In silico cancer immunotherapy trials uncover the consequences of therapy-specific response patterns for clinical trial design and outcome

Jeroen H.A. Creemers[1,2], Kit C.B. Roes[3], Niven Mehra[4], Carl G. Figdor[1,2], I. Jolanda M. de Vries[1], Johannes Textor[1,5*]

1) Department of Tumor Immunology, Radboud Institute for Molecular Life Sciences, Radboudumc, Nijmegen, The Netherlands
2) Oncode Institute, Nijmegen, The Netherlands
3) Department of Health Evidence, Section Biostatistics, Radboudumc, Nijmegen, The Netherlands
4) Department of Medical Oncology, Radboudumc, Nijmegen, The Netherlands
5) Data Science Department, Institute for Computing and Information Sciences, Radboud University, Nijmegen, The Netherlands

* Corresponding author:
Dr. J. Textor
Department of Tumor Immunology
Radboud Institute for Molecular Life Sciences, Radboudumc
Geert Grooteplein 26
6500 HB Nijmegen (P.O. Box 9101), The Netherlands
E-mail: johannes.textor@radboudumc.nl
Phone: 0031-24 361 76 00

ABSTRACT

Background

Late-stage cancer immunotherapy trials strive to demonstrate the clinical efficacy of novel immunotherapies, which is leading to exceptional responses and long-term survival in subsets of patients. To establish the clinical efficacy of an immunotherapy, it is critical to adjust the trial's design to the expected immunotherapy-specific response patterns.

Methods

In silico cancer immunotherapy trials are virtual clinical trials that simulate the kinetics and outcome of immunotherapy depending on the type and treatment schedule. We used an ordinary differential equation model to simulate (1) cellular interactions within the tumor microenvironment, (2) translates these into disease courses in patients, and (3) assemble populations of virtual patients to simulate *in silico* late-stage immunotherapy, chemotherapy, or combination trials. We predict trial outcomes and investigate how therapy-specific response patterns affect the probability of their success.

Results

In silico cancer immunotherapy trials reveal that immunotherapy-derived survival kinetics – such as delayed curve separation and plateauing curve of the treatment arm – arise naturally due to biological interactions in the tumor microenvironment. *In silico* clinical trials are capable of translating these biological interactions into survival kinetics. Considering four aspects of clinical trial design – sample size calculations, endpoint and randomization rate selection, and interim analysis planning – we illustrate that failing to consider such distinctive response patterns can significantly reduce the power of novel immunotherapy trials.

Conclusion

In silico trials have three significant implications for immuno-oncology. First, they provide an economical approach to verify the robustness of biological assumptions underlying an immunotherapy trial and help to scrutinize its design. Second, the biological basis of these trials facilitates and encourages communication between biomedical researchers, doctors, and trialists. Third, its application as an educational tool can illustrate design principles to scientists in training, contributing to improved designs and higher success rates of future immunotherapy trials.

INTRODUCTION

Immunotherapy is revolutionizing the treatment landscape for patients with advanced cancers. While the number of immuno-oncology drugs under investigation is rising rapidly – around 4700 agents are currently in the development pipeline - the need to further improve patient outcomes remains high[1]. Well-designed immunotherapy trials are crucial to establish advances in clinical outcomes robustly. Unfortunately, the odds for cancer treatments to successfully pass the development pipeline are unfavorable, and only a minority of the treatments (5-10%) will ultimately obtain market approval[2, 3, 4]. Even for cancer therapies that reach late-stage development, approval rates remain modest at around 27%[5]. The primary reason in most of these trials (i.e., 63.7%) is failure to demonstrate efficacy[5], which can be partly attributed to suboptimal trial design choices based on overly optimistic assumptions of the treatment effect. Such assumptions may be used to erroneously justify low numbers of patients or inappropriate endpoints and lower the power of these trials[5, 6].

Undoubtedly, design choices in immunotherapy trials are complex, and conventional design methods are not naturally well attuned to the unique characteristics of immunotherapies[7]. Their broad spectrum – ranging from immunomodulators to cell therapies, cancer vaccines, oncolytic viruses, and CD3-targeted bispecific antibodies – illustrates the variety in molecular mechanisms, leading to novel toxicity profiles, response patterns, and survival kinetics[8, 9, 10]. These observations render a 'one-design-fits-all' approach futile and stress the need for immunotherapy- or even combination-therapy-tailored designs.

Immunotherapies are known to induce a delayed clinical effect and long-term overall survival (OS) in only a subset of patients[11]. The survival curve reflects these phenomena by a delayed curve separation and a plateau of the treatment arm at later stages of the trial, respectively[12]. Many immunotherapies, thereby, violate a fundamental premise that underlies the design of many trials: the proportional hazard assumption (PHA) – essentially stating that the treatment effect should remain constant over time[13]. As a result, immunotherapy trials based on this principle can have an overestimated power[12, 13] and require a longer follow-up to demonstrate efficacy than initially planned[12], increasing the likelihood of a negative trial.

These issues led to the development of innovative methods such as novel radiological criteria to quantify tumor responses[9, 14, 15], (surrogate) endpoints to capture unique survival kinetics[10, 16, 17, 18, 19], biomarkers to enrich for potentially responding patients for treatment[20, 21, 22, 23], and statistical methods to retain a trial's power in the presence of non-conventional survival kinetics[24, 25, 26]. Despite the plethora of available methods, it still remains complicated to predict trial outcomes *in advance* and adjust the methodology accordingly. The stakes are high: accurate predictions could augment the likelihood of a positive trial, whereas a misjudgment could result in a negative trial, potentially compromising patient benefit, vast amounts of work, and (public) research funds.

In this study, we use late-stage *in silico* cancer immunotherapy trials to investigate how design decisions affect the trial outcome in the context of cancer immunotherapy, possibly combined with chemotherapy. The mechanism-based nature of these trials allows researchers to translate cellular processes in the tumor microenvironment and interventions thereon into immunotherapy-specific response patterns, survival kinetics, and outcomes of novel immunotherapy trials. An *in silico* immunotherapy trial is based on clear-cut biological assumptions and provides an intuitive means to predict risk profiles and treatment efficacy. Moreover, it equips researchers with a tool to verify trial designs and analysis strategies of upcoming trials before the trials' execution. First, we show that our simulations replicate late-stage immunotherapy or combination trials realistically and capture their typical survival kinetics. Then, we demonstrate various applications of these trial simulations, including the ability to predict the trial outcomes and calculate sample sizes for specific treatment and control groups. Finally, we illustrate the consequences of (not) considering immunotherapy-specific response patterns in settings selected for educational purposes, such as selecting survival endpoints and randomization ratios of upcoming trials and planning interim analyses.

METHODS

Mechanism-based model of the tumor microenvironment

We extended our previously published ordinary differential equation (ODE) model that describes cancer development in the tumor microenvironment and the subsequently induced anti-tumor response in patients[27]. Briefly, the model describes cancer onset and progression, starting with the malignant degeneration of a single cell into a tumor cell. This tumor cell divides, leading to a proliferating mass (ρ: growth rate) of tumor cells (T; equation 1; Figure 1A). An anti-tumor immune response is induced within the tumor microenvironment, leading to the killing of tumor cells (at killing rate ξ; equation 1). The killing of tumor cells is implemented using a quasi-steady state approximation proposed by Borghans *et al*.[28, 29]. For details on the implementation, we refer to our previous work[27]. The immune cells within the tumor microenvironment originate from tumor-draining lymph nodes, where naive cytotoxic T cells (N) turn into activated T cells (S) at priming rate α. Since a larger tumor mass is generally more immunogenic, the priming rate is scaled by the tumor size ($\frac{T}{10+T}$; equations 3 and 4). Primed T cells proliferate in the lymph nodes at rate p_s and migrate into the tumor microenvironment at rate m_s (equation 2). The natural life span of T cells in the tumor microenvironment is modeled by incorporating a natural death rate δ of T cells. The system of ODEs is presented below:

120
$$\begin{cases} \frac{dT}{dt} = \rho T^{\frac{4}{5}} - \frac{\xi I T}{1 + \frac{I}{h_I} + \frac{T}{h_T}} & (Eq.1) \\ \frac{dI}{dt} = m_s S - \delta I & (Eq.2) \\ \frac{dS}{dt} = \alpha \left(\frac{T}{10^7 + T}\right) N + p_s S - m_s S & (Eq.3) \\ \frac{dN}{dt} = -\alpha \left(\frac{T}{10^7 + T}\right) N & (Eq.4) \end{cases}$$

121

122 *Table 1: Overview of model parameters*

Interpatient variability	Symbol	Parameter	Default value or range	Description
Variable	ρ	Tumor growth rate (cells/day)	1.76 - 150	Growth rate of tumor cells
	Δρ	Growth rate decline	-0.6 - 0	Rate at which tumor cell proliferation declines over time
	ρ_δ	Growth rate decline decay	-2 - 0	Rate at which the tumor growth rate decline decreases.
Universal	ξ	Relative T cell killing rate (cells/ day)	0.001	Killing rate of cytotoxic T cells
	α	T cell priming rate (cells/day)	0.0025	Rate at which antigen-specific T cells are activated
	δ	T cell death rate (cells/day)	0.019	Death rate of T cells
	h	Michaelis constant (cells)	571	Ratio between tumor-immune cell complex formation and dissociation
	p_s	Production rate (cells/day)	1	Production rate of T cells from lymph nodes
	m_s	Migration rate (cells/day)	1	T cell migration rate from lymph node to tumor microenvironment

123

124 Since a constant time-independent tumor growth rate would unlikely be observed in a clinical
125 setting, we have added two additional parameters to the model that influence the growth rate of
126 tumors in a time-dependent manner, which are:

127 - The tumor growth rate decline (Δρ): a parameter that describes to which extent the
128 proliferation rate of tumor cells gradually declines over time.
129 - The decay rate of the tumor growth rate decline (ρ_δ; hereafter referred to as 'decline decay
130 rate'): a parameter that indicates at which pace the tumor growth rate decline decreases.

131 The model parameters and their values are listed in Table 1; their rationale is described previously[27].
132 Moreover, the values of the tumor growth rate decline and its decay rate were set to augment the
133 interpatient variability in tumor development and allow for more extensive disease trajectories. At
134 baseline, only one tumor cell and a pool of 10^6 naive T cells are presumed to be present, while activated
135 T cells are absent, yielding the following initial conditions for the simulations: T(0) = 1, I(0) = 0, S(0) =
136 0, and $N(0) = 10^6$.

Simulating untreated disease, chemotherapy, and immunotherapy in individual patients

Using this ODE model, we simulated cancer development and disease trajectories in patients. We varied the tumor properties (i.e., the tumor growth rate, the growth rate decline, and the decline decay rate) between patients extensively to guarantee interpatient variation in disease courses. Unless otherwise specified, the remaining model parameters are set to the same values for all patients (Table 1).

Each patient is simulated from cancer onset (i.e., malignant transformation of the first cell) for up to more than two years (i.e., 800 days). A simulated time step corresponds to one day. The diagnosis threshold of a tumor mass was set to 65 * 10^8 cells, corresponding to the size at which common malignancies are diagnosed. The lethal tumor burden is set to 10^{12} tumor cells (a tumor volume of approximately 10.6 dm^3).

Disease trajectories of patients with cancer can be steered with therapy. Given their prominent roles in the oncological treatment landscape, we included immune checkpoint inhibitors (ICI) and chemotherapy in the model. Both treatments function through their primary modes of action. ICI are implemented as follows: once a cancer reaches a diagnosis threshold, immune checkpoint inhibitors increase the killing rate of cytotoxic T cells (multiplication factor: 0-7), enabling them to eradicate tumor cells. The duration and potency of the ICI treatment eventually determines patient outcome.

In patients treated with chemotherapy, the immune system is still present; however, it is not boosted (as is the case during ICI treatment), hence the T cells are not potent enough to curb the tumor growth. Once a patient is diagnosed with cancer, chemotherapy can reduce the tumor growth rate with its cytotoxic capacity (multiplication factor: 0-1). Again, the duration and potency determine patient outcome. By default, the treatment duration for ICI and chemotherapy are two years and six months, respectively.

Simulating untreated cohorts in an *in silico* trial: model fitting

To expand our modeling approach from a single patient into a trial cohort, we simulated multiple patients with individualized disease courses based on unique tumor properties. As an illustrative example, we took a publicly available dataset of patients with advanced lung cancer from the North Central Cancer Treatment Group (NCCTG) and regarded the survival times of these patients as if they were untreated[30]. We fitted our model to the NCCTG dataset to show that our trial model can reproduce authentic survival kinetics as observed in clinical trials. Specifically, we searched for parameter combinations reflecting realistic survival times for each patient in our model. Since a single-parameter fitting approach could not generate a sufficiently wide range of survival times, we used a

multidimensional fitting approach. An overview of the fitting approach is depicted in Supplementary Figure 1.

Specifically, we fitted the tumor growth rate (ρ), the decline in tumor growth rate over time ($\Delta\rho$), and the decline's decay rate (ρ_δ) of patients to OS. The fitting approach comprised two processes: 1) sampling non-censored survival values and 2) translating these survival values to model parameters. We fitted a (parametric) Weibull distribution to the NCCTG lung cancer dataset (shape κ: 1.32, scale λ: 417.76; Supplementary figure 1A). The choice for a Weibull distribution is based on the Akaike information criterion and the fact that a Weibull model is a survival model from which the parameters (i.e., scale and shape) contain a mechanistic meaning and can, therefore, be interpreted. From the Weibull distribution, non-censored survival values were sampled (Supplementary figure 1B). To translate the survival times into model parameters, we generated a 3D grid containing survival times for ranges of parameters (ρ: 1.76 – 150, Δ: -0.6 - 0, ρ_δ: -2 - 0; Supplementary figure 1C). From this grid, combinations of these three parameters that led to a similar OS (i.e., the isosurfaces) were extracted with a marching cubes algorithm (Supplementary figure 1D). The isosurfaces corresponding to the non-censored survival times from the Weibull distribution were selected, and parameter combinations were sampled from these isosurfaces randomly (Supplementary figure 1E). To illustrate that these parameter combinations led to realistic survival kinetics as seen in clinical trials, we compared the original NCCTG lung cancer dataset with our simulation result (Supplementary figure 1F; Figure 1).

Simulating late-stage immunotherapy trials

Late-stage (i.e., phase III) clinical trials traditionally contain two arms: a control arm and a treatment arm. The control arm can be a placebo (i.e., untreated) or a standard of care therapy. To construct phase III *in silico* immunotherapy trials, we extended the simulations with treatment cohorts (mono-chemotherapy, mono-immunotherapy, chemoimmunotherapy, or induction chemotherapy followed by immunotherapy). These cohorts facilitate the comparison between various treatment regimens. A treatment cohort uses the same baseline parameter distribution as a control cohort. It differs in one critical aspect, though: once patients in the treatment arm reach a tumor burden that corresponds to the diagnosis threshold, patients can be treated with chemotherapy, ICI, or combination therapy, as described above. The distribution of survival parameters is, unless otherwise specified, derived from the most mature, digitized data from the CA184-024 trial, as shown below[31]. At inclusion into the trial, patients are randomly assigned to a study arm (randomization).

The primary endpoint of the trials is the 2-year OS. Given the absence of accrual times in *in silico* trials, the trial duration equals two years, which provides each patient in the trial with 24 months

follow-up at the time of analysis. If the OS endpoint is not reached for a patient, the patient is considered censored for the endpoint and regarded as such in subsequent analyses.

Power simulations

To illustrate how the analysis method can affect the outcome of immunotherapy trials, we use several simulation approaches to calculate the power of trials. Power simulations were performed as follows: per data point, 1000 clinical trials are simulated. The survival data from each trial is analyzed with a log-rank test (dependent on the proportional hazard assumption) or proportions test (Pearson's chi-squared test; independent on the proportional hazard assumption), and we count the number of positive trials (defined as $p < 0.05$). The percentage of positive trials indicates the power of the trial.

Data digitization & reconstruction

For some survival curves, the raw data was not available. Therefore, we extracted data points from the Kaplan-Meier curves with WebPlotDigitizer (https://apps.automeris.io/wpd/), and individual patient data was reconstructed with the IPDfromKM package in R.

Analyses

Analyses and visualizations were performed in R. The complete list of R packages used throughout this manuscript is provided in Supplementary Table 1. The code is accessible via https://github.com/jeroencreemers/in-silico-clinical-trials.

RESULTS

Generating trial populations based on tumor-immune dynamics

We used *in silico* cancer immunotherapy trials – a mechanism-based simulation platform of cancer-immune dynamics – to investigate the consequences of immunotherapy-specific response patterns on trial design principles[27]. Patients within these trials are simulated with an ODE model, which describes cancer development in a patient by modeling the interaction between tumor cells and the immune system[27]. In short, the model describes the following tumor-immune dynamics in the tumor microenvironment: immunogenic tumor growth leading to priming and clonal expansion of T cells in the lymph nodes, migration of effector T cells from lymph node to the tumor microenvironment, and formation of tumor-immune complexes to enable tumor cell killing (see Methods; Figure 1A). The model allows the treatment of patients with ICI and chemotherapy. ICI increase the killing rate of T cells and have a direct effect on the tumor-immune dynamics. Chemotherapy has a cytotoxic effect on the tumor, slowing down its growth rate. A detailed description of the model, including the rationale for parameter selection, has been published previously[27].

The *in silico* clinical trials describe cancer outcomes on three levels: (1) a cellular level, (2) a patient level, and (3) a trial population level. Cellular interactions in the tumor microenvironment are translated into clinical trial outcomes as follows: firstly, the ODE model is implemented, and model parameters are selected using a fitting approach (Figure 1B; see Methods). Next, individualized disease trajectories – either treated or untreated – of cancer patients are generated (Figure 1C). Eventually, patients are randomized into two cohorts to resemble conventional phase III trials: a control group (either placebo or chemotherapy) and a treatment group (immunotherapy, chemoimmunotherapy, or induction chemotherapy followed by immunotherapy; Figure 1D). Since the cellular dynamics (e.g., tumor burden over time or the efficacy of T cell killing) and survival outcomes of these patients are known and can be modified, *in silico* clinical trials are suited to answer questions like: “Assuming that a novel treatment *X* increases T cell killing by 5%, how does this translate to a survival benefit in patients? Moreover, how many patients are needed to establish this benefit in a clinical trial? When should one analyze the results?” (Figure 1E).

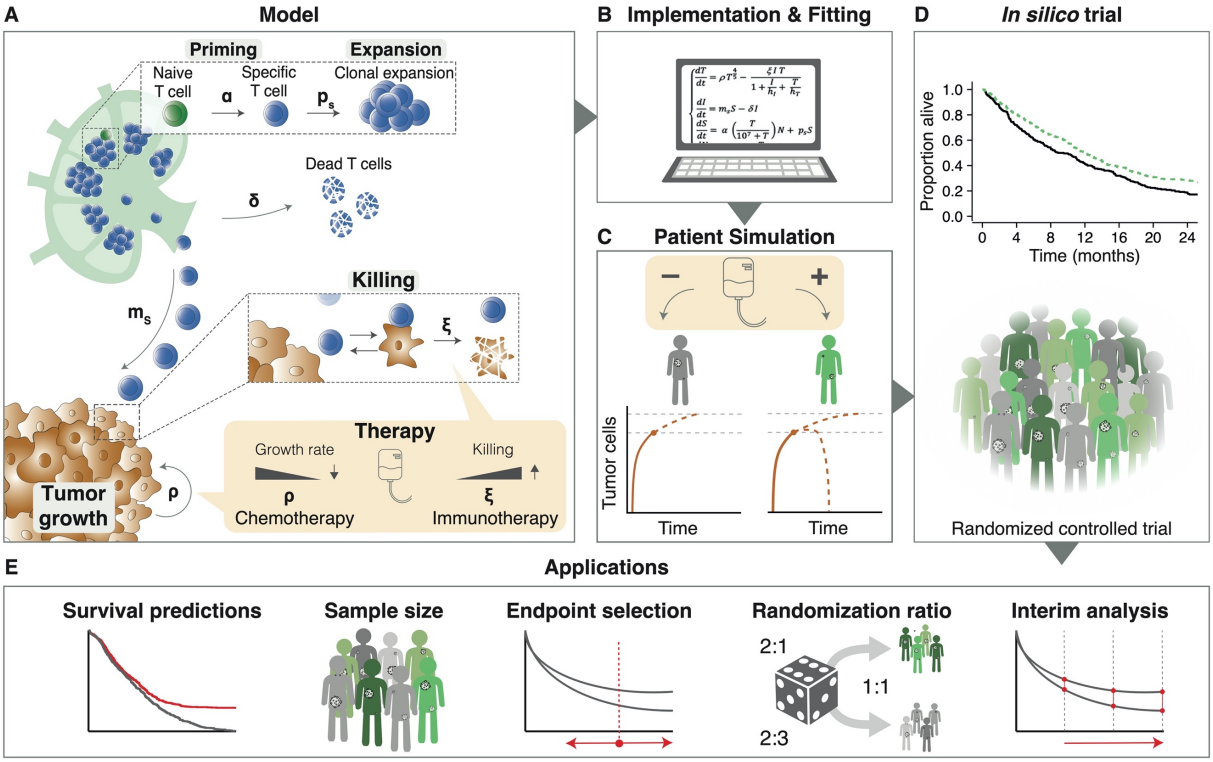

Figure 1: *In silico* late-stage immunotherapy trials and their applications.
(A) Cellular interactions between a tumor and the immune system are captured in an ODE model. This model describes immunogenic tumor growth leading to a T cell response originating from lymph nodes. Disease courses in patients could be steered by immunotherapy, chemotherapy, or a combination of both. Parameters: α = naive T cell priming rate, δ = effector T cell death rate, ξ = effector T cell killing rate, ρ = tumor growth rate, p_s = effector T cell proliferation rate, and m_s = effector T cell migration rate. **(B)** After implementation, we used clinical trial-derived survival data to fit the model parameters (see Suppl. Figure 1). **(C)** Patients received either no treatment (placebo), chemotherapy, immunotherapy, or both. Disease trajectories based on tumor-immune dynamics were simulated for each patient, resulting in individual survival outcomes. **(D)** Subsequently, cohorts of patients were constructed based on the fitted parameters to simulate actual immunotherapy trials. **(E)** Applications of these trials include predicting survival outcomes of trials, estimating appropriate sample sizes, selecting endpoints and randomization ratios, and investigating the timing of interim analyses.

In silico late-stage immunotherapy trials yield realistic survival outcomes

To illustrate that this *in silico* clinical trial approach can generate realistic survival kinetics as observed in late-stage immunotherapy trials, we fitted the simulation to three different datasets: (1) the NCCTG lung cancer survival dataset[30]; (2) the CA184-024 trial [ipilimumab + dacarbazine vs. dacarbazine in previously untreated metastatic melanoma[31]]; and (3) the CheckMate 066 trial [nivolumab + placebo vs. dacarbazine + placebo in treatment-naive metastatic melanoma patients without BRAF mutation[32]]. The choice for these trials is based on the size of the trials and the maturity of the data. The follow-up of the CA184-024 trial and the CheckMate 066 trial were five and three years, respectively. As the last two datasets were not publicly available, we extracted the data using image digitization (see Methods). As a reference for the *in silico* trials, we visualized the Kaplan-Meier estimators of these datasets (Figure 2A). Both trials were digitized correctly, as reflected by the nearly identical risk tables compared to the original manuscripts[31, 32]. Next, we fitted our trial simulation model on the NCCTG dataset and the control arms of the CA184-024 and CheckMate 066 trials (Figure 2B; black lines). Given the limited response rates of dacarbazine for metastatic melanoma (15%), the patients in the control arm were

regarded as untreated, and the model was fitted as such. For simplicity, we did not simulate dropout or censoring in the trials shown in this paper, although it could be added to the simulation. On average, the simulations capture the survival kinetics of the trials accurately, which is reflected by similar median overall survival values and reasonably corresponding risk tables. The final step to fully resemble late-stage immunotherapy trials in a simulation setting is replicating the treatment arms of the CA184-024 and CheckMate 066 trials. These simulated patients were treated with ICI upon diagnosis. ICI increased their T cell killing rate seven-fold and prolonged their survival, leading to OS benefit in the *in silico* trial that matched the original trial. Hence, these *in silico* trials couple the disease mechanism and mechanistic treatment effect to a realistic clinical trial outcome. Interestingly, although not incorporated explicitly in our model, the trials show survival kinetics arising as a consequence of the interaction between tumor and immune cells typically seen in immunotherapy trials: a delayed curve separation and a plateau of the survival curve of the treatment arm at later stages of the trial (Figure 2B (last two columns)).

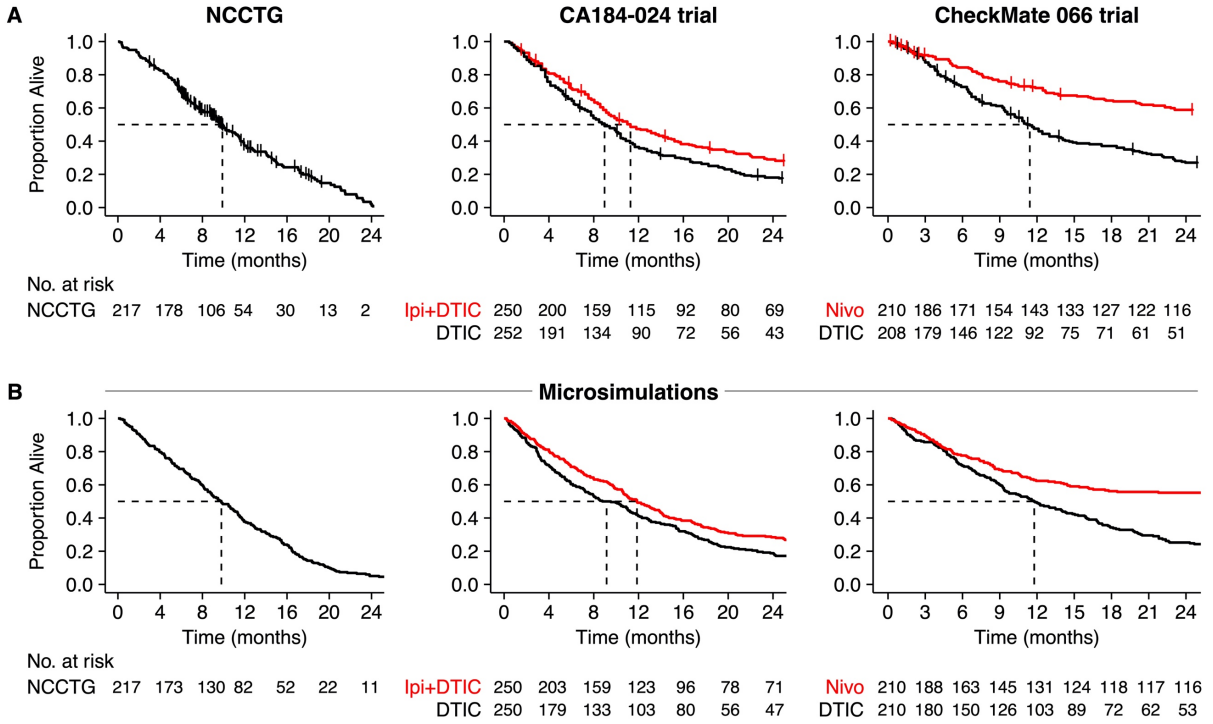

Figure 2: *In silico* cancer immunotherapy trials can generate a wide array of immunotherapy-specific survival kinetics. **(A)** Kaplan-Meier estimators of the NCCTG, CA184-024, and CheckMate 066 trials. While the NCCTG dataset is publicly available [30], the others are carefully reconstructed survival curves based on digitized data from their respective manuscripts [31, 32]. **(B)** These trial simulations can mimic representative survival kinetics as observed in actual immunotherapy trials. Specifically, typical immunotherapy-related survival kinetics – such as a delayed curve separation and a plateauing survival curve in the treatment arm – arise from these simulations as emergent behavior. Note: We matched the risk table intervals to the original manuscripts for comparative purposes.

In silico immunotherapy trials enable *a priori* prediction of trial outcomes and uncover immunotherapy-specific response patterns

The design and the success rate of any clinical trial depends, among others, on an accurate *a priori* prediction of the survival kinetics – i.e., the shape of the survival curves and the trial outcome. For late-stage immunotherapy trials, commonly observed immunotherapy-induced response patterns are a delayed curve separation and a plateauing tail of the survival curve of the treatment arm (Figure 2). These characteristic survival curve shapes reveal a violation of a vital premise at the basis of many clinical trials: the proportional hazard assumption (PHA). The PHA states that the 'instantaneous death rate' of a patient (i.e., the hazard rate) in both arms of the trial should be proportional, resulting in a constant hazard ratio. Many traditional design methods, ranging from sample size calculations to outcome analyses, depend on this theory. For late-stage immunotherapy trials, this induces two problems: (1) while a violation of the PHA needs to be addressed during trial planning, the hazard rates – and thereby the fact if the trial adheres to or violates the PHA – become available *after* analysis of the trial, and (2) if a trial does not adhere to a PHA, what will be the shape of the survival kinetics? Especially in an era where treatment and control arm regimens are becoming increasingly complex, adjusting the design and analysis methods to unknown survival kinetics is challenging.

In silico clinical trials can provide principled estimates of the shape of the survival curve, including the underlying hazard rates and hazard ratios before trial execution. The most traditional scenario would be a trial in which patients are randomized 1:1 to mono-chemotherapy or placebo. Given the direct chemotherapy effect, the PHA is generally assumed to hold for these trials. An *in silico* trial in which chemotherapy reduces the tumor growth rate to 70% for the duration of the trial indeed replicates these assumptions (Supplementary Figure 2): the survival curves separate from the start of the trial, and the hazard ratio is constant over time. However, what happens if the chemotherapy effect does not last for the entire trial but for – maybe more realistically – 6 months? Initial proportional separation of the survival curves is followed by a nearly parallel decay of both curves, leading to an early survival benefit for the chemotherapy arm (Figure 3A). Consequently, for any therapy with a non-constant treatment effect – even for chemotherapy trials – deviations from the PHA might be observed. When we switch to immunotherapy in the treatment arm, a violated PHA becomes immediately apparent. Recall that in our model, immunotherapy exerts its mechanistic effect indirectly on the tumor via an increase in the killing rate of T cells. Through approximately the first six months, the hazard rates remain constant over time, but after that, they start to decline in the immunotherapy group (red line), yielding a non-constant hazard ratio over time (Figure 3B).

The flexibility of *in silico* trials lies in their ability to incorporate complex treatment regimens. For example, let us assume one would be interested in estimating the survival curves and underlying hazard ratio over time of a chemoimmunotherapy vs. chemotherapy trial (Figure 3C), an immunotherapy + chemotherapy-placebo vs. chemotherapy vs. immunotherapy-placebo trial (Figure 3D), or a trial with induction chemotherapy followed by immunotherapy vs. immunotherapy (Figure

3E). Mechanism-based immunotherapy trials provide the means to translate biological assumptions regarding the disease and treatment effects into survival kinetics (including its hazard rate/ratio estimates). These survival kinetics, such as crossing survival curves (Figure 3D) or a temporary curve separation (Figure 3E), may be hard to predict otherwise and can be detrimental for the trial outcome if not dealt with appropriately.

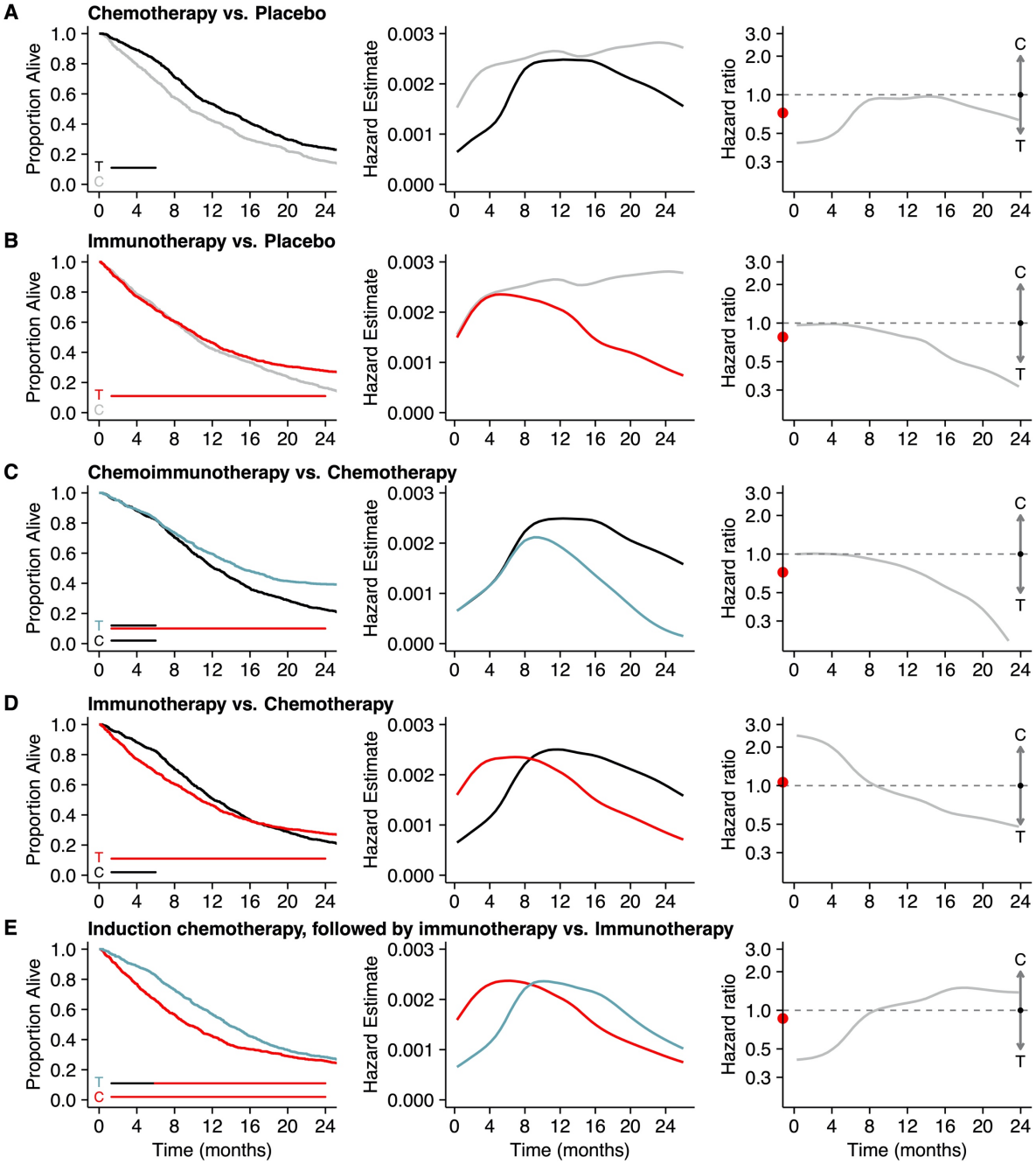

Figure 3: *In silico* clinical trials can predict immunotherapy-specific survival patterns based on biological assumptions. **(A-E)** Examples of 1:1 randomized trials with various (treatment) regimens (n=600 per arm). **(A)** A traditional chemotherapy trial (vs. placebo) only shows a proportional hazard ratio when the biological treatment effect targets the tumor directly and remains constant over time (compare to Supplementary Figure 2). **(B)** An *in silico* immunotherapy trial elicits typical immunotherapy-induced survival kinetics (i.e., delayed curve separation) and violates the proportional hazard assumption. **(C-E)** More intricate treatment or control regimens – **(C)** chemoimmunotherapy vs. chemotherapy, **(D)** immunotherapy + chemotherapy-placebo vs. chemotherapy + immunotherapy-placebo, or **(E)** induction chemotherapy followed by immunotherapy vs. immunotherapy – induce more complex survival kinetics, including **(D)** crossing survival curves or **(E)** only a temporary separation of the survival curves. The horizontal bars in column 1 indicate the duration of the treatment effect

(T = treatment, C = control). The red dot in column three indicates the overall hazard ratio. For an accurate prediction of the hazard rates and hazard ratios, saturated survival curves (n=100.000 patients per arm) were used, and the data was smoothed before plotting.

Ignoring immunotherapy-specific response patterns might cause an overestimation of an immunotherapy trial's power

To investigate the consequences of violating the PHA on the power of a clinical trial, we compared the power calculated using a PHA-dependent method (the Log Rank test) with a non-PHA-dependent method (Pearson's Chi-squared test) for different clinical scenarios. An essential difference between both methods is that the Log Rank test considers the entire survival curve, while Pearson's Chi-squared test only compares the number of events in both arms at 24 months. Logically, in a scenario that approximates the PHA the closest (such as a chemotherapy vs. placebo trial), a PHA-dependent method is superior (Figure 4A). However, a traditional immunotherapy trial violates the PHA, leading to a vast underestimation of the power of the trial when PHA-dependent methods are used for its planning (Figure 3B/4B). Given similar survival characteristics, this also holds for chemoimmunotherapy vs. chemotherapy trials (Figure 3C/4C). When survival curves cross during the trial, as seen in the immunotherapy vs. chemotherapy trial (Figure 3D), the PHA does not hold, and there is only a small survival benefit at 24 months for one of the arms of the trial. In this case, both methods predict the need for a large sample size to reach adequate power (Figure 4D). This insight indicates that it may be more rational to lengthen the trial's follow-up (i.e., a different endpoint) instead of increasing the number of participants; a finding that – on top of the absolute power – requires insight in the shape of the survival kinetics. As a general finding, ignoring immune-specific response patterns might induce an overestimation of a trial's current power, while in reality, a larger sample size (or a different endpoint) is necessary to establish the desired outcome.

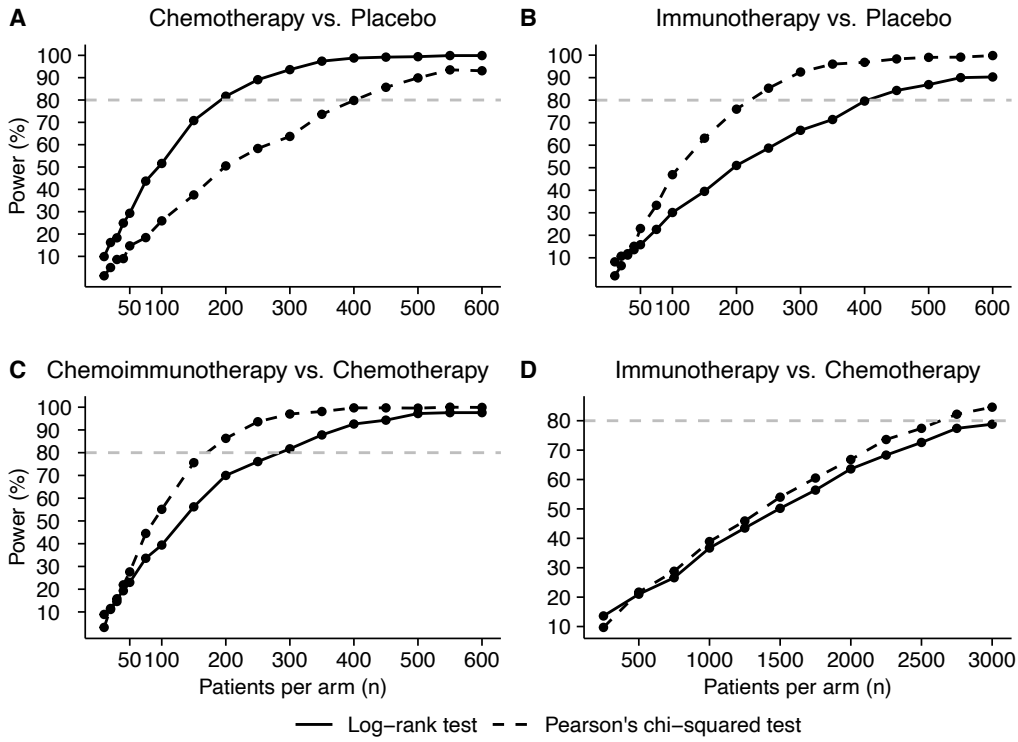

Figure 4: A mismatch between survival kinetics and analysis methods induces an underestimation of the power of an immunotherapy trial.
A power comparison using a PHA-dependent method (Log-rank test) or a PHA-independent method (Pearson's Chi-squared test) was made in different clinical scenarios: **(A)** a chemotherapy vs. placebo trial, **(B)** an immunotherapy vs. placebo trial, **(C)** a chemoimmunotherapy vs. chemotherapy trial, and **(D)** an immunotherapy vs. chemotherapy trial. Neglecting immunotherapy-specific response patterns while setting up a novel trial leads to a significant underestimation of the number of patients needed to show the efficacy of a treatment, thereby overestimating the power of the trial and reducing the probability of success.

In silico trials can validate endpoints and randomization ratios before trial execution

Clearly, the success rate of novel immunotherapy trials depends on more than its sample size alone. To establish an OS benefit of the treatment arm, it is crucial to analyze the trial once the data have reached a certain maturity – i.e., the treatment needs to be granted sufficient time to induce a survival benefit. We assumed that a delayed curve separation in immunotherapy trials would prolong the follow-up needed to establish an OS benefit of immunotherapy and thereby defer reaching maturity of the trial data. If the therapy is effective, data maturity can be regarded as the time point when a treatment effect can be observed. Hence, an optimal trial endpoint would be the earliest time at which this treatment effect can be detected with sufficient power. Therefore, we analyzed the power of differently sized trials with respect to their OS endpoint. Herein, we distinguished trials that were subject or were not subject to a delayed curve separation (immunotherapy and chemotherapy, respectively). In a classic chemotherapy trial, the treatment effect translates directly to a survival benefit in the treatment arm – the survival curves separate from the start. Therefore, the highest power will be obtained after the total duration of the treatment effect (Figure 5A, panel 1). In this case, the treatment effect lasts for six months, leading to the 6-months OS as the endpoint with the highest power. The delayed curve separation in immunotherapy trials renders it futile to analyze OS data early

on in the trial (Figure 5B, panel 1). A practical ramification is that in the presence of a delayed curve separation, the trial requires a sufficiently long follow-up and an adequate size to gain power and detect immunotherapy-specific treatment effects. Mechanism- and simulation-based power calculations with *in silico* trials can consider these specific kinetics when determining the sample size for upcoming trials.

Given the observation that both the size of an immunotherapy trial and its endpoint heavily influence the probability of finding the survival benefit of interest, we presumed that increasing the size of the treatment arm – i.e., an unequal randomization scheme – would similarly affect the power. Instead of varying the study size, we now varied the randomization ratio (second panel of Figure 5A/B). Interestingly, while the power logically depended on the OS endpoint, the randomization ratio did not greatly affect the power (Figure 5B). Considering that an unequal treatment allocation may provide ethical benefits, we confirm that the randomization ratio in immunotherapy trials is of secondary importance compared to its size or primary OS endpoint. In summary, our *in silico* immunotherapy trials replicate existing insights from trial design on how violation of the PHA affects power and analysis choices. Our ability to directly translate biological assumptions on treatment mechanisms into survival kinetics allows the researcher to reason in a principled manner about whether such violations of the PHA would or would not be expected in their specific trial design and how the problem could be addressed if it arises.

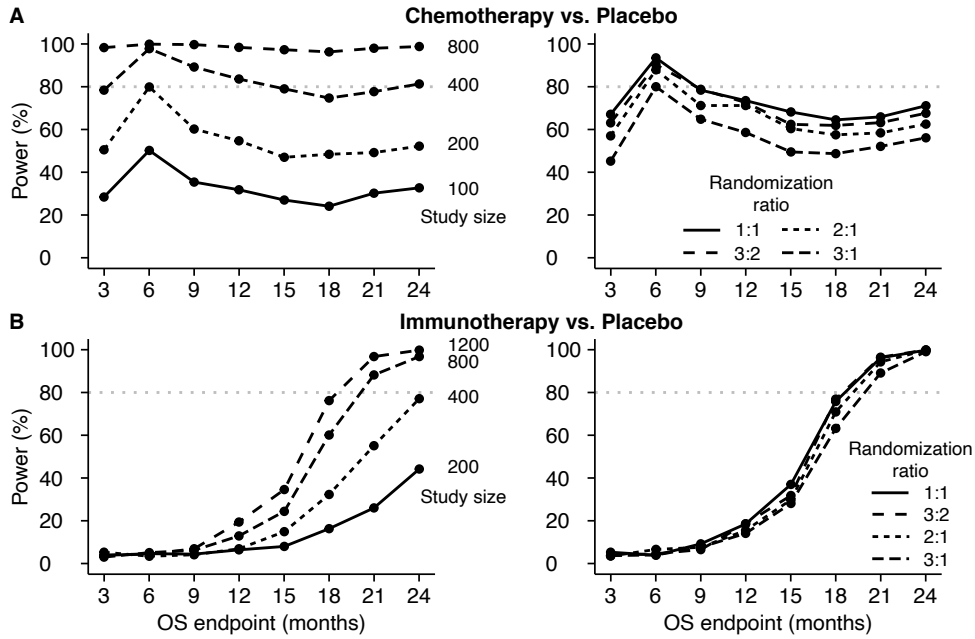

Figure 5: *In silico* trials guide decisions on OS endpoints and randomization ratios of upcoming immunotherapy trials. **(A-B)** *In silico* trials can be used to find the optimal endpoint (panel 1) or randomization ratio (panel 2) of novel trials. (A) Since the survival curves in classical chemotherapy trials separate from the trial onset, the highest power – and most optimal endpoint – is obtained at the end of the treatment interval (i.e., after six months in this example; see Figure 3A). Although less influential, a similar observation can be made for randomization ratios (study size panel 2: 300 patients). **(B)** Delayed curve separation in immunotherapy trials emphasizes that a premature final analysis of the primary OS endpoint is detrimental to the trial outcome. These trials permit validating the pre-specified survival outcomes of novel trials *a priori*. Commonly selected randomization ratios do not seem to be heavily influenced by immunotherapy-specific response patterns (study size panel 2: 1200 patient). Trial characteristics are similar to Figure 3A/B.

435

436 **Validating interim analyses to balance patient benefit and trial resources**

437 We have observed a clear tradeoff between the power of an immunotherapy trial on the one hand,
438 and the primary OS endpoint, and correspondingly the data maturity, on the other. Luckily, the two
439 are not entirely mutually exclusive: interim analyses have been developed for ethical purposes to
440 establish positive or harmful treatment effects early. However, there is a catch: the necessity to control
441 for multiple testing at each interim analysis lowers the significance threshold on the final analysis to
442 maintain the same overall type I error rate. This begs the question: "How many interim analyses should
443 you plan, and when should you plan them?" Again, principled answers to such questions can be
444 obtained with the help of *in silico* immunotherapy trials. To illustrate this, we simulated 1000
445 immunotherapy trials with 1200 patients per trial, randomized 1:1 over immunotherapy with a strong
446 treatment effect or a placebo (Figure 6A). In the absence of interim analyses, the vast majority of the
447 trials are predicted to end up positive. Adding interim analyses (O'Brien-Flemming approach) to the
448 equation induces a tradeoff. On the one hand, increasing the number of equally-spaced interim
449 analyses increases the probability of early detecting a positive treatment effect (e.g., approximately
450 60% of the trials are positive after 18 months in the case of three interim analyses; Figure 6A). On the
451 other hand, the overall probability of ending up with a negative trial due to more stringent analyses
452 (i.e., less power) also increases, especially in the case of immunotherapies with a weaker treatment
453 effect (±57% without an interim analysis vs. ±63% with three interim analyses; Figure 6B). In an actual
454 trial, the latter needs to be corrected by including additional patients to maintain the pre-planned
455 power. Furthermore, we observe that the timing of the interim analysis is crucial. Whereas an interim
456 analysis at 18 months provides additional value to the trial, interim analyses before 16 months are
457 predicted to be wasteful: both due to the presence of non-proportional hazards and less mature data.
458 As a control, we simulated trials without any treatment effect. By design, approximately 95% of the
459 trials should end up negative irrespective of the number of interim analyses, which seemed to be the
460 case (Figure 6C). Logically, the weaker the treatment effect, the higher the probability of erroneously
461 finding a harmful treatment effect – a characteristic that the simulation also exhibits (Figure 6B/C).

462

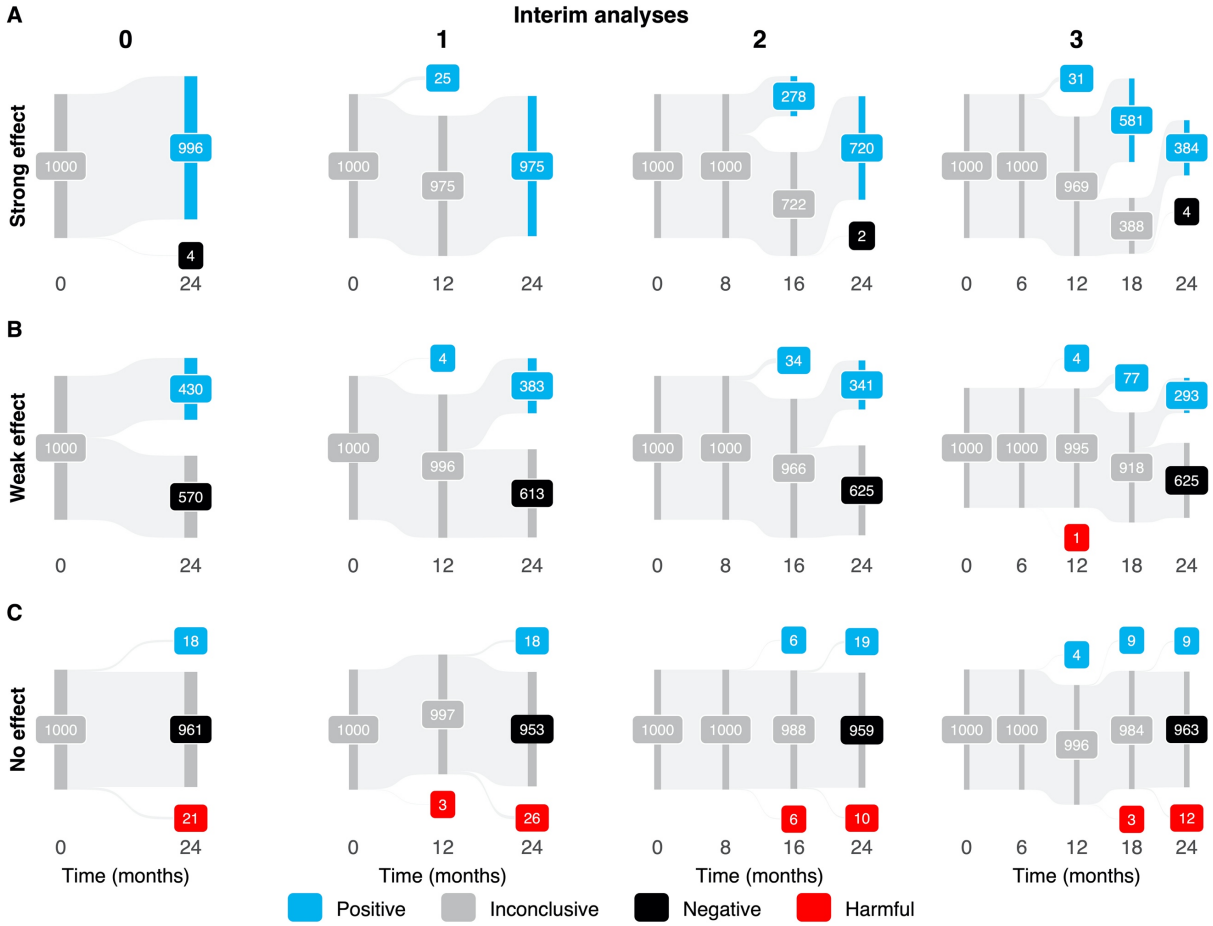

Figure 6: *A priori* validation of the interim analysis plan steers the selection of the optimal number and timing of the analyses, avoids futile or even harmful analyses, and might optimize the use of trial resources.
(A) In the case of immunotherapy with a potent effect, *in silico* trials help develop a rationale for the timing of the interim analyses. While an interim analysis at 12 months might not add value to the trial, analyses after 16 and 18 months, respectively, have a probability of approximately 28% and 60% to lead to early stopping with a positive result. **(B)** Multiple interim analyses can reduce the probability of confirming the desired treatment effect in case of a weak immunotherapy effect. **(C)** In the absence of any treatment effect (a control scenario), the number of interim analyses does not heavily influence the trial outcome. Each trial simulation contains 1200 patients (randomization ratio 1:1) to ensure adequate power of the trial. Trials are analyzed with a proportions test (Pearson's chi-squared test). Treatment effect (amplification factor of T cell killing rate): strong = 7, weak = 3, no effect = 0 (see Methods).

DISCUSSION

In this study, we used mechanism-based *in silico* cancer immunotherapy trials to predict survival kinetics and response profiles of novel immunotherapy trials. Complementary to conventional design methods, *in silico* trials provide the ability to investigate the implications of a researcher's biological (as opposed to statistical) hypotheses of a drug's mechanism of action for the design, conduct, analysis, and outcome of clinical trials. When comparing the simulated outcomes to actual immunotherapy trial outcomes, we showed that *in silico* trials are suited to translate complex biological mechanisms (such as observed during the treatment of patients with ICI) into realistic trial outcomes. Crucially, the survival kinetics that arose from these mechanism-based simulations reflected two pivotal components often found in immunotherapy trials: a delayed curve separation and a plateauing tail of

the survival curve at later stages of the trial. In line with genuine immunotherapy trials, we find that these immunotherapy-specific response patterns differ considerably from chemotherapy-based kinetics. Our findings confirm that diversity in survival kinetics profoundly impacts the outcomes of immunotherapy trials[33]. Consequently, these features need to be considered when deciding on the sample size, endpoint, randomization ratio, and the number and timing of interim analyses of a novel immunotherapy trial.

Over the past two decades, *in silico* clinical trials are gaining in popularity. These trials enable investigating, among others, how novel drugs, treatment schedules, dosing regimens, and interpatient heterogeneity affect the outcome of a clinical trial[34]. *In silico* clinical studies have a wide range of applicability from pediatric infectious[35] and orphan diseases[36] to diabetes[37], inflammatory autoimmune diseases[38], traumatic injury[39], psychiatric illness[40], and cancer. In oncology, several *in silico* clinical trials involving chemotherapy and tyrosine kinase inhibitors have been performed[41, 42]. Moreover, with the onset of checkpoint inhibitors, *in silico* immunotherapy trials have gained interest, leading to trials with anti-CTLA-4-antibodies and anti-PD-(L)1 antibodies[43, 44, 45]. The common denominator in these trials is that they primarily center around the therapies' dosing regimens and treatment schedules. Herein lies the main difference with our simulation approach: although the 'key ingredients' of these approaches are similar – they are based on a mathematical abstraction of a disease mechanism – our trials do not aim to optimize treatment schedules. Instead, we complement traditional design methodology by adding the means to predict trial outcomes and elucidate trial kinetics *a priori* to steer design decisions of novel immunotherapy trials. These trials differ from traditional trial design research in that these, often statistically-grounded, approaches simulate clinical trials based on population-level assumptions (e.g., with particular distributions of survival times, study durations, or with a specific censoring mechanism). Examples of these high-level simulation approaches include, but are certainly not limited to, studies aiming to calculate the sample size and power of clinical trials[46, 47, 48]. Since these methods lack a direct link to the underlying biological disease mechanism, interpreting their parameters for individual trial participants is difficult or even impossible. In contrast, *in silico* trials are founded on biological assumptions but then translate these assumptions into statistical concepts such as hazard ratio kinetics. In this manner, simulated trials encourage an interdisciplinary discussion about the design of an upcoming trial.

In silico clinical trials are applicable in several settings. First, they provide the means to verify clinical trial and treatment assumptions before investing extensive amounts of work and funds into the development and execution of a clinical trial and can, thereby, function as a proof of principle of the soundness of the hypotheses for an upcoming trial. Scrutinizing each aspect of the trial supports optimal design decisions and might reduce unanticipated outcomes. Moreover, this mechanism-based approach does not necessitate a deep understanding of complex mathematical theorems; instead, it

requires a biological understanding of a disease. This mechanistic basis is intuitive, which benefits the communication between clinical doctors and biomedical researchers on the one hand and statisticians and clinical trialists on the other. Additionally, *in silico* trials might serve as excellent educational tools. The ability to simulate a wide range – from basic to highly advanced – research questions can be exploited in teaching activities for entry-level clinicians to experienced trialists. A final implication, which holds for any trial simulation, is that they may provide insight when conventional clinical trials are unfeasible due to practical or ethical constraints (e.g., clinical trials in rare diseases, pediatrics, or critical care medicine).

Nonetheless, *in silico* clinical trials have to be considered in light of some limitations. The most critical limitation is universal to *any* – either *in vitro*, *in vivo*, or computational – scientific model: the immunotherapy trial outcomes depend heavily (if not entirely) on the biological assumptions of the model, meaning that incorrect interactions or erroneous parametrization of the model might induce inaccurate outcomes. The parameterization, in particular, might pose a problem: given the often novel treatment mechanisms, data to fine-tune the parameters of the model accurately might be scarce. In these cases, the simulation itself can be used as a sensitivity analysis to assess to what extent a certain parameter range influences the robustness of the predictions. In addition, while the model itself is intuitive to understand, translating biological principles into an ODE model and implementing it into a simulation requires thorough knowledge of computational methods, limiting its widespread applicability.

In summary, *in silico* cancer immunotherapy trials offer a versatile approach to simulate immunotherapy trials based on biological assumptions. Furthermore, as a simulation tool, they facilitate the verification of trial design decisions to optimize the probability of a successful immunotherapy trial and contribute to high-quality research for cancer patients.

REFERENCES

1. Upadhaya S, Hubbard-Lucey VM, Yu JX. Immuno-oncology drug development forges on despite COVID-19. *Nat Rev Drug Discov*, (2020).

2. Dowden H, Munro J. Trends in clinical success rates and therapeutic focus. *Nat Rev Drug Discov* **18**, 495-496 (2019).

3. Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. *Biostatistics* **20**, 273-286 (2019).

4. Thomas D*, et al.* Clinical Development Success Rates and Contributing Factors 2011-2020.). BIO, Informa Pharma Intelligence, QLS Advisors (2021).

5. Hwang TJ, Carpenter D, Lauffenburger JC, Wang B, Franklin JM, Kesselheim AS. Failure of Investigational Drugs in Late-Stage Clinical Development and Publication of Trial Results. *JAMA Intern Med* **176**, 1826-1833 (2016).

6. de Miguel M, Calvo E. Clinical Challenges of Immune Checkpoint Inhibitors. *Cancer Cell* **38**, 326-333 (2020).

7. Li A, Bergan RC. Clinical trial design: Past, present, and future in the context of big data and precision medicine. *Cancer* **126**, 4838-4846 (2020).

8. Cousin S, Seneschal J, Italiano A. Toxicity profiles of immunotherapy. *Pharmacol Ther* **181**, 91-100 (2018).

9. Seymour L*, et al.* iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol* **18**, e143-e152 (2017).

10. Hoos A*, et al.* Improved endpoints for cancer immunotherapy trials. *J Natl Cancer Inst* **102**, 1388-1397 (2010).

11. Chen TT. Statistical issues and challenges in immuno-oncology. *J Immunother Cancer* **1**, 18 (2013).

12. Mick R, Chen TT. Statistical Challenges in the Design of Late-Stage Cancer Immunotherapy Studies. *Cancer Immunol Res* **3**, 1292-1298 (2015).

13. Rahman R*, et al.* Deviation from the Proportional Hazards Assumption in Randomized Phase 3 Clinical Trials in Oncology: Prevalence, Associated Factors, and Implications. *Clin Cancer Res* **25**, 6339-6345 (2019).

14. Hodi FS*, et al.* Immune-Modified Response Evaluation Criteria In Solid Tumors (imRECIST): Refining Guidelines to Assess the Clinical Benefit of Cancer Immunotherapy. *J Clin Oncol* **36**, 850-858 (2018).

15. Wolchok JD*, et al.* Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clin Cancer Res* **15**, 7412-7420 (2009).

16. Anagnostou V*, et al.* Immuno-oncology Trial Endpoints: Capturing Clinically Meaningful Activity. *Clin Cancer Res* **23**, 4959-4969 (2017).

17. Mushti SL, Mulkey F, Sridhara R. Evaluation of Overall Response Rate and Progression-Free Survival as Potential Surrogate Endpoints for Overall Survival in Immunotherapy Trials. *Clin Cancer Res* **24**, 2268-2275 (2018).

18. Chen TT. Milestone Survival: A Potential Intermediate Endpoint for Immune Checkpoint Inhibitors. *J Natl Cancer Inst* **107**, (2015).

19. Kaufman HL*, et al.* Durable response rate as an endpoint in cancer immunotherapy: insights from oncolytic virus clinical trials. *J Immunother Cancer* **5**, 72 (2017).

20. Chan TA*, et al.* Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol* **30**, 44-56 (2019).

21. Dudley JC, Lin MT, Le DT, Eshleman JR. Microsatellite Instability as a Biomarker for PD-1 Blockade. *Clin Cancer Res* **22**, 813-820 (2016).

22. Patel SP, Kurzrock R. PD-L1 Expression as a Predictive Biomarker in Cancer Immunotherapy. *Mol Cancer Ther* **14**, 847-856 (2015).

23. Sha D, Jin ZH, Budczies J, Kluck K, Stenzinger A, Sinicrope FA. Tumor Mutational Burden as a Predictive Biomarker in Solid Tumors. *Cancer Discovery* **10**, 1808-1825 (2020).

24. Chen TT. Designing Late-Stage Randomized Clinical Trials with Cancer Immunotherapy: Can We Make It Simpler? *Cancer Immunol Res* **6**, 250-254 (2018).

25. Royston P, Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol* **13**, 152 (2013).

26. Xu Z, Zhen B, Park Y, Zhu B. Designing therapeutic cancer vaccine trials with delayed treatment effect. *Stat Med* **36**, 592-605 (2017).

27. Creemers JHA*, et al.* A tipping point in cancer-immune dynamics leads to divergent immunotherapy responses and hampers biomarker discovery. *J Immunother Cancer* **9**, (2021).

28. Borghans JA, de Boer RJ, Segel LA. Extending the quasi-steady state approximation by changing variables. *Bull Math Biol* **58**, 43-63 (1996).

29. Gadhamsetty S, Maree AF, Beltman JB, de Boer RJ. A general functional response of cytotoxic T lymphocyte-mediated killing of target cells. *Biophys J* **106**, 1780-1791 (2014).

30. Loprinzi CL*, et al.* Prospective evaluation of prognostic variables from patient-completed questionnaires. North Central Cancer Treatment Group. *J Clin Oncol* **12**, 601-607 (1994).

31. Maio M*, et al.* Five-year survival rates for treatment-naive patients with advanced melanoma who received ipilimumab plus dacarbazine in a phase III trial. *J Clin Oncol* **33**, 1191-1196 (2015).

32. Ascierto PA*, et al.* Survival Outcomes in Patients With Previously Untreated BRAF Wild-Type Advanced Melanoma Treated With Nivolumab Therapy: Three-Year Follow-up of a Randomized Phase 3 Trial. *JAMA Oncol* **5**, 187-194 (2019).

33. Chen TT. Predicting analysis times in randomized clinical trials with cancer immunotherapy. *BMC Med Res Methodol* **16**, 12 (2016).

34. Alfonso S, Jenner AL, Craig M. Translational approaches to treating dynamical diseases through in silico clinical trials. *Chaos* **30**, 123128 (2020).

35. Valitalo PA*, et al.* Novel model-based dosing guidelines for gentamicin and tobramycin in preterm and term neonates. *J Antimicrob Chemother* **70**, 2074-2077 (2015).

36. Carlier A, Vasilevich A, Marechal M, de Boer J, Geris L. In silico clinical trials for pediatric orphan diseases. *Sci Rep* **8**, 2465 (2018).

37. Klinke DJ, 2nd. Integrating epidemiological data into a mechanistic model of type 2 diabetes: validating the prevalence of virtual patients. *Ann Biomed Eng* **36**, 321-334 (2008).

38. Schmidt BJ, Casey FP, Paterson T, Chan JR. Alternate virtual populations elucidate the type I interferon signature predictive of the response to rituximab in rheumatoid arthritis. *BMC Bioinformatics* **14**, 221 (2013).

39. Brown D*, et al.* Trauma in silico: Individual-specific mathematical models and virtual clinical populations. *Sci Transl Med* **7**, 285ra261 (2015).

40. Magnusson MO*, et al.* Dosing and Switching Strategies for Paliperidone Palmitate 3-Month Formulation in Patients with Schizophrenia Based on Population Pharmacokinetic Modeling and Simulation, and Clinical Trial Data. *CNS Drugs* **31**, 273-288 (2017).

41. Perez-Garcia VM*, et al.* Computational design of improved standardized chemotherapy protocols for grade II oligodendrogliomas. *PLoS Comput Biol* **15**, e1006778 (2019).

42. Fassoni AC, Baldow C, Roeder I, Glauche I. Reduced tyrosine kinase inhibitor dose is predicted to be as effective as standard dose in chronic myeloid leukemia: a simulation study based on phase III trial data. *Haematologica* **103**, 1825-1834 (2018).

43. Jafarnejad M*, et al.* A Computational Model of Neoadjuvant PD-1 Inhibition in Non-Small Cell Lung Cancer. *AAPS J* **21**, 79 (2019).

44. Milberg O*, et al.* A QSP Model for Predicting Clinical Responses to Monotherapy, Combination and Sequential Therapy Following CTLA-4, PD-1, and PD-L1 Checkpoint Blockade. *Sci Rep* **9**, 11286 (2019).

45. Wang H*, et al.* In silico simulation of a clinical trial with anti-CTLA-4 and anti-PD-L1 immunotherapies in metastatic breast cancer using a systems pharmacology model. *R Soc Open Sci* **6**, 190366 (2019).

46. Bang H, Jung SH, George SL. Sample size calculation for simulation-based multiple-testing procedures. *J Biopharm Stat* **15**, 957-967 (2005).

47. Doostfatemeh M, Taghi Ayatollah SM, Jafari P. Power and Sample Size Calculations in Clinical Trials with Patient-Reported Outcomes under Equal and Unequal Group Sizes Based on Graded Response Model: A Simulation Study. *Value Health* **19**, 639-647 (2016).

48. Wilson DT, Hooper R, Brown J, Farrin AJ, Walwyn RE. Efficient and flexible simulation-based sample size determination for clinical trials with multiple design parameters. *Stat Methods Med Res* **30**, 799-815 (2021).

LIST OF ABBREVIATIONS

ICI: immune checkpoint inhibitor; NCCTG: North Central Cancer Treatment Group; ODE: ordinary differential equation; OS: overall survival; PHA: proportional hazard assumption.

DECLARATIONS

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The code is available at GitHub: https://github.com/jeroencreemers/in-silico-clinical-trials

Competing interests

Not applicable

Funding

JC was funded by the Radboudumc. CF received an ERC Adv Grant ARTimmune (834618) and an NWO Spinoza grant. IV received an NWO-Vici grant (918.14.655). JT was supported by a Young Investigator Grant (10620) from the Dutch Cancer Society and an NWO grant (VI.Vidi.192.084).

Authors' contribution

JHAC and JT conceived this study. JHAC performed the experiments and wrote the manuscript under the supervision of JT. All authors provided feedback on the manuscript and reviewed the manuscript prior to submission.

Acknowledgments

Not applicable.