately account for differences in dose schedules. In dose-escalating studies, looking at concentration or dose versus biomarker changes as surrogates of efficacy or proof of biological response could help dose selection and establish the maximum tolerable dose that will lead to maximum biological response. Further, it is worth mentioning that in order to collect sufficient information over an informative range of doses or exposure, an adaptive/Bayesian design could be a good choice. However, such study designs may cause logistical and operational challenges [27].

Therefore, the use of the drug exposure metrics depends on the study design, the drug mechanism of action, and the

nature of the relationship between exposure and response (i.e., short versus long-term effects). In oncology when there is a single read on efficacy (i.e., objective response rate, ORR), it may be more appropriate to use a simple metric that represents the drug exposure over the course of the treatment. In this case,  $C_{\text{avg}}$  or  $C_{\text{trough}}$  could be good metrics of choice since these provide an average measure of exposure; however for E–R safety analyses with acute AEs, the  $C_{\text{max}}$  prior to the AE event could be explored. Cumulative AUC is confounded with time on study and careful consideration should be given to the use of this metric, as well as the nature of the endpoint under study. AUC at steady state ( $AUC_{\text{ss}}$ ) is a valid exposure metric often associated with long-term effects. However, when looking at steady state metrics (i.e.,  $C_{\text{trough}}$ ,  $C_{\text{avg}}$ ,  $AUC_{\text{ss}}$ ,  $C_{\text{max,ss}}$ ) for a given dose schedule, those may be correlated and selecting one metric versus another will often not lead to different conclusions.

Another consideration is the use of model-predicted drug exposure versus observed concentration values. In the case of a drug with very high variability in exposure (i.e., > 70% residual error) and sparse PK sampling, model-predicted exposure profiles may be questionable and the use of observed  $C_{\text{trough}}$  values over the course of the treatment could be a more reliable exposure metric.

In summary, a variety of exposure metrics could be considered when performing E–R analysis; what should guide the selected drug metrics for E–R is multifactorial and includes the type of E–R analysis, endpoint under consideration, understanding of the drug pharmacokinetics and mechanism of action, the available drug exposure data, and the nature of safety and efficacy endpoints under analysis. Table 2 provides a summary of exposure metrics used for different E–R analyses and clinical response (efficacy/safety) endpoints.



**Table 2** Summary of types of analyses and exposure metrics to consider

Exposure metric	Analysis	Endpoints
Single exposure metric representing the course of treatment or exposure to the time of the event: * $C_{\text{trough}}$ , * $C_{\text{avg}}$ , * $AUC_{\text{ss}}$ , * $C_{\text{max}}$ , $C_{\text{obs}}$ , *time above a threshold concentration	KM Curves, Cox proportional hazards Logistic regression Empirical myelosuppression models	PFS, EFS, OS ORR, CR, MRD, AE grades, Myelosuppression grades
Longitudinal exposure (time-varying exposure): * $C(t)$ , * $C_{\text{trough}}(t)$ , * $C_{\text{ave}}(t)$	Multi-state models Parametric time-to-event Markov models Tumor growth dynamics Time-course of myelosuppression Disease progression models	SLD-TTE PFS, EFS, OS, AE events SLD neutropenia, lymphopenia, anemia, and thrombocytopenia, Time-course of biomarkers as surrogates of efficacy

*AE* adverse event, *AUC<sub>ss</sub>* area under the curve at steady-state, *CR* complete response, *C<sub>avg</sub>* average concentration calculated as the ratio of cumulative AUC over the timeframe for the cumulative exposure, *C<sub>max</sub>* maximum concentration, *C<sub>obs</sub>* observed concentration, *C<sub>trough</sub>* concentration prior to dose, *EFS* event free survival, *KM* Kaplan–Meier, *MRD* minimal residual disease, *ORR* objective response rate, *OS* overall survival, *PFS* progression free survival, *SLD* sum of the longest diameter, *TTE* time-to-event

\*Model predicted/estimated exposure metrics

## Safety and efficacy endpoints: categorical scoring systems

### Logistic regression

In addition to continuous clinical endpoints, categorical or ordered categorical (ordinal) endpoints, such as graded AEs or ORR, are often considered. Depending on the granularity of data collected and the objective of the analysis, logistic regressions and Markov chain models can be used to analyze these endpoints.

Logistic regression modeling is a widely used approach in E–R analysis that enables characterization of the relationship between drug exposure and ordered categorical clinical outcomes [20, 28–31]. In logistic regression analysis, a linear predictor with a link function is used. The linear regression yields a nonlinear relationship commonly through a logit predictor, but other links could be used, such as probit or complementary log–log models. Logistic regression models the probability of an outcome (binomial dependent variable) based on independent variables such as demographic factors, drug exposure, etc. This type of analysis informs about the likelihood of an event happening.

The characterization of relationships between exposure and clinical outcomes of both efficacy [21] (i.e., objective response, based on short-term tumor response) and safety [32] (i.e., AEs of clinical interest) provides a quantitative assessment of the benefit/risk profile and is often used for dose selection in late-stage (i.e., Phase 3) oncology drug development. In some situations, these results can be

utilized for dose recommendation in regulatory submissions (i.e., a rolling submission), if the treatment fits an unmet medical need and short-term clinical outcome (e.g., OR) is promising [33, 34].

The development of a logistic regression model generally includes 3 steps; (1) development of a base model, with an evaluation of the appropriate functional form of exposure metrics that may include interaction terms when the drug is given in combination with another drug; (2) development of a full model, including assessment of covariates of interest; (3) and development of the final model after considering the contribution of all potential variables of interest. Even though a final model can be simplified using backward elimination to achieve parsimony, the presence or absence of parameters of minimal impact in the probability will not result in model prediction differences [35].

A minimum of 10 events per predictive variable analyzed in the logistic regression model is recommended for accurate and precise estimation of the regression coefficients. Peduzzi et al. evaluated the effect of the number of events per estimated parameters in logistic regression. They concluded that less than 10 events per predictive variable can lead to major biases, questioning the validity of the logistic regression analysis if those conditions are not fulfilled [35, 36]. Likewise, the number of covariates to be included in the full model should consider the total events in the analysis dataset to avoid over-parameterization. The utilization of the full model provides the benefit of avoiding biased parameter estimation by accounting for all measured covariate effects. Moreover, the full model

avoids confounding among covariates and exposure metrics [20, 31]. A recent E–R efficacy analysis of nivolumab showed how baseline CL was a significant predictor of efficacy endpoints, OR, and OS. Subjects with higher CL had poor efficacy outcomes (i.e., higher risk of death and lower OR). Prior E–R analyses found only apparent E–R relationships that were misleading by ignoring the incorporation of baseline CL. It is important to note that baseline CL was not included as an exposure parameter but as a surrogate of disease status and other baseline confounders. In fact, exposure did not influence the outcome. Subjects within a given dose level with the lowest CL were more likely to be responders. No relationship was found comparing exposure levels and clinical response across the range of doses evaluated. The recent approval of flat dosing of nivolumab (240 mg Q2W, and 480 mg Q4W) away from originally approved body weight normalized dosing (3 mg/kg Q2W) was based on a flat E–R relationship that included both baseline CL and systemic exposure in the model, and that increasing drug exposure was unlikely to improve efficacy outcomes (thus, indicative of maximal response). This example highlights the importance of utilizing a full model approach for trials with limited dose-finding information to assess covariates' effect and exposure levels on relevant clinical outcomes [30, 37].

### Longitudinal logistic regression models with markov elements

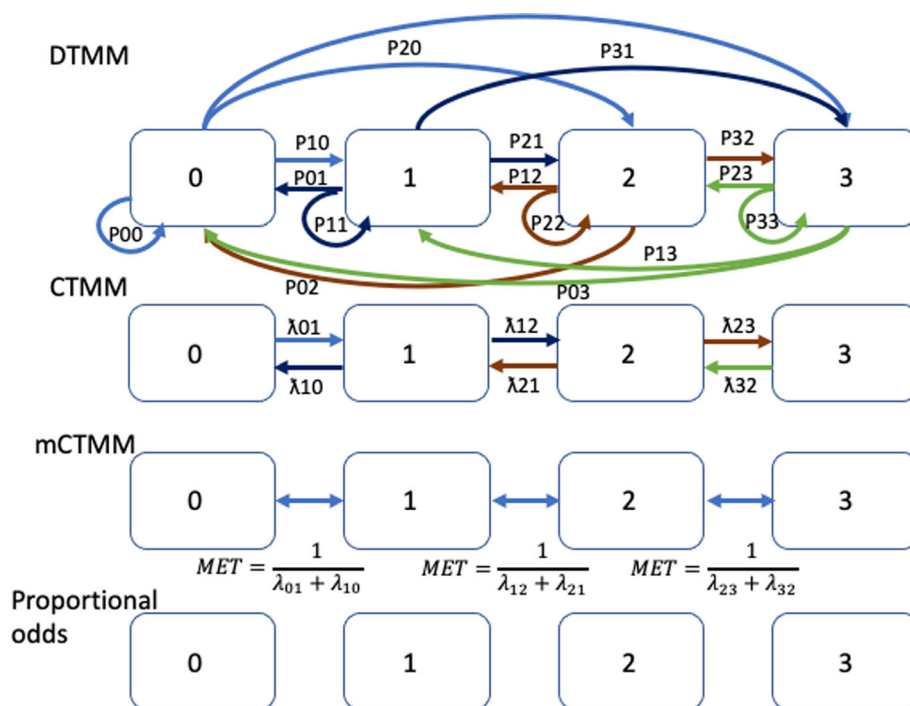
E–R analysis for categorical endpoints commonly focuses on correlating the highest AE grades observed during the course of treatment within a patient's measure of exposure. However, this approach ignores the time course of drug exposure and AE development, and, as such, loses valuable information. Markov models are useful when the outcomes are categorical measures that are monitored continuously over time.

Markov modeling has been widely used in ordered and non-ordered categorical analyses in many therapeutic areas of drug development [28, 38–41]. The Markov models considered in this review are first-order models, Markov chains of second or higher orders are models in which the probability of one event changing depends on 2 or more preceding ones and will not be discussed here. A central concept in Markov models is the transition probability, which models the probability of one event changing to another event (or staying the same). For example, for the 2-event state, like whether a patient develops a rash after starting treatment with a drug, there are 4 transition probabilities. If '0' is no rash and '1' is rash, then the transition probabilities are  $P_{00}$ ,  $P_{01}$ ,  $P_{10}$ ,  $P_{11}$  where the subscript refers to a change from the current state  $i$  to the future state  $j$ . Compared to logistic regression models where there are

no transition probabilities and characterization focuses on the probability of having an event at given drug exposure, Markov modeling provides a description of longitudinal clinical data over time and assumes that (i.) the distribution of future states depends only on the current state and (ii.) is not a function of the whole history of events (first-order Markov models). Given this defined Markov property [42, 43], Markov modeling can capture the onset of events, their duration, and severity as they change over time; it also allows the assessment of an exposure effect on the transition probability from one event state to another state and the severity of the event.

For E–R analyses using Markov models where longitudinal categorical data are modeled, drug exposure over time would be a more appropriate exposure metric than a single exposure measurement, such as  $C_{max}$  or AUC up to the event. Time-varying exposure would be more associated with the onset and duration of AEs when the dose was modified or discontinued due to adverse events. The effect of co-medication on transition probabilities can also be considered if co-medications were applied for the treatment of specific AEs and with the resolution of AEs. If a delay between exposure and clinical outcome was observed, effective concentrations derived from an effect compartment can be evaluated. Prognostic factors (e.g., biomarkers) could be included to assess their effects on the severity of an AE during model development.

Different types of Markov models have been reported and will be briefly presented here. The discrete-time Markov model (DTMM) combines proportional odds with a transition model that allows event changes more than 1 grade higher or lower. In this model, all possible transit probabilities can be estimated, with the assumption that the transit probabilities are independent of whatever the time interval is between the two assessments, making it an ideal candidate when we have uniform time interval assessments. The continuous-time Markov model (CTMM) combines proportional odds with a transition model that prohibits event changes more than 1 grade higher or lower. For example, a transition directly from State 3 to State 1 is prohibited, whereas the transition from State 1 to State 2 and then from State 2 to State 3 is allowed, and vice-versa. With a CTMM, the influence of the previous state on the probability of the current state changes with time (usually decreasing over time as the time interval between measurements increases). Thus, this model is preferred when the observation intervals are non-uniform across patients either due to study design or missing observations. Lastly, the minimal CTMM (mCTMM), which is a simplification of the CTMM, is characterized by independent transition rates between two consecutive states and governed by a single parameter, the mean equilibration time (MET). A schematic of these models is presented in Fig. 1. Schindler



**Fig. 1** Schematic representation of different Markov models and proportional odds model adapted from Schindler et al. [44]. The preference of selection of one type of Markov model vs. others depends upon the frequency of data collection (time interval of events) and a number of categories (states) of the events. DTMM might be preferred when data were collected with uniform time intervals, whereas CTMM might be suitable for data collected with

non-uniform time intervals across patients either due to study design or missing observations. CTMM models, in general, have fewer parameters relative to DTMM, as CTMM assumes that the transitions only occur between neighboring states. mCTMM is a simpler version of the CTMM model, in which mean equilibration time (MET) between two succeeding states is assumed to be constant across states

et al. showed a few examples of model performance between DTMM and mCTMM suggesting that mCTMM had the potential in describing the data reasonably well with a more parsimonious model structure relative to the CTMM [44]. In addition, the effect of covariates (e.g., exposure, biomarker) on the probability of each state can be described in a single relationship. Lu et al. compared the model performance between a proportional odds model, CTMM, and DTMM in which models were developed using weekly based time-course of muscle spasm AE data. Model performance showed that the odds model was influenced by time–frequency, DTMM was unable to describe unevenly spaced data, and CTMM seemed to perform well in all evaluated data frequencies (daily, weekly, and unevenly spaced) [45].

## Safety endpoints

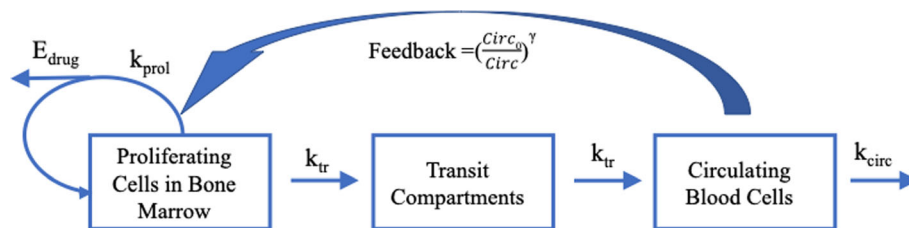
### Time-course of myelosuppression

Blood cell production in the bone marrow is a highly prolific process, making it susceptible to the inhibitory effects of anti-cancer agents with anti-proliferative

activities or with immuno-modulatory agents directly targeting markers expressed on the surface of hematopoietic stem cell progenitors or mature blood cells. In fact, myelosuppression manifested as decreases in circulating red blood cells, white blood cells, or platelets, is among the most frequent AEs observed for anticancer therapeutics [46, 47]. Quantitative understanding of the drug effects on blood cell production is important for the assessment or prediction of the myelosuppression risk, as well as the optimization of the dose and regimen, to reduce myelosuppression-related AEs.

The dose/E–R relationships for drug-induced myelosuppression have often been analyzed by empirical or mechanism-based modeling approaches [48, 49]. Empirical models are usually developed by a theoretical understanding of drug behaviors with very few assumptions of the data [50]. Semi-mechanistic disease models use simplified biological systems to describe the available data falling between empirical models and mechanistic models. Empirical modeling approaches involve either regression-based correlation analyses between a descriptor of the blood cell change (such as maximum % decrease from baseline, often called the nadir, or incidence of particular AEs) and the dose or a particular drug exposure parameter





**Fig. 2** Key structural features of Friberg and related models for myelosuppression.  $E_{drug}$  drug effect,  $k_{prol}$  proliferation rate constant,  $k_{tr}$  maturation rate constant,  $k_{circ}$  circulating cells elimination rate

constant,  $Circ_0$  circulating blood cells at baseline,  $Circ$  amount of circulating cells,  $\gamma$  impact factor on feedback

[such as the area under the curve (AUC) and time above a threshold concentration] or by empirically linking the dynamic change in blood cell counts to a time-variant drug exposure parameter through a particular function (e.g.,  $E_{max}$  models). Semi-mechanistic and mechanism-based modeling approaches utilize differential equations to describe the physiological process of hematopoiesis and the pharmacological perturbation by the inducing agents. In practice, the choice of a specific modeling approach in evaluating drug-induced myelosuppression often depends on the purpose of the analysis and the type of data available.

Empirical myelosuppression models are often expressed as the absolute or relative decrease from baseline at nadir, the maximum percentage of decrease from baseline, the duration below a threshold cell count, the area between a threshold line and the observed cell counts vs time curve, or the incidence of a graded hematological adverse event [51]. Typical exposure parameters such as AUC, time above a threshold concentration, and  $C_{max}$  may be explored in the analysis. The relationship between drug exposure and the myelosuppressive effects is modeled without regard to the time course of drug concentrations or blood cell counts. This type of correlation analysis may be used to determine the myelosuppression response or outcome associated with a certain dose/exposure level. This analysis is relatively easy to implement and does not require complete blood cell time course data obtained from extensive sampling. However, this type of modeling analysis has no or limited value in predicting myelosuppression time courses or responses beyond the tested dose range and regimens.

Empirical longitudinal models may also describe the time course of blood cell change following drug administration [52, 53]. Because there is a typical delay in the myelosuppressive effect in relation to systemic drug concentrations, empirical longitudinal modeling assumes a direct drug effect from a drug exposure parameter (e.g., cumulative AUC or  $C_{average}$ ), with or without adding a lag time parameter. This empirical modeling allows prediction of blood cell count time courses under certain conditions; however, given the often lack of physiological meaning for the PD parameters in these models, there are also

challenges to extrapolate the models to untested conditions in many cases, such as cross-species translation or predicting the effects for similar compounds.

Semi-mechanistic modeling of myelosuppression is based on an understanding of the hematopoiesis process and how the drug perturbs the process. Anti-cancer agents may cause bone marrow suppression through direct cytotoxicity on differentiated bone marrow cells, inhibition of progenitor or precursor cell proliferation, or disruption of growth factor signaling pathways involved in differentiation [54, 55]. There have been generations of mechanism-based mathematical models to describe drug-induced myelosuppression over the last two decades [56, 57]. The most commonly cited model is the Friberg model, which has been the basis of similar models with various modifications [48, 49, 58]. The Friberg model and its related models share the following key structural components (Fig. 2): (1) one or more proliferating compartments with a pool of proliferating cells that can be derived from self-renewable HSCs in the bone marrow; (2) a series of transit compartments representing nonproliferating cells at different maturation stages in the bone marrow; (3) a compartment representing circulating cells with natural turnover; (4) a negative feedback loop where circulating cells regulate the proliferation of bone marrow cells in the proliferating compartment. These structural features are represented by ordinary differential equations, with system-related parameters inherent to the body system and drug-specific parameters that vary by the inducing agents. The drug effects are incorporated into the models in a way consistent with the myelotoxicity mechanisms.

The Friberg and related models have been utilized to describe various types of drug-induced myelosuppression effects including leukopenia, neutropenia, thrombocytopenia, and anemia [51, 59–61]. These models have been able to capture the delay in myelosuppressive effects relative to systemic drug concentrations, and the recovery, rebound and return to baseline for circulating blood cells upon treatment cessation. Despite the complexity of the models, the total number of parameters remains identifiable.

With the mathematical representation of the physiological and pharmacological processes involved in

myelosuppression, the models allow the estimation of system-related and drug-specific parameters. The semi-mechanistic basis enhances the confidence in using these models to extrapolate beyond tested conditions. These models have been used to predict the time courses of circulating blood cell profiles from different doses or regimens. Also, the models can be used in the cross-species translation of myelosuppressive effects [62]. Integration of these semi-mechanistic models into population PK/PD modeling can assess the inter-patient variability and influential covariates of drug-induced myelosuppressive effects.

Since the introduction of this semi-mechanistic model, it has become the golden standard approach to model the myelosuppressive effects of chemotherapy. However, due to the lack of clinical data characterizing the upstream processes of granulopoiesis, some assumptions must be made, such as a constant transit time between maturation compartments. In addition, since the semi-mechanistic model is largely data-driven, the predictive ability of this model might be limited compared with the full mechanistic model [63].

Several researchers have extended the applications of this model by incorporating more complex mechanisms and relationships. For example, Quartino et al. integrated granulocyte colony-stimulating factor (G-CSF)—myelosuppression model to describe the dynamics of endogenous G-CSF and absolute neutrophil count (ANC) following chemotherapy [64]. The final model captured both the initial rise in endogenous G-CSF concentrations following chemotherapy-induced neutropenia and the subsequent return to baseline for G-CSF and ANC. This semi-mechanistic model adequately described the time-course of ANC where the feedback mechanism of G-CSF regulated the neutrophil production and maturation in the bone marrow.

### QTc interval prolongation

Since adoption in the ICH E14 guidance in 2015, concentration-QTc (C-QTc) analyses have rapidly replaced Thorough QT (TQT) studies for the assessment of QT prolongation risk during small molecule oncology development [65–68]. However, the complexities of oncology drug development, including the quick pace of development, differences in trial design, co-medications, and risk–benefit profile in the face of life-threatening disease, pose a number of unique challenges in applying this methodology to exclude a risk of QT prolongation or accurately quantifying the effect size when the drug has a known QT liability. The scope of this section relates primarily to small molecule development; as specified by ICH E14, large

targeted proteins and monoclonal antibodies have a low likelihood of direction channel interactions, and a thorough QT/QTc study (or C-QTc analysis replacing this study) is generally not necessary unless the potential for proarrhythmic risk is suggested by mechanistic considerations or data from clinical or nonclinical studies [66].

A scientific white paper on concentration-QTc modeling published in 2018 provides clear guidance and recommendations on standardizing C-QT analyses intended for assessing QTc prolongation risk under ICH E14 in healthy volunteers [68]. In addition to guidance on Phase 1 study design in order to support a C-QTc analysis, a pre-specified linear mixed-effects (LME) C-QTc model was proposed, including variations on this model depending on the available data, and provides guidance on exploratory and goodness-of-fit plots for model evaluation [68]. Assumptions of the pre-specified LME C-QTc model, including lack of drug effect on heart rate (HR), adequacy of the HR correction used for the QT interval (i.e., lack of trend on QTcF vs RR plot), lack of hysteresis between concentration and QTc effects, and a linear C-QTc relationship (vs non-linear relationships) should be explored and justified during model development [68].

Several elements that are important components of a C-QTc analysis in non-oncology drug development are typically unavailable for an oncology program—namely concentration and QTc data at a suprathreshold exposure, and inclusion of data from placebo subjects, as is typically available during single- or multiple-dose escalation cohorts in many non-oncology programs conducted in healthy volunteers [68]. The lack of placebo subjects in oncology trials typically requires that inferences from a C-QTc analysis are drawn based on baseline-corrected QTc (i.e.,  $\Delta$ QTc), rather than baseline-corrected, placebo-corrected QTc ( $\Delta\Delta$ QTc). In addition, the lack of placebo data introduces diurnal fluctuation in QTc as a potential confounding factor for drug effect on QTc intervals [69]. The collection of time-matched baseline (i.e., at the same time points to be collected post-dose) in order to account for diurnal fluctuation has been successfully implemented for C-QT analysis of single-arm trials [65, 68, 70]; where only a pre-dose baseline was available, the inclusion of categorical time effects in the C-QTc model has recently been proposed [71]. Where a compound is known to substantially affect HR (i.e., mean change > 10 bpm), the inclusion of a time-matched baseline can allow a patient-specific HR correction to be calculated from baseline QT/RR data, although consensus has not been achieved on the optimal approach [68, 72].

Due to safety concerns as well as the differences in risk–benefit profile in oncology, the highest tested dose and

exposure are frequently also the therapeutic exposure, and meeting the requirement for a suprathreshold exposure is not possible [65, 68]. This may change with recent regulatory requests to greater explore the dose–response relationship in early clinical studies. Currently, however, recent draft revisions to ICH S7B and E14 may enable a greater number of oncology small molecule development programs to meet the suprathreshold exposure requirement [73, 74]. When a compound meets the definition of a ‘double negative nonclinical assessment’ for QTc prolongation, a suprathreshold exposure at the ‘high clinical exposure’ (increase in exposure under the effect of intrinsic or extrinsic factors at the maximum therapeutic dose) is required to exclude a positive control, rather than  $\geq$  twofold the high clinical exposure as required under the previous guidance [66, 73, 74]. As discussed in the ICH E14 Q&A guidance documents, in the absence of a positive control or suprathreshold exposure, there is a reluctance to conclude a lack of effect on QT, however, if the upper bound of the two-sided 90% confidence interval around the estimated maximal effect on  $\Delta$ QTc is less than 10 ms, the treatment is unlikely to have an actual mean effect as large as 20 ms [66, 74].

The sample size is also an important consideration for C-QTc analysis. While general guidance of 4–8 subjects on drug and 2–4 subjects on placebo across at least 4 dose cohorts has been proposed based on low false-negative and false-positive rates [68], these studies generally included placebo subjects and/or a suprathreshold exposure typical of a non-oncology program [75–77]. A recent review of QT prolongation risk assessment of small molecule oncology NDAs from 2011 to 2019 found that where a C-QTc analysis was performed, sample size varied greatly, but was generally smaller when the C-QTc dataset was based on data from early phase studies ( $\sim$  20–300 patients), and larger where data from later phase studies (or pooled early and late phase) was used ( $\sim$  100–800 patients) [65]. It was noted that no clear trend was identified between sample size and the labeling recommendation category to which an NDA was assigned [65].

Other considerations that may be more frequently encountered in oncology include pooling of data from patients with different underlying malignancies across treatment arms or trials in order to increase sample size or dose range, which may increase the risk of confounders such as differences in health status and concomitant medications, as well as in study conduct or ECG acquisition or analysis. When pooling of data is required, between-study differences and potential bias should be evaluated and justified; this may include through exploratory plots and via the inclusion of a study effect variable on key model parameters [68, 78]. Looking at C-QTc relationships with parent compound and/or active metabolites might be required sometimes to characterize QTc effects [79].

## Efficacy endpoints

### Time to event analysis: survival

In analyzing survival data, two functions that are dependent on time are of particular interest: the survival and the hazard function. The survival function  $S(t)$  is the probability of surviving at least to time  $t$ . The hazard function  $h(t)$  is the instantaneous conditional probability of dying at time  $t$  having survived to that time. Survival curves can be estimated non-parametrically (i.e., Kaplan–Meier (KM) curves), semi-parametrically (i.e., Cox proportional hazard model), or parametrically (i.e., a lifetime model). The KM method estimates survival curves without the assumption of an underlying probability distribution in the presence of right-tailed censoring. Although easy to compute, it is hard to include covariate analysis, other than by grouping, in the models. Statistical significance between two survival curves can be made using a log-rank test, which tests the null hypothesis that there is no difference between the population survival curves (i.e., the probability of an event occurring at any time point is the same for the populations under comparison) against the alternative hypothesis that they are not the same.

Cox’s proportional hazards models are semi-parametric models making fewer assumptions than typical parametric methods. Cox models quantify how specific factors (covariates) influence the rate of a particular event happening at a given point in time. This rate is commonly referred to as the hazard rate. The hazard function can be written as a multiple linear regression of the logarithm of the hazard with the baseline hazard being the intercept term that varies with time:

$$h(t) = h_0(t) \cdot e^{(b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n)}$$

where  $t$  represents the survival time,  $h(t)$  is the hazard function determined by a set of covariates where the coefficients ( $b_1, b_2, \dots, b_n$ ) quantify the effect of covariates on the hazard,  $h_0$  is the baseline hazard. The quantities  $\exp(b_i)$  are called hazard ratios (HR). An  $HR > 1$  increases the hazard indicating a positive effect of the covariate with event probability and therefore negatively associated with the length of survival. An  $HR < 1$  reduces the hazard and increases the probability of survival. The main assumption of the Cox model is that the hazard is proportional: if an individual has a risk of death at some initial point that is twice as high as that of another individual, then at later times the risk of death remains twice as high, that is, the hazard ratio comparing two groups is constant over time. However, in oncology, the assumption of proportional hazard implicit into Cox’s models may not represent the ideal choice where the factor changing the hazard may vary with time.

Unlike the Cox regression model which does not specify the distribution function of the hazard, there are several parametric models such as Weibull, Gompertz, exponential, log-normal, and log-logistic models where the hazard function has to be specified. These parametric models allow us to estimate the effects of covariates on the hazard function, including variables that may change over time like drug concentration/dose, age, tumor growth dynamics, or time since surgical intervention.

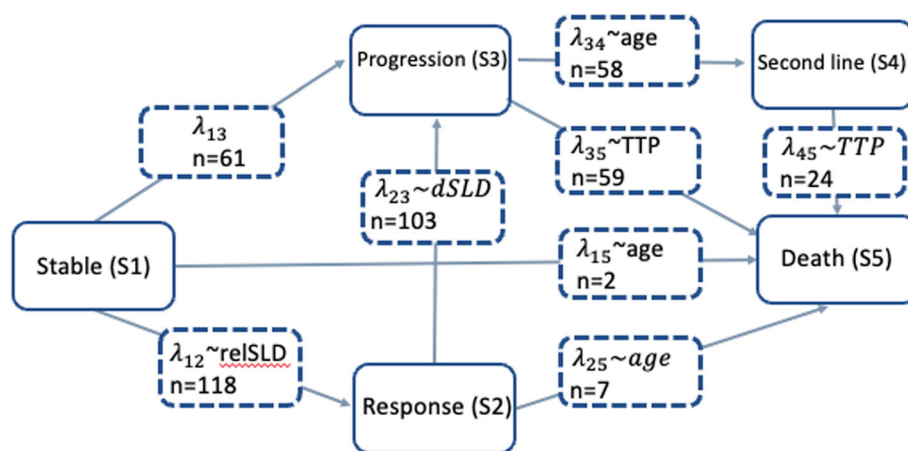
It is important to highlight the inherent selection bias and immortal time bias in survival metrics in oncology [10]. The TTE model approach with the estimation of a single survival function has its limitations. One alternative approach is the landmark method, which determines each patient's response at some fixed time point, with survival estimates calculated from that time point and associated statistical tests being conditional on patients' landmark responses excluding patients that die before the selected landmark timepoint [12]. Another approach to address the challenges of correctly describing the hazard function over time is the use of multistate models. Beyer et al. developed a multistate model with transition hazards estimated using semiparametric models [11]. Krishnan et al. proposed transition hazards using a parametric approach and mixture models [13]. Five states were considered according to Response Evaluation Criteria in Solid Tumors (RECIST) and using predictions derived from a longitudinal tumor growth inhibition model. The model did not allow to move back from one state to another when a response level has been achieved and maybe changed over time. Transition rates between states were estimated and defining the different hazard distributions (see Fig. 3).

In oncology, analysis of the association between drug exposure levels and late-phase clinical outcomes, such as OS and PFS (or event-free survival EFS) are often limited in scope for regulatory submissions. In light of the paucity of data and time constraints, industry practice and regulatory expectations to rationalize the dose rely on performing logistic or Cox proportional hazard regression analysis [20] of the registration-intent trial efficacy data. Even simpler exploratory approaches are sometimes considered to support the dose rationale, such as conducting a KM exploratory analysis of OS and/or PFS stratified by quartiles or tertiles of exposure levels [80, 81]. The general strategy is to demonstrate the absence of a relationship between exposure and the registrable endpoint within the range of exposure tested and infer from this that the dose is optimal since no substantial gain in efficacy can be achieved by increasing exposure levels further. Whether such analyses can be qualified as best practice is contentious as pragmatism is the main driver to rationalize the methodological framework used. Typically, the establishment of a recommended phase II dose is endorsed based on

combined analyses of emerging safety signals and surrogate efficacy endpoints with proven independent prognostic value for OS or PFS (e.g., tumor burden [82] or objective response [37, 83, 84] obtained from earlier trial data). Confirmatory trials rarely study more than one dose, and inherently suffer from underpowered statistical consideration for E–R characterization of OS and PFS. Furthermore, the ever-increasing pace for regulatory filing driven by commercial incentives and rapid access of novel treatment to patients in unmet needs further restrains the possibility to develop a data-driven E–R model of OS or PFS for regulatory submission given the time-consuming nature of these analyses. The downside of simple analyses lies in identifying spurious relationships due to imbalance of known prognostic factors of OS at baseline in strata of exposure [85, 86]. For example, ramucirumab E–R showed a positive association with OS in gastric cancer patients following a Cox regression [87, 88] with  $C_{\min,ss}$  or  $C_{\text{avg}}$ . However, the E–R model did not evaluate covariates such as C-reactive protein levels and tumor burden that are known prognostic factors of OS and could also impact PK as demonstrated in the case of tremelimumab [80], trastuzumab emtansine [81], and to a lesser extent with trastuzumab in the same indication [20]. The recommendation in such a case would be to first interpret with great caution any apparent trend of E–R emerging from simple analyses of trials evaluating a single-dose regimen. To explore further the risk of potential confounding in the event of a positive E–R finding, more sophisticated methods such as the parametric hazard model and multistate models should be attempted to quantify E–R in a multivariate framework. A full model or a selection of covariates of clinical relevance for the endpoint of interest and for the PK metric chosen should then be applied to quantify the relative contribution of covariates on the underlying E–R. As the last step, exposure should be removed from the model in case its causative link with OS or PFS assumed by the model is no longer supported by the data and can be fully explained by prognostic or predictive covariates remaining in the model. The best example of this good practice would be the already mentioned nivolumab E–R analysis in which baseline CL was included in the model [21].

Technical considerations, such as the choice of the PK metrics ( $C_{\max}$ , AUC) or when this metric is measured (first cycle vs steady-state), and data limitations (infrequent PK sampling, usually one dose tested in confirmatory trial) further constitutes challenges to a robust evaluation of E–R. Systemic exposure is consistently used in oncology as the driver of the E–R of OS and PFS while tumor penetration is often non-uniformed across solid tumors. Furthermore, some drugs, such as antibody–drug conjugates (ADC) or combination therapies, require understanding on which exposure level is assumed causally related (e.g.,





**Fig. 3** A diagram illustrating the multistate model describing the different states and transitions in the patient population studied.  $\lambda_{ij}$  represents the transition intensities between states,  $n$  for the transitions represents the number of observed transitions,  $n$  for the states is

warhead or total antibody exposure) and would require more complex analyses to disentangle the contribution of components. Combination therapy is another challenge from a modeling standpoint since data are generally limited. Monotherapy arms are seldomly available to fully characterize each single agent E–R relationship, and assumptions on the additivity or synergistic nature of the combinatory agents are left without data to infer their validity.

From a regulatory standpoint, simpler E–R approaches are customary to inform label claims justifying the dose as long as several conditions are fulfilled [86]. First, clinical data should demonstrate meaningful benefit-risk for patients for the registrable endpoint and a statistically compelling primary analysis of trial-level data at the given dose, with minimal dose reduction, delays, or omission after treatment initiation. Second, the regulatory strategy implies pre-specifying the model-based analysis plan to reinforce its confirmatory nature. If a 2-stage analysis is selected, individual post-hoc estimates from the popPK model will be incorporated as part of the analysis dataset and use to calculate the metrics of exposure used in the TTE analysis. Alternatively, a simultaneous modeling of PK and response could be carried out with longer running times, more intensive computational power requirements, and in many cases not a clear benefit in final model results unless the response is affecting the drug disposition and the proposed model is accounting for it. Ideally, data from Phase 2 trials with more than one dose level would be ideal to start model development and be prepared to streamline the critical path activities for filing regulatory dossiers. The reality is that due to the life-threatening condition of this therapeutic area and the possibility of expedited pathways for approval, phase 2 trials or expansions of Phase 1 trials

are often the basis for initial approvals. Therefore, there is usually no possibility of externally validating these analyses with independent (test) data sets. Thus, the analyst is generally not expected to perform external model validation.

are often the basis for initial approvals. Therefore, there is usually no possibility of externally validating these analyses with independent (test) data sets. Thus, the analyst is generally not expected to perform external model validation.

Fortunately, more complex models relating OS or PFS with systemic exposure levels, tumor growth dynamics and accounting for dropout, and dose reduction/delays or interruptions are developed once the regulatory submission timeline pressure unwinds [89]. These multivariate tools are far more valuable in their demonstrated track records of impacting the drug development strategy and post-marketing study designs. Leveraging the ability to integrate data from multiple studies and extrapolation/interpolation intrinsic properties [90], these models are used to bridge subcutaneous vs. intravenous dosing, convert flat vs. weight-based or BSA-based [30, 91], adults vs. pediatrics, ethnicity considerations, extend the dose interval [92], redefine therapeutic window for earlier line of therapies, inform patients treatment strategy [93], integrate historical data [94] and other advanced analytic framework [95].

### Tumor growth dynamics

RECIST is the current standard for determining how well a patient's tumor responds to treatment using assessments of growth/shrinkage captured in on-study X-rays, computerized tomography (CT), or magnetic resonance imaging (MRI) scans. RECIST is broadly accepted by oncology practitioners and regulatory bodies, and nearly all clinical trial treatment assessments for solid malignancies apply this framework. Still, standard RECIST methodology and its criteria for declaring treatment 'response' versus 'non-response' based on certain % of tumor shrinkage in the



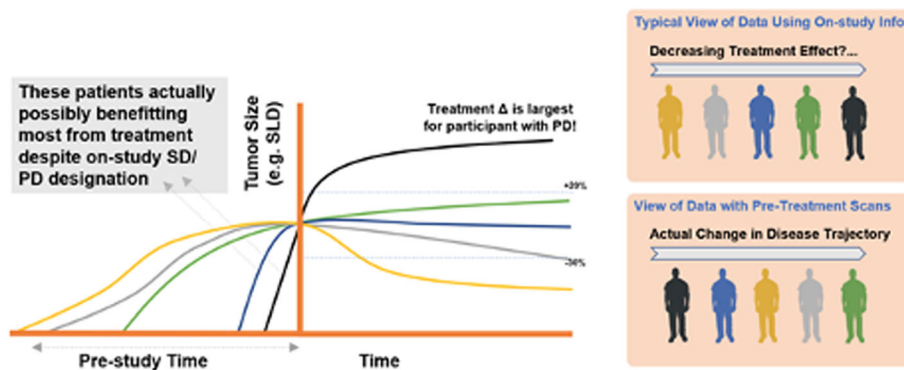
original reported tumor lesions, is often critiqued as inadequately representing overall disease burden [96], e.g. at times penalizing a deeper response with shorter time to a progression from nadir [97]. At the center of all RECIST-based assessments, including newer versions such as irRECIST [98], is a practice of categorization of data-rich longitudinal tumor size information into response strata of Progressive Disease (PD), Stable Disease (SD), Partial Response (PR), and Complete Response (CR). These categories are often further dichotomized into binary assignments of responders (PR + CR) versus non-responders (SD + PD) subgroups summarized at the population level as an ORR %. Hematologic malignancy studies use similar categorical response criteria to classify continuous assessments of disease burden, such as percent blasts in Acute Myeloid Leukemia (AML) [99], BCRABL/BCR ratio in Chronic Myeloid Leukemia (CML) [100], and M-protein in multiple myeloma (MM) [101]. In all cases, this categorization leads to loss of statistical power and is insensitive to both time dependencies and depth of response (or non-response) information captured in underlying time-course data. Indeed, if, as in MM, a discretized response spectrum has over six categories including “Very Good Partial Response (VGPR),” it may be a sign that the limits of discretization are being over-stretched to describe an underlying continuum! For this reason, longitudinal tumor burden modeling has become an increasingly applied tool for describing efficacy outcomes in clinical trials and relating tumor dynamics to predictive factors, including treatment dose/exposure [102].

Longitudinal modeling allows for a better understanding of the entirety of a patient’s tumor burden growth/shrinkage time-course to assess the possible impact of dose or schedule selection on disease response. Such analyses are attractive also because they permit derivation of simple secondary parameters to describe features of the profile (e.g. time to regrowth/nadir, depth of nadir, etc.) through interpolation—and sometimes extrapolation—of the observable data. Secondary parameters may be more intuitively linked with survival outcomes in TTE analyses, and thus play an important role in communicating modeling results with a clinical audience. Previously described tumor burden models have been employed in a wide array of drug development applications, and a significant number of these models successfully applied in late-stage development are simple, empirical models with a minimal combination of linear or exponential primary parameters describing growth (e.g., Kg) or shrinkage (e.g., Kd) of target tumor(s). Details on many of these kinetic tumor burden models have been previously published [82, 103–106] including several excellent review articles [107–109].

Central to the notion of longitudinal tumor burden modeling is the incorporation of multiple time-point

observations per patient to describe an overall disease trajectory. The type of data incorporated will inherently impact model fidelity/interpretability, and hence, the utility of its application. In oncology, where ethical/logistical considerations dictate the availability of tumor assessments before treatment initiation and after discontinuation, the influence of underlying data structure on tumor model inferences deserves particularly close attention [110]. Often, when a patient’s disease progresses due to tumor burden growth, the patient will discontinue study medication and contribute little or no more data beyond the treatment discontinuation date. Conversely, patients with responding disease tend to remain on study longer, thus contributing more and *longer* duration of scan data. Those patients with responsive disease, therefore, tend to receive more cumulative therapy and are more likely to experience safety-related dose modifications, which are common in oncology clinical trials due to the long-term systemic toxicities of many antineoplastic treatments. All these factors influence E–R interpretations and require careful consideration. The emphasis here, and for any astute longitudinal modeler, should always be on handling selective missingness of data, or informative censoring. In particular, this can problematically impact multiple aspects of E–R tumor burden modeling [110].

Informative censoring often contributes to issues of tumor model parameter identifiability in cases where very little on-study data was collected in one or more patient subgroups. In particular, the time-truncation of data from patients with rapidly progressing disease may limit the availability of data to describe accurate rates of tumor (re-)growth. As such, typical models with simple growth and decay terms to describe tumor kinetic profiles tend to be more empirical than mechanistic in nature, and any biological interpretation of a given parameter with regard to cancer cell replication, treatment resistance, and cell killing is often confounded. In non-linear mixed-effects models, this could manifest as high shrinkage of the estimated between-subject variability of one or more parameters. Parameter variability and parameter estimates could very well be appropriately estimated when the levels of shrinkage are high. However, graphical exploration of covariates using empirical Bayes estimates (EBEs) will not be able to guide covariate search. To mitigate on-study parameter identifiability issues, ideally, one would 1) acquire more data throughout the course of the trial, especially scans following disease progression, or 2) incorporate measurements of pre-treatment tumor growth into the tumor burden model [111] (Fig. 4). Such assessments could be an immensely valuable decision tool as they allow each trial participants’ pre-treatment trajectory to serve as internal control—assuming care is taken in assessing the same set of lesions as the ‘future’ RECIST



**Fig. 4** Deeper understanding of treatment effect and exposure–response patterns through tumor burden dynamics modeling: longitudinal modeling allows for robust, quantitative characterization of data-rich pre- and on-study tumor size information on the importance of understanding pre-baseline tumor size trajectory. This hypothetical example illustrates the concept of variability in the tumor growth trajectory at study start and consequent possibility to declare RECIST

target lesions. However, incorporating when “baseline tumor” was collected is relevant information as it is in general days to weeks before the treatment starts. In indications where PFS events are more often related to tumor burden growth (as opposed to survival), this approach then allows projection of control arm PFS using a single (active) arm study [112]. However, due to the nature of most sponsor-initiated studies and focus on the assessment of on-study treatment effects, obtaining and properly digitizing pre-study scans requires additional effort/resourcing and therefore has been rarely implemented.

Another common impact of informative censoring, particularly relevant to E–R applications of tumor burden modeling, involves potentially biased estimates of the E–R relationship when an aggregate exposure summary (i.e. steady-state or cumulative) is applied. A responding individual who remains on the study will necessarily have a higher aggregate/cumulative exposure even if there is no “true” E–R relationship simply because they have had more time on-study to accumulate exposure. A simple best practice in such analyses is to apply an instantaneous, time-varying, or early milestone/baseline exposure metric as the longitudinal model input. Careful consideration of data structure is still required in E–R analyses based on a steady-state exposure/dose intensity which may also be confounded as a result of time-varying dosing due to possible dose reductions and delays. Such analyses could actually imply an inverse exposure–efficacy relationship due to the commonly encountered correlation of time on-study, favorable treatment response, and corresponding increased likelihood of safety-related dose/exposure reductions from the longer duration of exposure. Hence, a general awareness of some of the key aspects of typical oncology trial conduct and these multiple potential

progressive disease in a patient likely benefiting more from treatment (e.g. black) than another who would be classified with a more favorable stable disease designation (e.g. green). More precise estimates of treatment effect and dose/exposure–response relationships can be brought through a quantitative understanding of a patient’s entire tumor size trajectory, including available pre-baseline scans

confounding phenomena is key in both study design and data analysis to avoid possible misinterpretation of spurious E–R relationships. Similar considerations of informative censoring should also factor into the use of on-study (post-baseline) covariates. Zhen et al. [113] model survival data and longitudinal changes in target lesions accounting for correlation between dropout and response. In this trial with locally advanced or metastatic urothelial carcinoma (UC) patients treated with durvalumab, at the time of the data cut, 100 of 186 subjects had dropped out of the study (67 of them were due to death). As is typical of oncology trials, the risk of patient dropout was strongly influenced by treatment response. Patients with rapid tumor progression dropped out early, whereas those whose disease improved had longer follow-up times. Similarly, Hansson et al. [28] and Schindler et al. [114] incorporated a dropout model enabling prospective simulations of tumor response over time as dropouts were not considered at random.

Tumor burden model structure selection and verification for a particular application should be driven by multiple considerations, including general goodness-of-fit, parsimony considerations, and whether the model is able to adequately describe the data for pre-stated objectives. Prior knowledge on the kinetics of disease burden for a given treatment modality may also factor into model selection. For example, cytotoxic chemotherapy treatments, which differ from immunomodulating treatments in their mechanisms of action, can be expected to show unique kinetics of disease response and progression [115–117]. In general, immunotherapy efficacy has been associated with delayed but durable responses that contrast with the more rapid but transient responses seen with cytotoxic agents. Studies have also shown initial tumor “pseudoprogression” followed by delayed response in some patients treated with

immunotherapy [98, 118]. The simpler two or three-parameter empirical tumor models derived primarily from clinical experience with chemotherapy agents often will not adequately capture this type of response pattern. Previous publications describing mixture models have been used to account for and categorize patients with hyper progressive disease as well as those with a delayed, durable response [119, 120].

When choosing a model, careful evaluation of extrapolation bias is recommended given that many existing tumor size models include unbounded exponential growth terms that fail to adequately extrapolate without significant prediction bias [121]. It is therefore critical to examine model performance in the extrapolation setting and to investigate the relationship between follow-up duration and extrapolation bias. Simulation bias tests should generally be performed when evaluating base model structures by first estimating models with time-truncated data, and then assessing the general ability of the model to generate unbiased predictions of ‘future’ data. When assessing model GOF diagnostics, informative censoring and extrapolation bias may also impact interpretation. Visual inspection of trends in conditional weighted residuals versus population predictions, for example, will tend to obscure model misspecification or biased prediction of observations from participants with rapidly growing tumors, which comprise a smaller proportion of the total data set than data from participants with shrinking tumors. VPCs may therefore be a valuable tool for model verification, but again, may also be impacted by extrapolation bias since exponential growth in the post-discontinuation phase can drive anomalous prediction intervals. It is crucial that VPCs also include censoring rules or a drop-out model which approximates clinical practice implemented in the corresponding trial protocols, e.g. truncation of simulated tumor size profiles after 20% growth from nadir—a typical rule governing RECIST progression of target lesions in solid tumor indications.

Advances in radiomics, which applies informatics, machine learning, and other big-data approaches to imaging data, have led to a growth in the number and types of features that may be captured for tumor burden modeling [122]. With appropriate application, this information holds great potential to enhance the clinical relevance of inferences drawn from tumor burden modeling. In alignment with the conventions of RECIST, the majority of published tumor burden modeling studies have been conducted using the summary metric “sum of longest diameters” (SLD) from radiologist-selected target lesions. However, several reports have indicated that volumetric data may be more informative [103]. Hierarchical modeling of individual lesion dynamics with between-tumor variability may yield even deeper insights into disease heterogeneity, as the

homogenizing effect of combining lesion information from different anatomic sites for SLD inherently reduces the information available to the modeler.

A key shortcoming of nearly all of the aforementioned tumor burden models, including individual lesion models, is that they rely on information from only pre-specified target lesions—which may or may not be adequately indicative of an overall disease burden. Per RECIST guidelines, individual target lesions are chosen as representative of a patient’s tumor burden for monitoring a relative treatment effect, but a more precise understanding of the tumor size to survival relationships may be established by accounting for the entire tumor/metabolic disease burden [123]. Irrespective of whether unidimensional, bidimensional, or volumetric radiographic assessments are employed, the clinical appropriateness of any given tumor size descriptor should be re-evaluated in different treatment settings and tumor types. While construction of quantitative models linking longitudinal tumor burden with instantaneous survival risk is a relatively recent endeavor, it is already clear that these relationships may vary by nature of different cancer types, anatomical locations of the lesions, and potentially by treatment modality. Hence, the choice of appropriate tumor burden descriptors is likely to be case-dependent and may involve either one or more derived primary or secondary tumor parameters.

### Special considerations in modeling hematologic malignancies

With the exception of lesion-level modeling, all recommendations in the above sections can be applied to hematologic malignancies, where total target tumor size is replaced by the appropriate continuous tumor burden metric for that particular malignancy: M-protein levels in secretory multiple myeloma patients, percent blasts in AML, and BCRABL/ABL ratio in CML. In the latter two, care must be taken to correctly transform the raw disease burden to a bounded assessment value based on the nature of the measurement. For example, in CML, assuming that mRNA levels are proportional to the number of genes in a cell, the BCRABL/ABL ratio can be represented as the ratio of a number of malignant cells to the weighted sum of normal cells (which have two copies of ABL) and malignant cells (which have one copy of ABL) [124].

Modeling lymphoma data may present additional complexities as response assessments are based on both lesion size (sum of products of diameters, SPD) and metabolic activity (FDG-PET avidity) [125]. If raw scans are available, assessment of metabolic tumor volume (MTV) [126], which is the total number of FDG-PET avid voxels in the scanned region of the patient’s body, are preferable to the dichotomized criteria or sum of products of diameters

alone, which does not consider whether the nodes in question are actually metabolically inactive (dead).

An additional feature of many hematologic malignancies is the concept of minimal residual disease (MRD) [127], which typically refers to technology with higher sensitivity to low disease burden than the conventional metrics. For example, 6-color flow cytometry in multiple myeloma can detect down to 0.01% levels of myeloma-transformed plasma cells in the bone marrow, as opposed to M-protein levels in peripheral blood which reach the lower limit of detection of 0.1 g/dL at underlying disease burdens ranging from 0.001 to 1%. Six-color MRD, which in the case of MM has been shown to predict incremental survival benefit with every tenfold decrease in MRD [128], are nonetheless dichotomized into MRD positive or negative categories, despite that these definitions may change yearly as more sensitive assays are developed [129]. For this reason and many aforementioned benefits above, we recommend fitting the tumor burden dynamic model to both the conventional continuous metric (e.g. M-protein levels in g/dL) and the continuous MRD metric (e.g. number of cancer cells per ml of sample) simultaneously, which provides greater identifiability particularly when M-protein is below the limit of quantitation (BLQ), which is often the only time MRD is assessed [127].

## Tumor biomarker and disease progression

Disease progression modeling is utilized to describe the time course of disease status and track disease severity over time. These models usually incorporate biomarker data and clinical outcomes to characterize natural disease progression [130].

In prostate cancer (PCa), prostate-specific antigen (PSA) has been recognized as a biomarker for diagnosis, prognosis, and monitoring of disease activity [131]. PCa is usually characterized as either low-risk non-aggressive (indolent) or high-risk aggressive tumors. While indolent PCa is benign prostatic hyperplasia (BPH) in general, aggressive PCa may lead to cancer-specific morbidity [132]. Identification of the most aggressive PCa cases among all patients diagnosed with PCa could help in the selection of the patients who might benefit from radical therapy. de Charry et al. developed a semi-mechanistic model of PSA longitudinal growth to help differentiate aggressive and indolent PCa at diagnosis [133]. The individual preoperative PSA data from patients with PCa and those with benign prostatic hyperplasia were analyzed using a population kinetic approach and a semi-mechanistic nonlinear mixed-effects model [133]. This analysis demonstrated a greater PSA increase rate by cancer cells than by non-cancer cells, while PSA production rate was

greater by benign tissue than by malignant tissue. A significant relationship between the PSA production rate by cancer cells and the probability of D'Amico high-risk group was also identified with logistic regression. Moreover, multivariate tests demonstrated that the PSA production rate by cancer cells, Gleason score, and positive surgical margin status were all significant independent predictive factors regarding relapse-free survival (RFS). This semi-mechanistic model provided a possible means to determine whether a patient is likely to have aggressive PCa before surgery.

In addition to PSA, the count of circulating tumor cells (CTCs) has also emerged as a promising surrogate marker in patients with metastatic castration-resistant prostate cancer (mCRPC) [134]. In 2004, the FDA approved the use of the CellSearch® system for detecting CTCs in cancer patients. It is to date the only approved laboratory test for CTCs and is being used to enumerate the number of CTCs of epithelial origin in a 7.5 mL blood sample [135]. Wilbaux et al. developed a semi-mechanistic model to quantify the dynamic relationships between the kinetics of CTC counts and PSA concentrations during treatment in patients with mCRPC [136]. Their joint model incorporated drug effect kinetics for chemotherapy and hormonal therapy through two different K-PD compartments as no drug concentrations data were available. The treatment effects on both PSA and CTCs were assumed to be mediated through a common latent variable that was interpreted as tumor burden. PSA kinetics were described by an indirect response model, while the CTC kinetics in the total body blood was modeled by a cell lifespan model assuming that the rate of cell loss was equal to the rate of cell production delayed by the lifespan. The dynamic change of CTC counts was considered as a random sampling from a negative binomial distribution. By simulating the kinetics of PSA, CTC counts, and the tumor burden, CTC counts turned out to be more sensitive to the variation of the tumor burden. This model was the first to quantify the dynamic links between the kinetics of PSA and CTC counts during treatment in patients with mCRPC. Although limitations exist, this model demonstrated the potential of using CTC counts as a predictor of treatment response or disease progression in patients with mCRPC.

## Characterization of clinical consequences of product immunogenicity

In recent years, biological products, such as monoclonal antibodies and other therapeutic proteins, have been widely used for the treatment of various cancers. Occasionally, the administered biological product may provoke an immune response (known as immunogenicity) in some patients



receiving repeated dosing which may lead to excessive cytokine release and/or formation of anti-drug antibodies (ADAs) [137, 138]. Some unwanted adverse events, including anaphylaxis, cytokine release syndrome, infusion reaction, and other non-acute reactions, may occur as the result of the immune response. The cross-reaction between ADAs and their endogenous counterparts may interfere with certain physiological processes, causing additional safety concerns. Furthermore, the immunological reaction may compromise the drug effect. For example, the presence of ADAs usually affects the drug clearance and decreases drug exposure. Additionally, neutralizing ADAs may interfere with the interaction between the therapeutic protein and its target. Therefore, using modeling approaches to quantitatively characterize the clinical consequences of product immunogenicity may provide insights into the risk and benefit profiles in patient subgroups (e.g., ADA positive and negative) and guide the optimal use of the product [139].

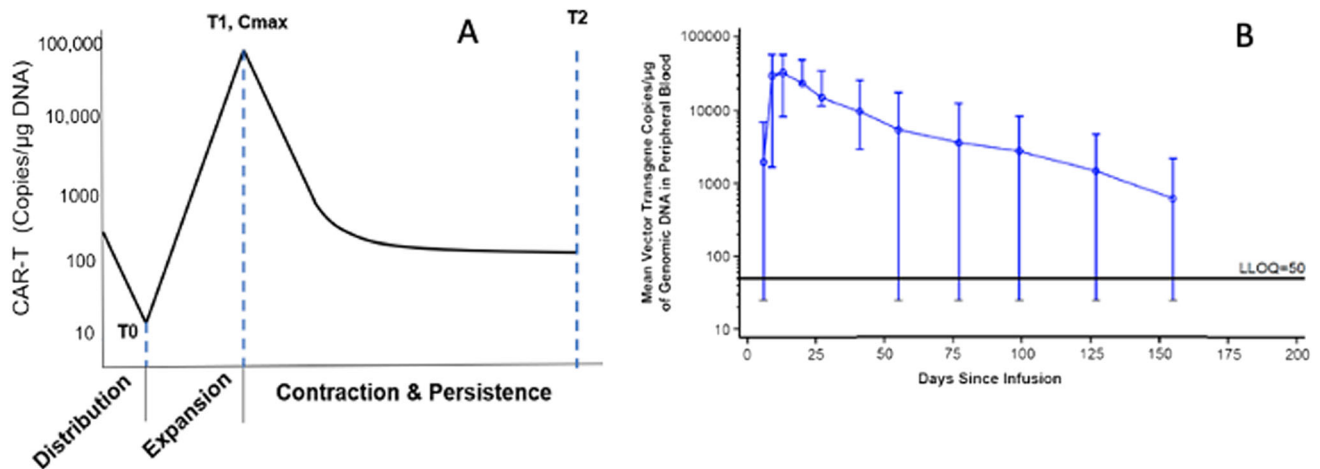
Appropriate quantification of the clinical impact of immunogenicity relies on the accurate acquisition of the data. ADAs titers in serum/plasma have been considered as the major biomarker to track patients receiving treatment with a biological product. Time to the first appearance of ADA (i.e., onset), sustained duration, and level of ADA (i.e., duration) may vary among individuals following chronic treatment. Unfortunately, the ADA level cannot be monitored on a continuous basis. The sampling schedule, therefore, becomes a critical component to characterize the potential ADA changes and to link the ADA changes to clinical consequences. A sensitive, specific, and selective bioassay for ADA is another key factor to ensure data quality. In general, ADA is detected by immunoassay using an ADA-antibody, which may interfere with the coexisting therapeutic protein and endogenous substances in biological matrices. An assay's drug tolerance, which measures the assay sensitivity in the presence of the therapeutic proteins, as compared to typical concentration levels in patients is important to understand the reliability of the detected ADA [140]. The Global Bioanalysis Consortium (GBC) set up an international team to explore the impact of immunogenicity on PK assessments. The result of the work was a white paper where they presented strategies to assess if changes in drug concentration are due to ADA-mediated changes in clearance or instead a consequence of ADA interference with the bioanalytical assay [141]. There is a multitude of factors that could influence the immunogenicity of biologics. These factors could be classified into disease-, patient-, or product-related factors [142–145]. Immunogenicity as a consequence of a biologic therapeutic administration may result in multiple polyclonal antibodies against multiple epitopes circulating in serum. Each ADA species has its own specificity and binding affinity. As a

result, these different treatment-emergent ADA responses may have different effects on the PK/PD of the drug therapeutic: a neutralizing effect on activity by interfering with the drug's ability to bind to its pharmacologic target, a non-neutralizing effect on activity paired with an effect on the PK, or a non-neutralizing effect on activity with an enhanced elimination of the drug therapeutic [145].

The findings on clinical consequences of immunogenicity for a biological product may be misleading if an inappropriate modeling approach is used. Randomized, well-controlled clinical trials are routinely used to characterize the efficacy and safety profiles of a biological product before it gains marketing authorization. An analysis, for instance, may be conducted to compare overall survival in patients with (i.e., ADA positive group) or without ADA (i.e., ADA negative group) detected anytime during the treatment. However, this analysis may lead to misleading outcomes. The time to the first occurrence of ADA and duration for detectable ADA sustained in plasma appears to be random among individuals during the treatment phase. It should be noted that some patients in the ADA-positive group are the patients with the late-occurring ADA formation, who must already survive long enough with continuous treatment. Thus, the true ADA effect on overall survival may be attenuated (i.e., biased), if the overall survival is directly compared between the two groups. Rather, a landmark analysis comparing patients with or without early (e.g., within 4 weeks after the treatment is initiated) detectable ADA may provide a relatively better angle to characterize the potential impact of ADA formation on drug effect, if adequate data are available. The analysis performed for atezolizumab provides a good example. As demonstrated in OAK study in locally advanced and metastatic non-small cell lung cancer patients, 21% of patients were tested positive for ADA by week 4. The analysis suggested that the presence of ADA early after the treatment initiation (i.e., 4 weeks) may significantly affect the overall survival in this patient population [146].

However, careful consideration must be given to any post-randomization variable, including ADA formation. Kong et al. studied the potential impact of ADA formation on long term benefit in a randomized controlled trial with atezolizumab, in which ADA was not observed in the control arm [147]. ADA status was only observable in atezolizumab-treated patients, and variables that are a consequence of the treatment either preclude observation or affect the interpretation of the clinical endpoint of interest. The authors propose a weighted approach for estimating effects based on Weight Placebo Patients (WPP) approach [148] where the post-baseline strata are observed for every subject. The idea behind this is if ADA can be organized in categories, is the treatment benefit clinically





**Fig. 5** **A** Kinetic Phases of CAR-T therapies. Adapted from Liu et al. [153]. **B** CAR-T expansion and persistence in peripheral blood adapted from Berdeja et al. [154]

meaningful for all categories of ADA? and is the treatment effect similar between certain or all categories of ADA?. Atezolizumab treated ADA-positive patients showed worse OS relative to ADA-negative patients. Furthermore, ADA-positive patients also showed differences in several prognostic baseline variables compared to ADA-negative patients. The authors concluded after correcting the data for the prognostic baseline variables that hazard ratios were similar for ADA-positive and ADA-negative patient populations.

## Modeling in cell therapies

Cell and Gene Therapy (CGT) have now entered a new area of treating diseases. Although scientific efforts have been in progress for over a century, recent advancements in the last decade have significantly shifted the treatment paradigm with several regulatory approvals [149]. As of February 2023, there were 27 CGT products that were approved by FDA [149, 150]. The proven hypothesis, that either alteration of certain cells (Cell therapy) or certain genes into cells (Gene therapy) may lead to treatment, has made CGT very effective in individualized treatment. CGT is formulated differently from traditional large batch production of a small or large molecule (antibody) drug, and in many instances, CGT is designed specifically for individual patients (for example, autologous cell therapy).

Among several CGT modalities, Chimeric Antigen Receptor (CAR) T cell-based therapy has shown clinical benefits in different cancer types [151, 152]. In this treatment, a patient's T-cells are modified in the laboratory and specific CARs are introduced to the surface of the T-cell in order to target very specific malignant cancer cells. For the purpose of this review, we will focus on current

advancements in the clinical pharmacology of autologous CAR-T therapy only.

The PK of CGT drugs is characterized using quantitative polymerase chain reaction (qPCR) assay for viral products and using flow cytometry or qPCR for autologous cell therapies. A typical small or large molecule drug after IV administration exhibits maximum drug concentration ( $C_{max}$ ) at the end of infusion, whereas for CAR-T cells, the  $C_{max}$  is generally achieved after a few days of IV administration due to cellular proliferation and expansion upon interaction with the target antigens in circulation (or at the site of action in tumor or bone marrow). The apparent kinetics of CAR-T may consist of four phases including distribution, expansion, contraction, and persistence [153], see Fig. 5A and B.

In humans, the expansion/proliferation of CAR-T cells is found to be highly variable and primarily driven by an individual's immune system activity. Cilta-cel (Carvykti<sup>TM</sup>) CAR transgene concentrations showed maximum peripheral expansion at a median of 12.7 days (range 8.7–54.6) with observed persistence lasting over > 100 days in peripheral blood [154]. Among patients with 6 months follow-up, most had cilta-cel CAR transgene concentration below the level of quantification (< 50 CAR gene copies per  $\mu\text{g}$  DNA) in peripheral blood. Several factors, such as prior lines of therapy, tumor burden, and disease status, may impact an individual's immune system activity. In addition, CAR-T cell drug product characteristics, such as CD4:CD8 ratio, %CAR + T cells, the constitution of effector, and memory cells [155] may also contribute to cellular expansion/proliferation, resulting in highly variable exposure parameters such as  $C_{max}$  and area under the curve (AUC) with inter-individual variability as large as 165–190% [156].

Since expansion and persistence of CAR-T cells are dependent on the cellular composition of the individual donor's original T cells and the individual's immune system, it is difficult to extrapolate or predict the extent of inter-individual PK exposure (cellular expansion/persistence) in humans. In addition, given the limitation of pre-clinical models in mimicking the complexity of the human immune system, allometric-based principles cannot be applied in translating CAR-T exposure from preclinical species to humans. There have been recent attempts to develop translational models [157], but the lack of relevant preclinical species limits such preclinical to clinical predictions. Current regulatory guidance [158] on gene and cell therapy also states that unlike for small and large molecules, allometric-based scaling cannot be applied to predict starting dose in humans. Typical CAR-T starting dose selection has been based on prior knowledge of cell therapy and target expression.

Since CAR-T exposure exhibits high inter-individual variability, depending on the range of dose levels studied, and the number of subjects at each dose level, dose dependence on PK parameters and impact on response may be difficult to identify. A population cellular kinetic model was developed using clinical data of Kymriah<sup>TM</sup> [159] where the dose range administered ranged from 0.2 to 5.4 Mcells/kg. The authors did not observe any relationship between dose and exposure. Even though an increase in  $C_{max}$  was found to increase the incidence and severity of key safety outcomes (cytokine release syndrome, CRS), due to lack of a clear dose–exposure relationship, such correlations could not be used to identify an optimal dosing regimen. In another study using Abecma<sup>®</sup> [155], the authors proposed a dose-dependent increase in expansion/persistence based on a prospectively designed Phase 1/2 study where dose escalation was conducted in cohorts at dose levels of 50, 150, 450, and 800 Mcells. Despite a dose-dependent increase in exposure, a plateau in exposure was apparent at 450–800 Mcells. Abecma<sup>®</sup> transgene levels were positively associated with objective tumor response (partial response or better), where responders achieved  $\sim 4.6$ -fold higher  $C_{max}$  and  $\sim 5.6$ -fold higher  $AUC_{(0-28 \text{ days})}$  in comparison to non-responders [156]. It is unclear if high inter-individual variability with overlapping exposures and lack of a clear dose–exposure relationship influenced the approved dose range of 300–460 Mcells for Abecma<sup>®</sup>.

In summary, drug exposure after CAR-T therapy has shown high inter-individual variability with overlapping exposures between dose levels. Although higher exposure may be associated with higher responses, the lack of a clear dose–exposure relationship limits the determination of optimal dose levels. For future trials, a dose-escalation study should be considered in early development with rich

PK collection throughout the study to inform better characterization of the dose/exposure–response relationship.

## General considerations

Adequate care needs to be taken in performing E–R analyses using data collected from late-phase oncology clinical trials. The majority of Phase 2 and 3 oncology clinical trials are carried out with one dose level, having minimal E–R information available from the dose-escalation part of the Phase 1 study, often in refractory patients where at the most we can get an idea of E–R for safety endpoints, but we don't expect any efficacy as the target population is normally not the intended population for the projected filing indication. With only a single dose, the range of exposures is often limited, and may make detection of E–R relationships difficult. FDA has long been pushing for a more thoughtful approach to dosing. With the recent release of a draft guidance requesting sponsors to study a range of doses in early development, this may make E–R modeling more sensitive at detecting such relationships.

Performing a comprehensive pharmacometrics analysis of a drug product at the time of NDA/BLA filing is often challenging but probably the most cost-efficient approach when considering subsequent filings in subsequent indications and special populations. We have seen repeatedly with many therapies that disease burden as characterized by multiple risk factors, such as Eastern Cooperative Oncology Group (ECOG) performance score (PS), percentage of prior surgery, and increased number of metastatic sites, may vary significantly among enrolled patients at baseline and usually changes in the course of the treatment. In late-stage cancer, patients with a high disease burden may be associated with increased clearance for some compounds. Usually, these patients tend to show more aggressive disease progression and shorter survival times. Thus, the disease burden becomes a confounder that affects both drug exposure and survival. All these factors in combination may distort the underlying causal relationship associated with the treatment effect. Inappropriate modeling approaches are more likely to happen when very little information has been collected during early-stage development or no input has been requested from the pharmacometrician, which could lead to an erroneous presentation of the underlying E–R relationship when the pivotal trial reads out.

Often pharmacometricians play “catch-up” by having limited time to perform the retrospective analyses required to comply with regulatory agencies' recommendations with very little influence on requesting adequately powered dose-finding studies during the drug development program. Usually, the modeling comes at the end, when perhaps the

necessary assessments were not collected or incompletely collected, and the information available is insufficient to perform an in-depth analysis of the safety and efficacy endpoints and their correlations with exposure. The pharmacometrics community has the tools, the knowledge, and well-described models summarized in this review that can help characterize E–R in drug development and hence better understand the drug behavior and therapy management. However, tools, knowledge, and models are only half of the work, well-designed trials, and informative data collections are the other half and it is very challenging to do both parts in isolation. A collaborative interaction is required among clinicians, statisticians, pharmacologists, pharmacometricians, and decision-makers to succeed in prospectively defining a clinical development strategy tailored towards the identification of the optimal dose regimen and benefit/risk ratio of a given drug product for the intended-to-treat population of patients.

**Author contributions** All authors wrote one or multiple sections of the manuscript and all authors reviewed the totality of the manuscript.

## Declarations

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Minchella K, Xu H, Al-Huniti N (2016) Exposure-response methods and dose approval of new oncology drugs by FDA from 2005 to 2015. *J Clin Orthod* 34:2530–2530
- Guidance for industry: exposure-response relationships-study design, data analysis, and regulatory applications. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER); (2003). <https://ci.nii.ac.jp/naid/10016901109/>
- Bello CL, LaBadie RR, Ni G, Boutros T, McCormick C, Ndongo MN (2012) The effect of dacomitinib (PF-00299804) on CYP2D6 activity in healthy volunteers who are extensive or intermediate metabolizers. *Cancer Chemother Pharmacol* 69:991–997
- U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). Guidance for Industry Exposure-Response Relationships—Study Design, Data Analysis, and Regulatory Applications. FDA. Accessed 10/9/2020 April 2003. <https://www.fda.gov/media/71277/download>
- Bullock JM, Lin T, Bilic S (2017) Clinical pharmacology tools and evaluations to facilitate comprehensive dose finding in oncology: a continuous risk-benefit approach. *J Clin Pharmacol*. <https://doi.org/10.1002/jcph.908>
- Bi Y, Liu J, Furmanski B, Zhao H, Yu J, Osgood C et al (2019) Model-informed drug development approach supporting approval of the 4-week (Q4W) dosing schedule for nivolumab (Opdivo) across multiple indications: a regulatory perspective. *Ann Oncol* 30(4):644–651. <https://doi.org/10.1093/annonc/mdz037>
- Office of the Commissioner. Project Optimus. In: U.S. Food and Drug Administration. 18 Feb 2022. Accessed 28 Feb 2022. <https://www.fda.gov/about-fda/oncology-center-excellence/project-optimus>
- Houk BE, Bello CL, Poland B, Rosen LS, Demetri GD, Motzer RJ (2010) Relationship between exposure to sunitinib and efficacy and tolerability endpoints in patients with cancer: results of a pharmacokinetic/pharmacodynamic meta-analysis. *Cancer Chemother Pharmacol* 66:357–371
- Di Paolo A, Bracarda S, Arrigoni E, Danesi R (2017) Sunitinib in metastatic renal cell carcinoma: the pharmacological basis of the alternative 2/1 schedule. *Front Pharmacol* 8:523
- Anderson JR, Cain KC, Gelber RD (2008) Analysis of survival by tumor response and other comparisons of time-to-event by outcome variables. *J Clin Oncol* 26(24):3913–3915. <https://doi.org/10.1200/jco.2008.16.1000>
- Beyer U, Dejardin D, Meller M, Rufibach K, Burger HU (2018) A multistate model for early decision making in oncology. *arXiv [stat.AP]*. <https://doi.org/10.1002/bimj.201800250>
- Lara PN Jr, Redman MW, Kelly K, Edelman MJ, Williamson SK, Crowley JJ et al (2008) Disease control rate at 8 weeks predicts clinical benefit in advanced non-small-cell lung cancer: results from Southwest Oncology Group randomized trials. *J Clin Oncol* 26:463–467
- Krishnan SM, Friberg LE, Bruno R, Beyer U, Jin JY, Karlsson MO (2021) Multistate model for pharmacometric analyses of overall survival in HER2-negative breast cancer patients treated with docetaxel. *CPT Pharmacometrics Syst Pharmacol* 10:1255–1266
- Modi S, Saura C, Yamashita T, Park YH, Kim S-B, Tamura K et al (2020) Trastuzumab deruxtecan in previously treated HER2-positive breast cancer. *N Engl J Med* 382:610–621
- Uy GL, Aldoss I, Foster MC, Sayre PH, Wieduwilt MJ, Advani AS et al (2021) Flotetuzumab as salvage immunotherapy for refractory acute myeloid leukemia. *Blood* 137:751–762
- Liu C (2016) Dose Adjustment Integrated Exposure-Response Analysis (DAIER) for dose optimization. Accessed 13 June 2016. <https://www.aacr.org/wp-content/uploads/2019/11/6.13.16-FDA-AACR-Dose-Finding-for-Online.pdf>
- Leven C, Padelli M, Carré J-L, Bellissant E, Misery L (2019) Immune checkpoint inhibitors in melanoma: a review of pharmacokinetics and exposure-response relationships. *Clin Pharmacokinet* 58:1393–1405
- Liu C, Yu J, Li H, Liu J, Xu Y, Song P et al (2017) Association of time-varying clearance of nivolumab with disease dynamics and its implications on exposure response analysis. *Clin Pharmacol Ther* 101:657–666
- Kawakatsu S, Bruno R, Kågedal M, Li C, Girish S, Joshi A et al (2021) Confounding factors in exposure-response analyses and mitigation strategies for monoclonal antibodies in oncology. *Br J Clin Pharmacol* 87:2493–2501

20. Feng Y, Wang X, Bajaj G, Agrawal S, Bello A, Lestini B et al (2017) Nivolumab exposure-response analyses of efficacy and safety in previously treated squamous or nonsquamous non-small cell lung cancer. *Clin Cancer Res* 23:5394–5405
21. Wang X, Feng Y, Bajaj G, Gupta M, Agrawal S, Yang A et al (2017) Quantitative characterization of the exposure-response relationship for cancer immunotherapy: a case study of nivolumab in patients with advanced melanoma. *CPT Pharmacometrics Syst Pharmacol* 6:40–48
22. Wang R, Shao X, Zheng J, Saci A, Qian X, Pak I et al (2020) A Machine-learning approach to identify a prognostic cytokine signature that is associated with nivolumab clearance in patients with advanced melanoma. *Clin Pharmacol Ther* 107:978–987
23. Wang R, Zheng J, Shao X, Ishii Y, Roy A, Bello A et al (2019) Development of a prognostic composite cytokine signature based on the correlation with nivolumab clearance: translational PK/PD analysis in patients with renal cell carcinoma. *J Immunother Cancer* 7:348
24. Dai HI, Vugmeyster Y, Mangal N (2020) Characterizing exposure-response relationship for therapeutic monoclonal antibodies in immuno-oncology and beyond: challenges, perspectives, and prospects. *Clin Pharmacol Ther* 108:1156–1170
25. U.S. Package Insert of Herceptin®. Accessed 10 Sept 2020. [https://www.accessdata.fda.gov/drugsatfda\\_docs/label/2018/103792s53451bl.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/label/2018/103792s53451bl.pdf)
26. Yang J, Zhao H, Garnett C, Rahman A, Gobburu JV, Pierce W et al (2013) The combination of exposure-response and case-control analyses in regulatory decision making. *J Clin Pharmacol* 53(2):160–166. <https://doi.org/10.1177/0091270012445206>
27. Braun TM (2002) The bivariate continual reassessment method. Extending the CRM to phase I trials of two competing outcomes. *Control Clin Trials* 23:240–256
28. Hansson EK, Amantea MA, Westwood P, Milligan PA, Houk BE, French J et al (2013) PKPD modeling of VEGF, sVEGFR-2, sVEGFR-3, and sKIT as predictors of tumor dynamics and overall survival following sunitinib treatment in GIST. *CPT Pharmacometrics Syst Pharmacol* 2:e84
29. Zhao X, Suryawanshi S, Hruska M, Feng Y, Wang X, Shen J et al (2017) Assessment of nivolumab benefit-risk profile of a 240-mg flat dose relative to a 3-mg/kg dosing regimen in patients with advanced tumors. *Ann Oncol* 28:2002–2008
30. Bei D, Osawa M, Uemura S, Ohno T, Gobburu J, Roy A et al (2020) Benefit-risk assessment of nivolumab 240 mg flat dose relative to 3 mg/kg Q2W regimen in Japanese patients with advanced cancers. *Cancer Sci* 111(2):528–535. <https://doi.org/10.1111/cas.14252>
31. Feng Y, Roy A, Masson E, Chen T-T, Humphrey R, Weber JS (2013) Exposure-response relationships of the efficacy and safety of ipilimumab in patients with advanced melanoma. *Clin Cancer Res* 19:3977–3986
32. Zhao X, Shen J, Ivaturi V, Gopalakrishnan M, Feng Y, Schmidt BJ et al (2020) Model-based evaluation of the efficacy and safety of nivolumab once every 4 weeks across multiple tumor types. *Ann Oncol* 31:302–309
33. U.S. Food & Drug Administration (2018) FDA approves nivolumab plus ipilimumab combination for intermediate or poor-risk advanced renal cell carcinoma. *Case Med Res*. <https://doi.org/10.31525/fda1-ucm604685.htm>
34. Chuk MK, Chang JT, Theoret MR, Sampene E, He K, Weis SL et al (2017) FDA approval summary: accelerated approval of pembrolizumab for second-line treatment of metastatic melanoma. *Clin Cancer Res* 23:5666–5670
35. Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD (2000) Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med* 19:1059–1079
36. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR (1996) A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 49:1373–1379
37. Agrawal S, Feng Y, Roy A, Kollia G, Lestini B (2015) Nivolumab dose selection: challenges, opportunities and lessons learned for cancer immunotherapy. *J Immunother Cancer*. <https://doi.org/10.1186/2051-1426-3-s2-p141>
38. Lacroix BD, Lovern MR, Stockis A, Sargentini-Maier ML, Karlsson MO, Friberg LE (2009) A pharmacodynamic Markov mixed-effects model for determining the effect of exposure to certolizumab pegol on the ACR20 score in patients with rheumatoid arthritis. *Clin Pharmacol Ther* 86:387–395
39. Zingmark P-H, Kågedal M, Karlsson MO (2005) Modelling a spontaneously reported side effect by use of a Markov mixed-effects model. *J Pharmacokinetic Pharmacodyn* 32:261–281
40. Karlsson M (2000) A pharmacodynamic Markov mixed-effect model for the effect of temazepam on sleep. *Clin Pharmacol Ther* 68(2):175–188. <https://doi.org/10.1067/mcp.2000.108669>
41. Hénin E, You B, VanCutsem E, Hoff PM, Cassidy J, Twelves C et al (2009) A dynamic model of hand-and-foot syndrome in patients receiving capecitabine. *Clin Pharmacol Ther* 85(4):418–425. <https://doi.org/10.1038/clpt.2008.220>
42. Markov AA, Nagorny NM (1988) The theory of algorithms. <https://doi.org/10.1007/978-94-017-3477-6>
43. Markov AA (1954) The theory of algorithms. *Trudy Matematicheskogo Instituta Imeni VA Steklova* 42:3–375
44. Schindler E, Karlsson MO (2017) A minimal continuous-time Markov pharmacometric model. *AAPS J* 19:1424–1435
45. Lu T, Yang Y, Jin JY, Kågedal M (2020) Analysis of longitudinal-ordered categorical data for muscle spasm adverse event of vismodegib: Comparison between different pharmacometric models. *CPT Pharmacometrics Syst Pharmacol* 9:96–105
46. Barreto JN, McCullough KB, Ice LL, Smith JA (2014) Antineoplastic agents and the associated myelosuppressive effects: a review. *J Pharm Pract* 27:440–446
47. Daniel D, Crawford J (2006) Myelotoxicity from chemotherapy. *Semin Oncol* 33:74–85
48. Friberg LE, Karlsson MO (2003) Mechanistic models for myelosuppression. *Invest New Drugs* 21:183–194
49. Fornari C, O'Connor LO, Yates JWT, Cheung SYA, Jodrell DI, Mettetal JT et al (2018) Understanding hematological toxicities using mathematical modeling. *Clin Pharmacol Ther* 104:644–654
50. Yue C, Ducharme MP (2016) Empirical models, mechanistic models, statistical moments, and noncompartmental analysis. In: Shargel L, Yu AB (eds) *Applied biopharmaceutics & pharmacokinetics*. McGraw Hill, New York
51. Minami H, Sasaki Y, Watanabe T, Ogawa M. Pharmacodynamic Modeling of the Entire Time Course of Leukopenia after a 3-Hour Infusion of Paclitaxel. *Japanese Journal of Cancer Research*. 2001. pp. 231–238. doi:<https://doi.org/10.1111/j.1349-7006.2001.tb01086.x>
52. Karlsson MO, Molnar V, Bergh J, Freijs A, Larsson R (1998) A general model for time-dissociated pharmacokinetic-pharmacodynamic relationship exemplified by paclitaxel myelosuppression. *Clin Pharmacol Ther* 63:11–25
53. Karlsson MO, Port RE, Ratain MJ, Sheiner LB (1995) A population model for the leukopenic effect of etoposide. *Clin Pharmacol Ther* 57:325–334
54. Dy GK, Adjei AA (2013) Understanding, recognizing, and managing toxicities of targeted anticancer therapies. *CA Cancer J Clin* 63:249–279
55. Hu W, Sung T, Jessen BA, Thibault S, Finkelstein MB, Khan NK et al (2016) Mechanistic investigation of bone marrow suppression associated with palbociclib and its differentiation from cytotoxic chemotherapies. *Clin Cancer Res* 22:2000–2008



56. Krzyzanski W, Jusko WJ (2002) Multiple-pool cell lifespan model of hematologic effects of anticancer agents. *J Pharmacokinet Pharmacodyn* 29:311–337
57. de Schultink AHMV, de Vries Schultink AHM, Suleiman AA, Schellens JHM, Beijnen JH, Huitema ADR (2016) Pharmacodynamic modeling of adverse effects of anti-cancer drug treatment. *Eur J Clin Pharmacol* 72:645–653. <https://doi.org/10.1007/s00228-016-2030-4>
58. Friberg LE, Henningson A, Maas H, Nguyen L, Karlsson MO (2002) Model of chemotherapy-induced myelosuppression with parameter consistency across drugs. *J Clin Oncol* 20:4713–4721
59. Woo S, Krzyzanski W, Jusko WJ (2008) Pharmacodynamic model for chemotherapy-induced anemia in rats. *Cancer Chemother Pharmacol* 62:123–133
60. Kheifetz Y, Scholz M (2019) Modeling individual time courses of thrombopoiesis during multi-cyclic chemotherapy. *PLoS Comput Biol* 15:e1006775
61. Sun W, O'Dwyer PJ, Finn RS, Ruiz-Garcia A, Shapiro GI, Schwartz GK et al (2017) Characterization of neutropenia in advanced cancer patients following palbociclib treatment using a population pharmacokinetic-pharmacodynamic modeling and simulation approach. *J Clin Pharmacol* 57(9):1159–1173. <https://doi.org/10.1002/jcph.902>
62. Friberg LE, Sandström M, Karlsson MO (2010) Scaling the time-course of myelosuppression from rats to patients with a semi-physiological model. *Invest New Drugs* 28:744–753
63. Craig M (2017) Towards quantitative systems pharmacology models of chemotherapy-induced neutropenia. *CPT Pharmacometrics Syst Pharmacol* 6:293–304
64. Quartino AL, Karlsson MO, Lindman H, Friberg LE (2014) Characterization of endogenous G-CSF and the inverse correlation to chemotherapy-induced neutropenia in patients with breast cancer using population modeling. *Pharm Res* 31:3390–3403
65. Cohen-Rabbie S, Berges AC, Rekić D, Parkinson J, Dota C, Tomkinson HK (2021) QT prolongation risk assessment in oncology: Lessons learned from small-molecule new drug applications approved during 2011–2019. *J Clin Pharmacol* 61:1106–1117
66. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). ICH E14 guideline: the clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs questions & answers (R3) 2015 [2/6/2016]. 61134]. 2017 Jun. <https://www.fda.gov/media/71379/download>
67. Faucette S, Wagh S, Trivedi A, Venkatakrishnan K, Gupta N (2018) Reverse translation of US food and drug administration reviews of oncology new molecular entities approved in 2011–2017: lessons learned for anticancer drug development. *Clin Transl Sci* 11:123–146
68. Garnett C, Bonate PL, Dang Q, Ferber G, Huang D, Liu J et al (2018) Scientific white paper on concentration-QTc modeling. *J Pharmacokinet Pharmacodyn* 45:383–397
69. Molnar J, Zhang F, Weiss J, Ehlert FA, Rosenthal JE (1996) Diurnal pattern of QTc interval: how long is prolonged? Possible relation to circadian triggers of cardiovascular events. *J Am Coll Cardiol* 27:76–83
70. Belderbos BPSI, de Wit R, Chien C, Mitselos A, Hellemans P, Jiao J et al (2018) An open-label, multicenter, phase Ib study investigating the effect of apalutamide on ventricular repolarization in men with castration-resistant prostate cancer. *Cancer Chemother Pharmacol* 82:457–468
71. Orihashi Y, Ohwada S, Kumagai Y (2021) Concentration–QTc analysis for single arm studies. *J Pharmacokinet Pharmacodyn* 48:203–211
72. Garnett CE, Zhu H, Malik M, Fossa AA, Zhang J, Badilini F et al (2012) Methodologies to characterize the QT/corrected QT interval in the presence of drug-induced heart rate changes or other autonomic effects. *Am Heart J* 163:912–930
73. New approaches for an integrated nonclinical-clinical QT/proarrhythmic. 1 Nov 2021. Accessed 18 July 2021. <https://www.fda.gov/drugs/news-events-human-drugs/new-approaches-integrated-nonclinical-clinical-qtproarrhythmic-risk-assessment-10152020-10162020>
74. The International Council for Harmonisation of Technical Requirements for pharmaceuticals for human use. E14 and S7B clinical and nonclinical evaluation of QT/QTc interval prolongation and proarrhythmic potential—questions and answers (2020) <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e14-and-s7b-clinical-and-nonclinical-evaluation-qtqt-c-interval-prolongation-and-proarrhythmic>
75. Garnett C, Needleman K, Liu J, Brundage R, Wang Y (2016) Operational characteristics of linear concentration-QT models for assessing QTc interval in the thorough QT and phase I clinical studies. *Clin Pharmacol Ther* 100(2):170–178. <https://doi.org/10.1002/cpt.361>
76. Ferber G, Lorch U, Täubel J (2015) The power of phase I studies to detect clinical relevant QTc prolongation: a resampling simulation study. *BioMed Res Int*. <https://doi.org/10.1155/2015/293564>
77. Ferber G, Zhou M, Darpo B (2015) Detection of QTc effects in small studies-implications for replacing the thorough QT study. *Ann Noninvasive Electrocardiol* 20(4):368–377. <https://doi.org/10.1111/anec.12227>
78. Ferber G, Zhou M, Dota C, Garnett C, Keirns J, Malik M et al (2017) Can bias evaluation provide protection against false-negative results in QT studies without a positive control using exposure-response analysis? *J Clin Pharmacol* 57(1):85–95. <https://doi.org/10.1002/jcph.779>
79. Krishnatry AS, Hanze E, Bergsma T, Dhar A, Prohn M, Ferron-Brady G (2021) Exposure-response analysis of adverse events associated with molibresib and its active metabolites in patients with solid tumors. *CPT Pharmacometrics Syst Pharmacol*. <https://doi.org/10.1002/psp4.12724>
80. Baverel P, Roskos L, Tatipalli M, Lee N, Stockman P, Taboada M et al (2019) Exposure-response analysis of overall survival for tremelimumab in unresectable malignant mesothelioma: the confounding effect of disease status. *Clin Transl Sci* 12(5):450–458. <https://doi.org/10.1111/cts.12633>
81. Li C, Wang B, Chen S-C, Wada R, Lu D, Wang X et al (2017) Exposure–response analyses of trastuzumab emtansine in patients with HER2-positive advanced breast cancer previously treated with trastuzumab and a taxane. *Cancer Chemother Pharmacol* 80:1079–1090. <https://doi.org/10.1007/s00280-017-3440-4>
82. Chatterjee MS, Elassaiss-Schaap J, Lindauer A, Turner DC, Sostelly A, Freshwater T et al (2017) Population pharmacokinetic/pharmacodynamic modeling of tumor size dynamics in pembrolizumab-treated advanced melanoma. *CPT Pharmacometrics Syst Pharmacol* 66(1):29–39. <https://doi.org/10.1002/psp4.12140>
83. Gupta N, Labotka R, Liu G, Hui A-M, Venkatakrishnan K (2016) Exposure-safety-efficacy analysis of single-agent ixazomib, an oral proteasome inhibitor, in relapsed/refractory multiple myeloma: dose selection for a phase 3 maintenance study. *Invest New Drugs* 34:338–346
84. Fostvedt LK, Hibma JE, Masters JC, Vandendries E, Ruiz-Garcia A (2019) Pharmacokinetic/pharmacodynamic modeling to support the re-approval of gemtuzumab ozogamicin. *Clin Pharmacol Ther* 106(5):1006–1017. <https://doi.org/10.1002/cpt.1500>



85. Turner DC, Kondic AG, Anderson KM, Robinson AG, Garon EB, Riess JW et al (2018) Pembrolizumab exposure-response assessments challenged by association of cancer cachexia and catabolic clearance. *Clin Cancer Res* 25(10):5841–5849. <https://doi.org/10.1158/1078-0432.ccr-18-0415>
86. Wang Y, Zhu H, Madabushi R, Liu Q, Huang S, Zineh I (2019) Model-informed drug development: current US regulatory practice and future considerations. *Clin Pharmacol Ther* 105(4):899–911. <https://doi.org/10.1002/cpt.1363>
87. Tabernero J, Ohtsu A, Muro K, Van Cutsem E, Oh SC, Bodoky G et al (2017) Exposure-response analyses of ramucirumab from two randomized, phase III trials of second-line treatment for advanced gastric or gastroesophageal junction cancer. *Mol Cancer Ther* 16(10):2215–2222. <https://doi.org/10.1158/1535-7163.mct-16-0895>
88. Jones RL, Mo G, Baldwin JR, Peterson PM, Ilaria RL, Conti I et al (2019) Exposure–response relationship of olaratumab for survival outcomes and safety when combined with doxorubicin in patients with soft tissue sarcoma. *Cancer Chemother Pharmacol* 83:191–199. <https://doi.org/10.1007/s00280-018-3723-4>
89. Suleiman AA, Frechen S, Scheffler M, Zander T, Kahraman D, Kobe C et al (2015) Modeling tumor dynamics and overall survival in advanced non-small-cell lung cancer treated with erlotinib. *J Thorac Oncol* 10:84–92
90. Harnisch L, Shepard T, Pons G, Della Pasqua O (2013) Modeling and simulation as a tool to bridge efficacy and safety data in special populations. *CPT Pharmacometrics Syst Pharmacol*. <https://doi.org/10.1038/psp.2013.6>
91. Gupta N, Zhao Y, Hui A-M, Esseltine D-L, Venkatakrishnan K (2015) Switching from body surface area-based to fixed dosing for the investigational proteasome inhibitor ixazomib: a population pharmacokinetic analysis. *Br J Clin Pharmacol* 79(5):789–800. <https://doi.org/10.1111/bcp.12542>
92. Sehgal K, Bulumulle A, Brody H, Gill RR, Macherla S, Quilleri A et al (2021) Association of extended dosing intervals or delays in pembrolizumab-based regimens with survival outcomes in advanced non-small-cell lung cancer. *Clin Lung Cancer* 22:e379–e389
93. Pierre V, Guo X, Gonzalez-Garcia I, Morsli N, Yovine AJ, Li W et al (2020) Overall survival modeling and association with serum biomarkers in durvalumab-treated patients with head and neck cancer. *J Clin Oncol*. [https://doi.org/10.1200/jco.2020.38.15\\_suppl.6549](https://doi.org/10.1200/jco.2020.38.15_suppl.6549)
94. Claret L, Jin JY, Ferté C, Winter H, Girish S, Stroh M et al (2018) A model of overall survival predicts treatment outcomes with atezolizumab versus chemotherapy in non-small cell lung cancer based on early tumor kinetics. *Clin Cancer Res*. <https://doi.org/10.1158/1078-0432.ccr-17-3662>
95. Tardivon C, Desmée S, Kerioui M, Bruno R, Wu B, Mentré F et al (2019) Association between tumor size kinetics and survival in patients with urothelial carcinoma treated with atezolizumab: implication for patient follow-up. *Clin Pharmacol Ther* 106(4):810–820. <https://doi.org/10.1002/cpt.1450>
96. Chojniak R, Yu LS, Younes RN (2006) Response to chemotherapy in patients with lung metastases: how many nodules should be measured? *Cancer Imaging* 6:107–112
97. Johnson K, Gomez A, Burton J, White D, Chakravarty A, Schmid A et al (2019) Directional inconsistency between Response Evaluation Criteria in Solid Tumors (RECIST) time to progression and response speed and depth. *Eur J Cancer* 109:196–203. <https://doi.org/10.1016/j.ejca.2018.11.008>
98. Wolchok JD, Hoos A, O’Day S, Weber JS, Hamid O, Lebbé C et al (2009) Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clin Cancer Res* 15:7412–7420
99. Cheson BD, Bennett JM, Kopecky KJ, Büchner T, Willman CL, Estey EH et al (2003) Revised recommendations of the international working group for diagnosis, standardization of response criteria, treatment outcomes, and reporting standards for therapeutic trials in acute myeloid leukemia. *J Clin Oncol* 21(24):4642–4649. <https://doi.org/10.1200/jco.2003.04.036>
100. Cross NCP, White HE, Müller MC, Saglio G, Hochhaus A (2012) Standardized definitions of molecular response in chronic myeloid leukemia. *Leukemia* 26(10):2172–2175. <https://doi.org/10.1038/leu.2012.104>
101. Mateos M-V, Kumar S, Dimopoulos MA, González-Calle V, Kastiris E, Hajek R et al (2020) International Myeloma Working Group risk stratification model for smoldering multiple myeloma (SMM). *Blood Cancer J* 10:102
102. Al-Huniti N, Feng Y, Yu JJ, Lu Z, Nagase M, Zhou D et al (2020) Tumor growth dynamic modeling in oncology drug development and regulatory approval: past, present, and future opportunities. *CPT Pharmacometrics Syst Pharmacol* 9:419–427
103. Maitland ML, Wilkerson J, Karovic S, Zhao B, Flynn J, Zhou M et al (2020) Enhanced detection of treatment effects on metastatic colorectal cancer with volumetric CT measurements for tumor burden growth rate evaluation. *Clin Cancer Res* 26:6464–6474
104. Wang Y, Sung C, Dartois C, Ramchandani R, Booth BP, Rock E et al (2009) Elucidation of relationship between tumor size and survival in non-small-cell lung cancer patients can aid early decision making in clinical drug development. *Clin Pharmacol Ther* 86(2):167–174. <https://doi.org/10.1038/clpt.2009.64>
105. Chanu P, Musib L, Wang X, Cheeti S, Girish S, Bruno R et al (2021) Vismodegib efficacy in advanced basal cell carcinoma maintained with 8-week dose interruptions: a model-based evaluation. *J Invest Dermatol* 141:930–933
106. Hamid O, Bruno R, Fasso M, O’Hear C, Wu B (2020) Safety, clinical activity, and biological correlates of response in patients with metastatic melanoma: results from a phase I trial of atezolizumab-response. *Clin Cancer Res* 26(10):2436
107. Bruno R, Bottino D, de Alwis DP, Fojo AT, Guedj J, Liu C et al (2020) Progress and opportunities to advance clinical cancer therapeutics using tumor dynamic models. *Clin Cancer Res* 26:1787–1795
108. Magni P, Germani M, De Nicolao G, Bianchini G, Simeoni M, Poggessi I et al (2008) A minimal model of tumor growth inhibition. *IEEE Trans Biomed Eng* 55(12):2683–2690. <https://doi.org/10.1109/tbme.2008.913420>
109. Ribba B, Holford NH, Magni P, Trocóniz I, Gueorguieva I, Girard P et al (2014) A review of mixed-effects models of tumor growth and effects of anticancer drug treatment used in population analysis. *CPT Pharmacometrics Syst Pharmacol* 3:e113
110. Bonate PL, Suttle B (2013) Effect of censoring due to progressive disease on tumor size kinetic parameter estimates. *AAPS J* 15:832–839
111. Singleton KW, Porter AB, Hu LS, Johnston SK, Bond KM, Rickertsen CR et al (2020) Days gained response discriminates treatment response in patients with recurrent glioblastoma receiving bevacizumab-based therapies. *Neuro-Oncology Adv*. <https://doi.org/10.1093/oaajnl/vdaa085>
112. Bottino D, Hang Y, Neuwirth R, Patel C, Sedarati F, Faller DV et al (2019) Abstract A073: using prebaseline and on-treatment tumor assessments to compare time to tumor size progression of hormonal therapy sapanisertib vs an extrapolated hormonal therapy alone from a single-arm phase 2 study. *Clin Trials*. <https://doi.org/10.1158/1535-7163.targ-19-a073>
113. Zheng Y, Narwal R, Jin C, Baverel PG, Jin X, Gupta A et al (2018) Population modeling of tumor kinetics and overall survival to identify prognostic and predictive biomarkers of efficacy for durvalumab in patients with urothelial carcinoma. *Clin*

- Pharmacol Ther 103(4):643–652. <https://doi.org/10.1002/cpt.986>
114. Schindler E, Amantea MA, Karlsson MO, Friberg LE (2017) A pharmacometric framework for axitinib exposure, efficacy, and safety in metastatic renal cell carcinoma patients. *CPT Pharmacometrics Syst Pharmacol* 6:373–382
  115. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R et al (2009) New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 45:228–247
  116. Blumenthal GM, Karuri SW, Zhang H, Zhang L, Khozin S, Kazandjian D et al (2015) Overall response rate, progression-free survival, and overall survival with targeted and standard therapies in advanced non-small-cell lung cancer: US Food and Drug Administration trial-level and patient-level analyses. *J Clin Oncol* 33(9):1008–1014. <https://doi.org/10.1200/jco.2014.59.0489>
  117. Anagnostou V, Yarchoan M, Hansen AR, Wang H, Verde F, Sharon E et al (2017) Immuno-oncology trial endpoints: capturing clinically meaningful activity. *Clin Cancer Res* 23(17):4959–4969. <https://doi.org/10.1158/1078-0432.ccr-16-3065>
  118. Hodi FS, Stephen Hodi F, Hwu W-J, Kefford R, Weber JS, Daud A et al (2016) Evaluation of immune-related response criteria and RECIST v1.1 in patients with advanced melanoma treated with pembrolizumab. *J Clin Oncol* 34(13):1510–1517. <https://doi.org/10.1200/jco.2015.64.0391>
  119. Camelliti S, Le Noci V, Bianchi F, Moscheni C, Arnaboldi F, Gagliano N et al (2020) Mechanisms of hyperprogressive disease after immune checkpoint inhibitor therapy: what we (don't) know. *J Exp Clin Cancer Res*. <https://doi.org/10.1186/s13046-020-01721-9>
  120. Champiat S, Ferrara R, Massard C, Besse B, Marabelle A, Soria J-C et al (2018) Hyperprogressive disease: recognizing a novel pattern to improve patient management. *Nat Rev Clin Oncol* 15:748–762
  121. Moore H (2016) The American Conference on Pharmacometrics 2016 (ACoP7) W–W29
  122. Lu L, Derclé L, Zhao B, Schwartz LH (2021) Deep learning for the prediction of early on-treatment response in metastatic colorectal cancer from serial medical imaging. *Nat Commun* 12:6654
  123. Wibmer AG, Morris MJ, Gonen M, Zheng J, Hricak H, Larson S et al (2021) Quantification of metastatic prostate cancer whole-body tumor burden with 18F-FDG PET parameters and associations with overall survival after first-line abiraterone or enzalutamide: a single-center retrospective cohort study. *J Nucl Med* 62:1050–1056
  124. Michor F, Hughes TP, Iwasa Y, Branford S, Shah NP, Sawyers CL et al (2005) Dynamics of chronic myeloid leukaemia. *Nature* 435:1267–1270
  125. Cheson BD, Pfistner B, Juweid ME, Specht L, Rosen ST, Gascoyne R, et al (2005) Revised response criteria for malignant lymphomas from the members of the International Harmonization Project (IHP) of the Competence Network Malignant Lymphoma, Represented by: *Blood*. 18–18. <https://doi.org/10.1182/blood.v106.11.18.18>
  126. Meignan M, Cottreau AS, Versari A, Chartier L, Dupuis J, Boussetta S et al (2016) Baseline metabolic tumor volume predicts outcome in high-tumor-burden follicular lymphoma: a pooled analysis of three multicenter studies. *J Clin Orthod* 34:3618–3626
  127. Lu Z, James A, Hill J, Rosales M, Bahceci E, Bonate P (2019) Pharmacokinetic-pharmacodynamic analysis of minimal residual disease with data at boundary: gilteritinib a tyrosine kinase inhibitor for the treatment of acute myeloid leukemia. *Clinical Pharmacology & Therapeutics*. Wiley, pp S62–S63
  128. Rawstron AC, Gregory WM, de Tute RM, Davies FE, Bell SE, Drayson MT et al (2015) Minimal residual disease in myeloma by flow cytometry: independent prediction of survival benefit per log reduction. *Blood* 125:1932–1935
  129. Kostopoulos IV, Ntanasis-Stathopoulos I, Gavriatopoulou M, Tsitsilonis OE, Terpos E (2020) Minimal residual disease in multiple myeloma: current landscape and future applications with immunotherapeutic approaches. *Front Oncol* 10:860
  130. Cook SF, Bies RR (2016) Disease progression modeling: key concepts and recent developments. *Curr Pharmacol Rep* 2:221–230
  131. Hussain M, Goldman B, Tangen C, Higano CS, Petrylak DP, Wilding G et al (2009) Prostate-specific antigen progression predicts overall survival in patients with metastatic prostate cancer: data from Southwest Oncology Group Trials 9346 (Intergroup Study 0162) and 9916. *J Clin Oncol* 27:2450–2456
  132. Ji G, Huang C, Song G, Xiong G, Fang D, Wang H et al (2017) Are the pathological characteristics of prostate cancer more aggressive or more indolent depending upon the patient age? *Biomed Res Int* 2017:1438027
  133. de Charry F, Colombaro O, You B, Ruffion A, Paparel P, Wilbaux M et al (2016) Identification of most aggressive carcinoma among patients diagnosed with prostate cancer using mathematical modeling of prostate-specific antigen increases. *Clin Genitourin Cancer* 14:210–217.e1
  134. Danila DC, Fleisher M, Scher HI (2011) Circulating tumor cells as biomarkers in prostate cancer. *Clin Cancer Res* 17:3903–3912
  135. Allard WJ, Jeffrey Allard W, Matera J, Craig Miller M, Repollet M, Connelly MC et al (2004) Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clin Cancer Res* 10(20):6897–6904. <https://doi.org/10.1158/1078-0432.ccr-04-0378>
  136. Wilbaux M, Tod M, De Bono J, Lorente D, Mateo J, Freyer G et al (2015) A joint model for the kinetics of CTC count and PSA concentration during treatment in metastatic castration-resistant prostate cancer. *CPT Pharmacometrics Syst Pharmacol* 4(5):277–285
  137. Vaisman-Mentesh A, Gutierrez-Gonzalez M, DeKosky BJ, Wine Y (2020) The molecular mechanisms that underlie the immune biology of anti-drug antibody formation following treatment with monoclonal antibodies. *Front Immunol* 11:1951
  138. Manis JP, Feldweg AM (2019) Overview of therapeutic monoclonal antibodies. *US Pharm* 44:31
  139. FDA Guidance: Immunogenicity Assessment for Therapeutic Protein Products (2014) U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). <https://www.fda.gov/media/85017/download>
  140. FDA Guidance: Immunogenicity Testing of Therapeutic Protein Products—Developing and Validating Assays for Anti-Drug Antibody Detection (2019) {U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER)}. <https://www.fda.gov/media/119788/download>
  141. Sailstad JM, Amaravadi L, Clements-Egan A, Gorovits B, Myler HA, Pillutla RC et al (2014) A white paper—consensus and recommendations of a global harmonization team on assessing the impact of immunogenicity on pharmacokinetic measurements. *AAPS J* 16:488–498
  142. Committee For Medicinal Products for Human Use (2007) Guideline on immunogenicity assessment of biotechnology-

- derived therapeutic proteins. European Medicines Agency, London
143. Mire-Sluis AR, Barrett YC, Devanarayan V, Koren E, Liu H, Maia M et al (2004) Recommendations for the design and optimization of immunoassays used in the detection of host antibodies against biotechnology products. *J Immunol Methods* 289:1–16
  144. Tovey MG (2011) Detection and quantification of antibodies to biopharmaceuticals: practical and applied considerations. Wiley, New York
  145. Chirmule N, Jawa V, Meibohm B (2012) Immunogenicity to therapeutic proteins: impact on PK/PD and efficacy. *AAPS J* 14:296–302
  146. U.S. Package Insert of Tencentric®. Accessed 10 Sept 2020. [https://www.accessdata.fda.gov/drugsatfda\\_docs/label/2020/761034s029ibl.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/label/2020/761034s029ibl.pdf)
  147. Kong S, Heinzmann D, Lauer S, Tian L (2021) Weighted approach for estimating effects in principal strata with missing data for a categorical post-baseline variable in randomized controlled trials. *Stat Biopharm Res* 1–11
  148. Bornkamp B, Bermann G (2020) Estimating the treatment effect in a subgroup defined by an early post-baseline biomarker measurement in randomized clinical trials with time-to-event endpoint. *Stat Biopharm Res* 12:19–28
  149. Golchin A, Farahany TZ (2019) Biological products: cellular therapy and FDA approved products. *Stem Cell Rev Rep* 15(2):166–175. <https://doi.org/10.1007/s12015-018-9866-1>
  150. Center for Biologics Evaluation, Research. Approved cellular and gene therapy products. In: U.S. Food and Drug Administration. 1 Mar 2022. Accessed 30 April 2022. <https://www.fda.gov/vaccines-blood-biologics/cellular-gene-therapy-products/approved-cellular-and-gene-therapy-products>
  151. Pettitt D, Arshad Z, Smith J, Stanic T, Holländer G, Brindley D (2018) CAR-T cells: a systematic review and mixed methods analysis of the clinical trial landscape. *Mol Ther* 26(2):342–353. <https://doi.org/10.1016/j.ymthe.2017.10.019>
  152. Reddy SR, Llukmani A, Hashim A, Haddad DR, Patel DS, Ahmad F et al (2021) The role of chimeric antigen receptor-T cell therapy in the treatment of hematological malignancies: advantages, trials, and tribulations, and the road ahead. *Cureus*. <https://doi.org/10.7759/cureus.13552>
  153. Liu C, Ayyar VS, Zheng X, Chen W, Zheng S, Mody H et al (2021) Model-based cellular kinetic analysis of chimeric antigen receptor-T cells in humans. *Clin Pharmacol Ther* 109(3):716–727. <https://doi.org/10.1002/cpt.2040>
  154. Berdeja JG, Madduri D, Usmani SZ, Jakubowiak A, Agha M, Cohen AD et al (2021) Ciltacabtagene autoleucel, a B-cell maturation antigen-directed chimeric antigen receptor T-cell therapy in patients with relapsed or refractory multiple myeloma (CARTITUDE-1): a phase 1b/2 open-label study. *Lancet* 398:314–324
  155. Raje N, Berdeja J, Lin Y, Siegel D, Jagannath S, Madduri D et al (2019) Anti-BCMA CAR T-Cell Therapy bb2121 in relapsed or refractory multiple myeloma. *N Engl J Med* 380(18):1726–1737. <https://doi.org/10.1056/nejmoa1817226>
  156. U.S. Package Insert of Abecma® (2021). <https://www.fda.gov/vaccines-blood-biologics/abecma-idecabtagene-vicleucel>
  157. Singh AP, Zheng X, Lin-Schmidt X, Chen W, Carpenter TJ, Zong A et al (2020) Development of a quantitative relationship between CAR-affinity, antigen abundance, tumor cell depletion and CAR-T cell expansion using a multiscale systems PK-PD model. *mAbs* 12:1688616. <https://doi.org/10.1080/19420862.2019.1688616>
  158. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). Guidances for Industry: Cellular and Gene Therapy. In: FDA. (2021). <https://www.fda.gov/vaccines-blood-biologics/biologics-guidances/cellular-gene-therapy-guidances>
  159. Stein AM, Grupp SA, Levine JE, Laetsch TW, Pulsipher MA, Boyer MW et al (2019) Tisagenlecleucel model-based cellular kinetic analysis of chimeric antigen receptor-T cells. *CPT Pharmacometrics Syst Pharmacol* 8(5):285–295. <https://doi.org/10.1002/psp4.12388>

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