Citation: CPT Pharmacometrics Syst. Pharmacol. (2018) 7, 219–227; doi:10.1002/psp4.12276

ORIGINAL ARTICLE

The Virtual Anemia Trial: An Assessment of Model-Based *In Silico* Clinical Trials of Anemia Treatment Algorithms in Patients With Hemodialysis

Doris H. Fuertinger[1,2]*, Alice Topping[1], Franz Kappel[3], Stephan Thijssen[1] and Peter Kotanko[1,4]

***In silico* approaches have been proposed as a novel strategy to increase the repertoire of clinical trial designs. Realistic simulations of clinical trials can provide valuable information regarding safety and limitations of treatment protocols and have been shown to assist in the cost-effective planning of clinical studies. In this report, we present a blueprint for the stepwise integration of internal, external, and ecological validity considerations in virtual clinical trials (VCTs). We exemplify this approach in the context of a model-based *in silico* clinical trial aimed at anemia treatment in patients undergoing hemodialysis (HD). Hemoglobin levels and subsequent anemia treatment were simulated on a per patient level over the course of a year and compared to real-life clinical data of 79,426 patients undergoing HD. The novel strategies presented here, aimed to improve external and ecological validity of a VCT, significantly increased the predictive power of the discussed in *silico trial*.**

CPT Pharmacometrics Syst. Pharmacol. (2018) **7**, 219–227; doi:10.1002/psp4.12276; published online 31 January 2018.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ *In silico* clinical trials have been proposed over the last years as a novel strategy to increase the repertoire of trial designs and have been acknowledged as a useful tool by the scientific community, industrial research and development departments, as well as by regulatory agencies.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ The design of empirical clinical trials is guided by the principles of internal, external, and ecological validity. We explored different ways to set up *in silico* clinical trials that specifically address these criteria and quantified associated improvements in prediction quality.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ This study outlines a methodology to integrate internal, external, and ecological validity considerations in VCT designs. It further demonstrates the feasibility of VCTs in patients undergoing HD, a population plagued by excessive morbidity and mortality and high costs. Second, we have shown the importance of adding components that reflect the ecosystem of care (the stochastic modules) to VCTs. Third, we have shown that personalized avatars are superior to models that utilize Monte Carlo simulations; this is an important insight that may help to guide future research.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

☑ We believe that the presented results have the potential to serve as a blueprint for future trial designs. Of note, our setup enables VCTs of a range of ESAs without the immediate need for clinical trials, thus supporting the development of novel ESAs. Our large avatar population may prove very useful for the development of anemia treatment algorithms in clinically challenging HD subpopulations.

Randomized controlled trials (RCTs) are assigned the highest level of evidence for therapeutic studies, and it is beyond doubt that RCTs have contributed immensely to the medical progress and have advanced patient care. By randomly allocating subjects to two or more treatment groups, RCTs randomize confounding factors. Consequently, a well-designed and properly conducted RCT will give unbiased results and have little risk of systematic errors (i.e., have a high internal validity). However, RCTs may face weaknesses that limit their generalizability, because RCT participants may not be representative of the wider population of interest (i.e., have poor external validity). In addition, RCT results must also generalize to the real-life settings in which the trial results will later be applied (i.e., should have a high ecological validity). Other frequently noted shortcomings are the need to recruit a sufficiently large number of patients to conduct a properly powered study, associated high costs, the need to establish a sophisticated trial infrastructure, and the long duration from study inception to completion. With multiple treatments in the pipeline, pharmaceutical companies and academic institutions compete for a limited pool of patients. In oncology, it has been

[1]Renal Research Institute, New York, New York, USA; [2]Fresenius Medical Care Deutschland GmbH, Bad Homburg, Germany; [3]Institute for Mathematics and Scientific Computing, Karl-Franzens University, Graz, Austria; [4]Icahn School of Medicine at Mount Sinai, New York, New York, USA. *Correspondence: D H Fuertinger (doris.fuertinger@fmc-ag.com)
Received 9 October 2017; accepted 28 December 2017; published online on 31 January 2018. doi:10.1002/psp4.12276

estimated that only 20% of patients are eligible for clinical trials, because many patients are excluded due to poor performance status or inability to meet specific eligibility requirements. This limitation has been illustrated by a study showing that filling all pancreatic-cancer trials in the United States in 2011 would have required the participation of 83% of patients with resectable pancreatic ductal adenocarcinoma,[1] yet, only 5% of patients volunteer for trials.[2] Other areas of medicine face comparable challenges: in a recent hemodialysis (HD) RCT sponsored by the National Institutes of Health, of 6,276 screened patients only 245 (3.9%) were eventually randomized.[3]

To mitigate and overcome this challenges, alternative trial strategies have been developed, such as adaptive design clinical trials,[4] and, in 2010, the US Food and Drug Administration released an adaptive trial design guidance document for the industry. Despite these innovations, there is a continued need to advance the field. Specifically, *in silico* (i.e., performed entirely on a computer) trials have been proposed as a way to increase the repertoire of trial designs.[5–8] Such virtual trials have been acknowledged as a useful tool by the scientific community[9–15] as well as by regulatory agencies[16] and questions of the design of clinical trial simulations have been previously addressed by multiple authors (see refs. 17,18).

RCTs are designed with the intent to provide internal, external, and ecological validity. In this report, we demonstrate step-by-step how these validity considerations can be integrated in the design of virtual clinical trials (VCTs). We exemplify this approach in the context of a model-based *in silico* clinical trial aimed at anemia treatment in patients with chronic HD, the Virtual Anemia Trial (VIAT). Specifically, we pursue a stepwise strategy that subsequently addresses the topics of internal, external, and ecological validity to improve the predictive power of the VIAT.

Physiology-based mechanistic models are the foundation of many predictive biosimulations. The core of the VIAT is a comprehensive physiology-based mathematical model describing the development of red blood cells (RBCs; erythropoiesis) and the effect of erythropoiesis stimulating agents (ESAs) on this process.[19] One strength of well-designed and validated physiology-based mathematical models is their intrinsic internal validity as the causal relation between an intervention and the corresponding outcome, is clearly defined and easily comprehensible. Further, by definition, a deterministic model design, as the one used in the VIAT, generates reproducible results so that a specific intervention always results in the same outcome.

To date, most *in silico* trials utilize Monte Carlo type sampling (and resampling) techniques to create virtual patient populations.[20–26] Although conceptually attractive, Monte Carlo simulations fall short in representing the (patho)physiology of an actual individual subject, because, in general, only population-derived estimates of patient characteristics are represented. We overcame this limitation by applying advanced mathematical and computational techniques to create a large population of *in silico* representations ("avatars") of real patients undergoing HD receiving ESA treatment for anemia. Hence, a cornerstone of our VIAT is the integration of individual real-life clinical patient data in the modeling process to improve external validity. To that end, we randomly sampled almost 7,000 patients undergoing HD from a nationally representative US HD population comprising over 37,000 individuals. In the next step, we created one avatar for each sampled patient. Finally, stochastic "clinic modules" (informed by real-life operational data) were designed to create an *in silico* test environment that reflects operational processes and challenges in dialysis clinics. These modules included information on laboratory schedules and processing times, patient nonadherence, and hospitalizations, among others. By integrating such clinic modules into the *in silico* trial simulations, we enhanced the ecological validity. Finally, we present results on a VIAT conducted in these avatars utilizing the clinic modules with the eventual goal to improve anemia therapy for real patients.

METHODS

A comprehensive physiology-based mathematical model of erythropoiesis was used as the basis for *in silico* simulations of an anemia treatment protocol. Different approaches to conduct a VCT were tested and compared with retrospective clinical data. The hemoglobin (Hgb) levels and the subsequent anemia treatment were simulated on a patient level for the course of an entire year in all *in silico* trials. Results of VCTs were compared to data from 79,426 patients undergoing HD in Fresenius Kidney Care (FKC) clinics between September 2015 and August 2016. In these clinics, use of the tested anemia treatment protocol is part of standard care.

Population and data eligibility

Patients were included in the analysis if they were above 18 years of age, received at least one administration of methoxy polyethylene glycol-epoetin beta (Mircera, Roche, Basel, Switzerland) between September 2015 and August 2016, and had clinical routine laboratory data for at least 4 months during the considered time period. Patients were censored when they received any other ESA after their first administration of Mircera. Further, patients with International Classification of Disease, 9th revision codes in their electronic health record that indicated an increased risk for bleedings (e.g., patients with gastric ulcers, esophageal varices, colon polyps, etc.) were excluded.

Virtual clinical trial

An anemia treatment protocol was implemented *in silico*, as used in a large number of FKC clinics, in which the ESA was administered by the caregivers in the dialysis facility. Per protocol, ESA dosing schedules followed biweekly or monthly administration patterns, in general. The Hgb level was measured weekly and ESA doses were adjusted monthly or, under certain circumstances, biweekly. The Hgb laboratory schedules and dose calculations were implemented *in silico*, as stated in the anemia treatment protocol. Clinical modules were predefined based on knowledge of the operational process and medical requirements and designed to reflect important parts of the clinical routine that influenced how anemia therapy was conducted. The stochastic modules were developed in a combination of "art and science" based on published literature, laboratory

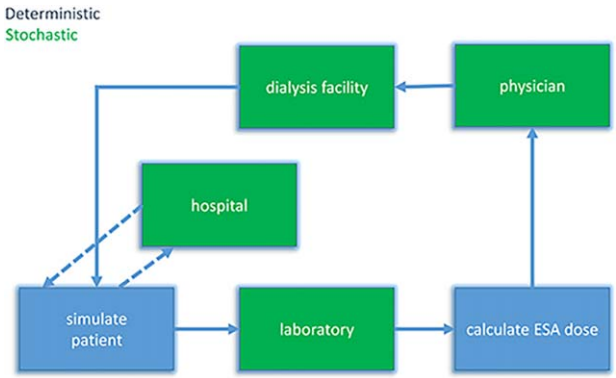

Figure 1 Setup schematic of Virtual Anemia Trial (VIAT) 3.0. Blue boxes indicate "deterministic" modules: the simulation of hemoglobin values for each Avatar using the physiology based mathematical model and the calculation of a new dose following the anemia therapy protocol. Green boxes indicate modules of "stochastic" nature. VIAT 3.0 comprises modules simulating the (random) impact of the involved laboratory (e.g., measurement noise), physician (e.g., blood transfusion orderings), dialysis facility (e.g., nonadherence of patients to therapy), and hospital (e.g., hospital stays). ESA, erythropoiesis stimulating agent.

performance information, and clinical experience of the physicians on the team. The dialysis facility module utilized data on missed treatments from the literature[27]; current rates were confirmed with the dialysis provider. The hospital module considered published hospitalization rates from multiple sources[28]; current rates were confirmed with the dialysis provider. For the laboratory module, we used Hgb measurement accuracy and precision data provided by the laboratory. Estimates of blood sample shipping delays were obtained from the dialysis provider. The physician module was developed in close collaboration of the two physicians on our team (P.K. and S.T.) who provided estimates of physicians' reaction to Hgb levels at, for instance, the lower end of the Hgb distribution (e.g., the likelihood for ESA dose escalations outside of usual anemia algorithms in the face of very low Hgb levels). A flowchart of the final setup (Virtual Anemia Trial 3.0) is depicted in **Figure 1**. To obtain a sound quantitative understanding of the impact of individual modules and their combined use, we used multiple iterations to test a variety of module combinations (data not shown). The relative importance of the stochastic modules was laboratory (measurement variability; shipping delays) > physician (odds to overwrite recommended ESA doses and to order blood transfusions) > hospitalization (odds for a patient to be hospitalized) > dialysis facility (odds of no-shows). It turned out that all modules contribute to the improved performance, so we decided to keep all of them.

Statistical analysis

Data are presented as mean ± SD where normally distributed, as medians (interquartile range) where not normally distributed. Categorical variables are expressed as frequencies and percentages of the group from which they were derived.

Comparison of Hgb results and ESA doses were conducted primarily using measurements of central tendencies and visual depiction of Hgb and ESA distributions. Several metrics were calculated from the raw data to better understand differences in the populations based on monthly and annual data and are found in **Tables 1 and 2**. Although Hgb laboratory measurements were completed weekly for each patient, ESA doses were administered in general at 2 or 4-week cycles. However, in rare cases, ESA administration cycles can increase up to 6 or 8 weeks necessitating the creation of monthly, per treatment, and per patient-year metrics to better understand the relationship between Hgb and ESA dosing. In some cases, the monthly average of Hgb results for each patient was used to compare Hgb distributions, especially when comparing to cumulative monthly ESA doses per patient.

Three random samples of the FKC patient data (after exclusion) were established using the same number of patients as the avatar patient population ($N = 6{,}659$) in order to create effective comparisons of patient groups. A single random group was selected and simulated data from all *in silico* trials were compared with this group. Results of all randomizations and patient characteristics of the groups are found in the **Supplementary Information**.

Per HD treatment is a metric used to compare the amount of ESA used at the treatment level and was calculated from the cumulative monthly ESA dose divided by the median number of treatments in each month for each patient. Treatment information was not available for the simulated patient data and, thus, median number of treatments per month was calculated using FKC data. Two separate values were used for FKC and virtual patient populations. This was necessary as patient data was simulated for exactly 1 year without attrition in the Virtual Anemia Trials. The FKC patient population, however, did experience attrition (64% of the FKC patients contributed an entire year of data) and, thus, the sum of the time each patient contributed to the study in each group and expected number of treatments per month was not equal across groups. The average cumulative monthly dose of ESA was divided by

Table 1 Overall ESA doses and hemoglobin results

Data	No. of patients	Average ESA dose[a]	ESA dose per-HD treatment[b]	ESA dose per patient year	Average Hgb
VIAT 1.0	6,659	138.7 ± 65.4	16.1 (8.1, 32.3)	2037.2	11.4 ± 3.4
VIAT 2.0	6,659	95.9 ± 54.2	8.1 (4, 16.1)	1267.6	11 ± 1
VIAT 3.0	6,659	97.1 ± 54.8	8.1 (4, 16.1)	1263.8	10.9 ± 1.1
Clinical data	6,659	92.5 ± 51.4	8.3 (6.2, 16.5)	1335.6	10.8 ± 1.1

ESA, erythropoiesis stimulating agent; Hgb, hemoglobin; HD, hemodialysis; VIAT, Virtual Anemia Trial.
[a]Mean ± SD.
[b]Median (25th percentile, 75th percentile).

Table 2 ESA doses per category of average monthly hemoglobin

Hgb category	Data	% of patients	No. of ESA doses per month[a]	Monthly cumulative ESA dose[a]	ESA dose per HD-treatment[a]
<8.0	VIAT 1.0	8.4	2.2 ± 0.5	372.9 ± 161.5	30.1 ± 13
	VIAT 2.0	0.4	2.5 ± 0.6	494.9 ± 99.2	39.9 ± 8
	VIAT 3.0	0.6	2 ± 0.6	339.1 ± 156.6	27.3 ± 12.6
	Clinical data	1.1	1.7 ± 1.1	335.9 ± 207.2	27.8 ± 17.1
8.0–10.0	VIAT 1.0	21.5	2.1 ± 0.4	337.2 ± 149.5	27.2 ± 12.1
	VIAT 2.0	10.4	2.1 ± 0.6	279.1 ± 160.4	22.5 ± 12.9
	VIAT 3.0	12.2	1.8 ± 0.6	241.1 ± 152.5	19.4 ± 12.3
	Clinical data	18.1	1.9 ± 0.9	236 ± 167	19.5 ± 13.8
10.0–11.0	VIAT 1.0	24.1	1.6 ± 0.9	148.3 ± 120.6	12.3 ± 10
	VIAT 2.0	46.2	1.7 ± 0.7	202.5 ± 148.9	16.3 ± 12
	VIAT 3.0	43.0	1.4 ± 0.7	132.5 ± 113.1	10.7 ± 9.1
	Clinical data	39.3	1.4 ± 0.7	137.3 ± 115.8	11.1 ± 9.3
11.0–12.0	VIAT 1.0	18.5	0.9 ± 0.8	133.9 ± 93.5	10.8 ± 7.5
	VIAT 2.0	32.5	0.7 ± 0.7	96.5 ± 70.5	7.8 ± 5.7
	VIAT 3.0	32.8	0.8 ± 0.7	96.8 ± 72.4	7.8 ± 5.8
	Clinical data	32.6	0.9 ± 0.8	108.4 ± 83.8	9 ± 6.9
12.0–14.0	VIAT 1.0	14.1	0.1 ± 0.4	106.7 ± 49.9	8.6 ± 4
	VIAT 2.0	9.7	0.1 ± 0.3	101.6 ± 45.6	8.2 ± 3.7
	VIAT 3.0	10.6	0.1 ± 0.3	99.4 ± 47.6	8 ± 3.8
	Clinical data	8.6	0.3 ± 0.6	108.9 ± 77.6	9 ± 6.4
>14.0	VIAT 1.0	13.4	0 ± 0.1	99.3 ± 45.1	8 ± 3.6
	VIAT 2.0	0.9	0 ± 0.1	150 ± 0	12.1 ± 0
	VIAT 3.0	0.8	0 ± 0.2	130.4 ± 28	10.5 ± 2.3
	Clinical data	0.3	0.1 ± 0.4	125 ± 75.6	10.3 ± 6.2

ESA, erythropoiesis stimulating agents; HD, hemodialysis; Hgb, hemoglobin; VIAT, Virtual Anemia Trial.
[a]Mean ± SD.

the median number of treatments for each patient group (12.1 for the FKC comparator group with attrition and 12.4 for virtual patients). The ESA per patient year was calculated by taking the total units of administered ESA and total number of patient-years contributed for each group for the entire study period. Moreover, spline curves were created to understand the relationship between Hgb and ESA in patient populations. The spline curves were derived from a general additive model[29,30] of cumulative monthly ESA dose and monthly average Hgb for each patient each month.

RESULTS

A comprehensive physiology-based mathematical model of erythropoiesis and patient-level clinical data was used to create 6,659 avatars. An anemia protocol used in a large cohort of patients undergoing HD was first tested using a Monte Carlo sampling approach to create virtual patients and then compared to the same tests conducted in avatars that were generated based on individual patient data routinely measured in patients undergoing HD. Further, we introduced stochastic modules that reflected the daily clinical routine to increase the ecological validity of the VIATs. In all *in silico* trials, individual Hgb levels, and the corresponding anemia treatment were simulated for an entire year. Results of the different *in silico* trial setups were compared to a year of clinical data from a random comparator group sampled from 79,426 patients undergoing HD ("reference population") who were treated for anemia using the same ESA protocol.

Patient characteristics

In order to alleviate statistical problems related to sample size differences between avatars and the patients undergoing HD reference population, we randomly selected a comparator group of 6,659 individuals from the reference population. Hence, all performance assessments of the anemia protocol are based on equally sized populations. **Table 3** shows the descriptive baseline characteristics of the reference patient population, the comparator group (subgroup of the reference population), and the avatar population. All three groups were balanced with respect to their clinical and laboratory data. Differences between the groups were that the avatar population was slightly younger (mean age 64.1 years vs. 65.6 years in the reference population), had been treated for a longer time on dialysis (median vintage 3.4 years vs. 2.3 years in the reference population), and the number of white patients undergoing HD were fewer in the avatar group (57% vs. 61% in the reference population).

Model adaptation: Avatar creation

Personalized anemia avatars were created using individual patient data. The model fit for individual patients was of excellent quality. The mean absolute percentage error (MAPE) between model simulation and empirical data from the avatar patients had a median value of 3.8% (range, 0.9%–13.7%). The MAPE distribution of the 6,659 anemia

Table 3 Characteristics of patients with avatars, the reference population, and the comparator group

	Avatars	Reference population	Comparator group
No. of patients	6,659	79,426	6,659
Male, %	54	55	55
Race, white, %	57	61	62
Age, years	64.1 ± 13.9	65.6 ± 14	65.9 ± 13.8
Body mass index, kg/m^2	29.4 ± 7.6	29.1 ± 7.5	29.4 ± 7.4
Vintage, years (range)	3.4 (1.7–5.8)	2.3 (0.7–5)	2.4 (0.7–4.9)
Comorbid diabetes, %	64	64.50	64.70
Hgb, mg/dL	10.6 ± 0.5	10.7 ± 0.7	10.7 ± 0.7
Pretreatment weight, kg	81.8 (68.6–98.1)	81.4 (68.1–98)	81.6 (68.4–98.1)
Pretreatment SBP, mmHg	150.1 ± 17.6	148.5 ± 18.4	148.3 ± 18.4
Post-treatment SBP, mmHg	138.2 ± 16	137.9 ± 16.4	137.6 ± 16.4
Treatment time, minutes	221.3 ± 25.4	222.7 ± 28.6	222 ± 29.1
Ultrafiltration volume, kg	2.44 ± 0.93	2.3 ± 0.9	2.34 ± 0.92
Interdialytic weight gain, %	3 (2.3–3.6)	2.9 (2.3–3.5)	2.9 (2.3–3.5)
Albumin, g/dL	3.9 ± 0.32	3.8 ± 0.4	3.8 ± 0.4
Iron dose, mcg/month	250 (150–500)	250 (150–500)	250 (150–500)
Dialysate sodium, mEq/L	137 (137–138)	137 (137–138)	137 (137–138)
Neutrophils to lymphocytes ratio	3.6 (2.7–5)	3.7 (2.7–5.2)	3.7 (2.7–5.2)
Transferrin saturation	34 ± 8.1	33.6 ± 8.9	33.7 ± 8.7

Hgb, hemoglobin; SBP, systolic blood pressure.

avatars is shown in **Figure 2**. Although the distribution of the model error was subtly right-skewed, the MAPE was below 6.9% for 90% of the patients.

All model parameters that were determined for individual patients were physiologically reasonable for patients undergoing HD. For further information on the avatars and their creation see the **Supplementary Information**.

Virtual Anemia Trial 1.0: Using virtual patients created by Monte Carlo sampling

A cohort of 6,659 virtual patients was created by randomly sampling unique parameter values from a parameter space defined *a priori*. The parameter distributions used to describe the physiologically reasonable parameter space had been previously determined as being meaningful for patients undergoing HD (Fuertinger, D.H., Kappel, F., Zhang, H., Thijssen, S. & Kotanko, P., unpublished data). For each virtual patient, weekly Hgb values were simulated and decisions on ESA dose adjustments were made following an anemia treatment protocol used in a large number of US dialysis clinics. The results showed poor alignment with the respective clinical data (**Figure 3a**). The distributions of the predicted Hgb values (mean ± SD, 11.3 ± 3.4 g/dL) did not reflect the empirical Hgb data (10.8 ± 1.1 g/dL); compare also to **Table 1**. Moreover, the clinically prescribed ESA doses and the simulated prescribed doses showed an almost inverse pattern (**Figure 3b**). Although the lowest ESA dose was the one most commonly administered in the real clinics, the simulation predicted that the highest dose would be applied most often. This behavior resulted in a 52.5% overestimation of ESA use per patient-year in the VIAT 1.0 compared to the clinical data.

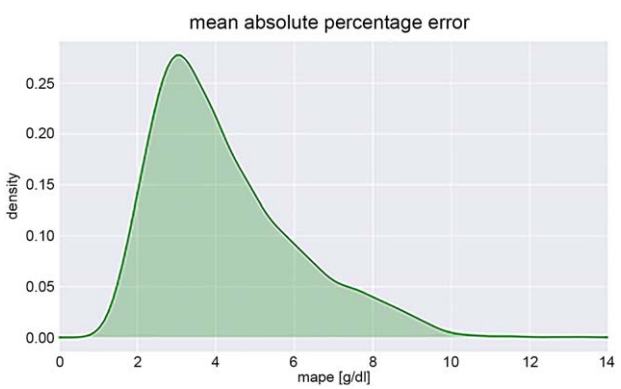

Figure 2 Mean absolute percentage errors (MAPE) between model simulation and individual patient data. Density of MAPE of hemoglobin levels between empirical patient data and all avatars during the model adaptation period.

Virtual Anemia Trial 2.0: Using personalized avatars to improve external validity

A virtual patient population of 6,659 avatars was created using routinely collected clinical data from individual patients undergoing HD treated for anemia. This cohort of virtual patients was subjected to the same anemia treatment protocol used in the dialysis clinics. Weekly Hgb values were simulated for each avatar and ESA dose adjustments were made following the same protocol used in the VIAT 1.0. This time, the predicted ESA use per patient-year underestimated the clinical data by a mere 5.1% (**Table 1**). Moreover, ESA dose distributions mirrored empirical characteristics closely (**Figure 3d**). However, the simulated Hgb distribution showed some clear deviations in the 10–11 g/dL range (**Figure 3c**). A closer look at the spline curve for the monthly average Hgb values at the monthly cumulative ESA doses (both determined on a per patient level) revealed a wide gap between predicted and real ESA use in the lower Hgb

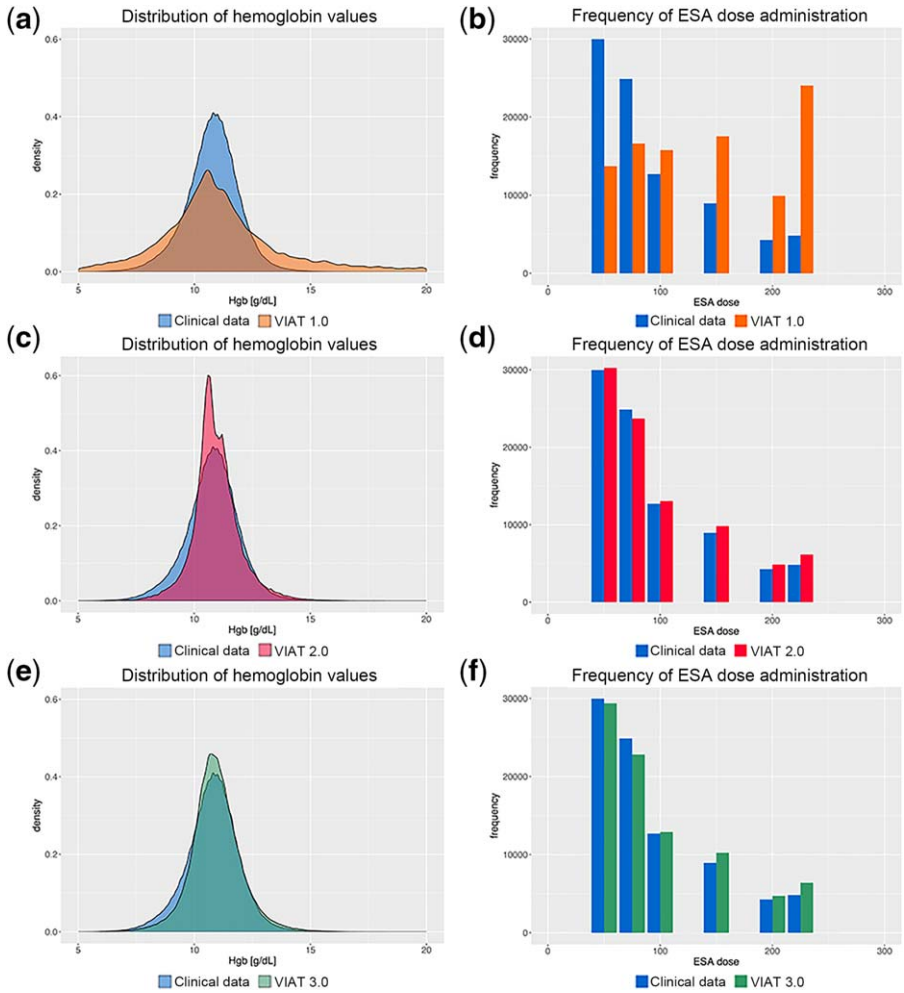

Figure 3 Comparison of clinical data of the comparator group and simulations obtained from Virtual Anemia Trials (VIAT). Empirical data is shown in blue across all panels; comparisons to results from VIAT 1.0, VIAT 2.0, and VIAT 3.0 are depicted in the top, middle, and bottom row, respectively. Panels (**a**), (**c**), and (**e**) exhibit the distribution of empirical hemoglobin (Hgb) values routinely measured in the clinics over the course of a year (blue) and simulated Hgb values for an entire year as obtained from VIAT 1.0 (orange), VIAT 2.0 (red), and VIAT 3.0 (green). Panels (**b**), (**d**), and (**f**) show the corresponding frequencies of empirical (blue) and simulated (orange, red, and green) erythropoiesis stimulating agent (ESA) doses.

range (<10 g/dL; **Figure 4a**). The mean ESA dose per HD treatment was 22.5 mcg in the VIAT 2.0 and 19.5 mcg in the empirical data for patients, with an average monthly Hgb of 8–10 g/dL (**Table 2**). For patients with an average monthly Hgb of <8 g/dL, the gap was even more pronounced with a difference in the per HD treatment ESA dose of 12.1 mcg (39.9 mcg vs. 27.8 mcg). Although this group is small and, thus, did not noticeably increase the overall ESA use of the population, this particular characteristic of the *in silico* trial is disconcerting, as this specific patient group is of high concern clinically.

Virtual Anemia Trial 3.0: Adding stochastic modules to improve ecological validity

The avatar population was subjected to the identical anemia treatment protocol as in trials 1.0 and 2.0. In the present trial, however, we incorporated clinic modules in the simulations to increase ecological validity. These modules were of stochastic nature and reflected important aspects of clinical routine. For instance, a laboratory module was designed to add noise patterns to simulated Hgb levels to reflect both the measurement noise of the laboratory device and the varying fluid status of the patient. Possible shipment delays of blood samples were included, and we accounted for the fact that a small fraction of blood samples that arrive at the laboratory are unusable. A schematic of the VIAT 3.0 setup is presented in **Figure 1**. For a more detailed description of the clinic modules, see the Methods section. Of note, the stochastic elements were added in the simulation after the avatars had been created and these modules were not part of the avatar generation process.

Incorporating several clinically relevant modules improved the predictions of the *in silico* trial considerably. The artificial Hgb distributions resembled clinical data exceptionally well. The mean Hgb was slightly higher than in the empirical data, with an overall narrower distribution of observed

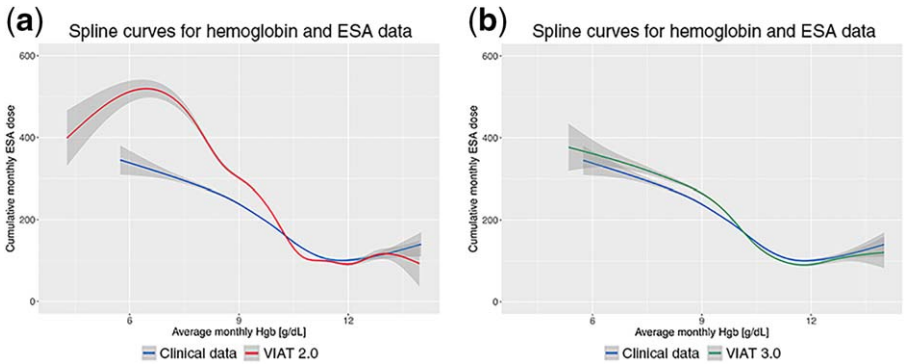

Figure 4 Spline curves representing the relationship between hemoglobin (Hgb) and erythropoiesis stimulating agents (ESA). Blue curves represent Hgb-ESA splines corresponding with clinical data in both panels (**a,b**). Results from Virtual Anemia Trials (VIAT) 2.0 and 3.0 is overlaid in red and green, respectively, **a** and **b**. Spline curves were derived from a general additive model using the cumulative monthly ESA dose and monthly average Hgb for each patient each month.

Hgb values in the *in silico* data (mean ± SD = 10.9 ± 1.1 g/dL vs. 10.8 ± 1.1 g/dL). Further, predicted ESA showed good alignment with clinical data. The median ESA dose per treatment was underestimated by 2.5%, and the cumulative ESA dose per patient-year was underestimated by 5.4% (**Table 1**). Moreover, spline curves of predicted monthly average Hgb vs. monthly cumulative ESA dose further highlighted the excellent agreement between simulated and empirically observed data over the entire Hgb range (**Figure 4b**). Specifically, the difference between predicted vs. clinic data in the low Hgb range that was apparent in the VIAT 2.0 was no longer present after including the stochastic clinic modules. Average per treatment ESA doses were materially identical in *in silico* predictions and clinical data in low Hgb ranges, with 19.4 mcg vs. 19.5 mcg in the 8–10 g/dL Hgb range and 27.3 mcg vs. 27.8 mcg in the <8g/dL Hgb range (VIAT 3.0 vs. clinic data; compare also **Table 2**).

DISCUSSION

A comprehensive physiology-based mathematical model of erythropoiesis together with a large avatar cohort and specifically designed clinic modules was used to design and execute a sophisticated Virtual Anemia Trial. In this report, we show that proactively addressing the questions of external and ecological validity in the design of an *in silico* trial can significantly improve its predictive power. Our final setup of the VCT (VIAT 3.0) clearly outperformed other designs both with respect to the big picture (e.g., Hgb level distribution and ESA use per patient-year) as well as more granular metrics (e.g., number of ESA administrations in different Hgb buckets).

Our study has several strengths. First, we used routinely collected clinical data from a large cohort of patients undergoing chronic HD (>37,000 patients), and we used their data over a baseline period of 90 days to create personalized avatars ($N = 6{,}659$). Importantly, patients were randomly selected for avatar creation, and we did not omit populations that may have been excluded from traditional RCTs, such as elderly, frail, and multimorbid patients. We consider this an important aspect, because this random selection helped us to ensure external validity.

Further, ecological validity is a well-recognized weak spot in traditional RCTs. For the final setup of our VCT (VIAT 3.0), we attempted to reflect some of the key clinical, operational, and laboratory intricacies by integrating clinic modules. These clinic modules were designed by analyzing the real-world challenges and using actual real-life data from FKC clinics to determine probability estimates for hospitalization patterns, laboratory processing times, and patient nonadherence, among others. Our results clearly indicate that improving ecological validity by integrating these clinic modules improved the predictive quality of the VCT. Last, we tested a widely used anemia treatment protocol, and the results of the VIATs were then compared with clinical data obtained from patients who were treated using the same protocol in real life (reference population $N = 79{,}426$, random comparator group $N = 6{,}659$).

Although our results are of interest from a pure academic standpoint, they may also help to address clinical and pharmaceutical research and development needs, such as identifying optimal drug dosing and treatment schemes. With large avatar populations at hand, the development of future anemia treatment algorithms can be streamlined and accelerated by using mathematical modeling and *in silico* clinical trials. Our modeling work is not just an exercise to predict real-world data. The overarching goal of our efforts is to improve anemia management, an aspect central to patient care and outcomes. To achieve that goal, a series of interim goals must be reached: first, we need an avatar population that mirrors a real patient population to run virtual trials of anemia treatment algorithms; we believe that our results are a crucial step in that direction. Second, we must better understand our patients' (patho)physiology, something our modeling approach can deliver. For example, resistance to ESA is a major clinical concern and has been associated with increased patient morbidity and mortality.[31] However, ESA resistance is defined solely based on the presence of an ESA requirement above some threshold (several definitions exist in the literature). Although important,

there is no routinely feasible way for the attending physician to differentiate poor bone marrow response to ESA from shortened RBC life span.[32] The differential diagnosis between these two broad categories (bone marrow pathology vs. short RBC life span) is clinically relevant as it informs subsequent diagnostic and therapeutic steps (e.g., search for hemolysis in the case of very short RBC life span; search for sources of inflammation; to name a few.)

Third, it is reasonable to hypothesize that the model output will aid the identification of patients at high risk for morbidity and mortality. To test this hypothesis, the patient-specific model-derived parameter estimates need to be integrated into prediction models, an exciting area of future research.

In addition to the above, realistic simulations of clinical trials serve many purposes: they can provide valuable information regarding safety and limitations of treatment protocols, eliminate ineffective treatment schemes at an early stage, and support the planning of clinical studies. By preselecting promising dosing strategies, the chance of a successful clinical trial can be increased and, with the support of *in silico* results, the size of a subsequent trial might be decreased. Thus, a well-designed VCT shortens the time to assess improved and novel therapies and consequently shortens the time to deployment in the clinical setting. As indicated above, we consider it a major advantage that VCTs allow for studies in patient groups who are frequently excluded from traditional clinical trials because of ethical and safety concerns (e.g., children, pregnant women, and severely sick patients with multiple comorbidities) and who, as a result, may subsequently also be denied the benefits of the new therapies.

In summary, although RCTs remain the golden standard for regulatory approval, the results from *in silico* models and VCTs could support the development of treatment algorithms; the novel strategies presented here help to improve external and ecological validity of *in silico* trials.

This improvement has been exemplified step-by-step in a VCT conducted using avatars of patients undergoing chronic HD treated for anemia. Future efforts will aim at using the discussed strategies to optimize treatment algorithms that will then be rolled out to patient populations and, it is hoped, will improve clinical care and outcomes.

ACKNOWLEDGMENTS. The authors thank Drs Jeff Hymes and Frank Maddux for providing access to data from Fresenius Kidney Care clinics. Further, we thank Amanda Stennett, Patricia McCarley, and Peter Yu for providing real-life operational data from Fresenius Kidney Care clinics to create realistic clinic modules.

CONFLICT OF INTEREST. All authors have been employed by or have received consulting fees from the Renal Research Institute at the time of the study. D.F. is currently employed by Fresenius Medical Care Germany. P.K. holds stock options in Fresenius Medical Care. S.T. holds performance shares in Fresenius Medical Care.

AUTHOR CONTRIBUTIONS. D.H.F., A.T., F.K., S.T., and P.K. wrote the manuscript. D.H.F., S.T., and P.K. designed the research. D.H.F. performed the research. A.T. analyzed the data. D.H.F. and A.T. contributed new reagents/analytical tools.

1. Wujcik, D. & Wolff, S.N. Recruitment of African Americans to National Oncology Clinical Trials through a clinical trial shared resource. *J. Health Care Poor Underserved* **21**(1 suppl.), 38–50 (2010).
2. Hoos, W.A., James, P.M., Rahib, L., Talley, A.W., Fleshman, J.M. & Matrisian, L.M. Pancreatic cancer clinical trials and accrual in the United States. *J. Clin. Oncol.* **31**, 3432–3438 (2013).
3. Sergeyeva, O. *et al*. Challenges to enrollment and randomization of the Frequent Hemodialysis Network (FHN) daily trial. *J. Nephrol.* **25**, 302–309 (2012).
4. Bhatt, D.L. & Mehta, C. Adaptive designs for clinical trials. *N. Engl. J. Med.* **375**, 65–74 (2016).
5. Clermont, G., Bartels, J., Kumar, R., Constantine, G., Vodovotz, Y. & Chow, C. In silico design of clinical trials: a method coming of age. *Crit. Care Med.* **32**, 2061–2070 (2004).
6. Kansal, A.R. & Trimmer, J. Application of predictive biosimulation within pharmaceutical clinical development: examples of significance for translational medicine and clinical trial design. *Syst. Biol. (Stevenage).* **152**, 214–220 (2005).
7. Leil, T.A. & Ermakov, S. Editorial: The emerging discipline of quantitative systems pharmacology. *Front. Pharmacol.* **6**, 129 (2015).
8. Bangs, A. Predictive biosimulation and virtual patients in pharmaceutical R and D. *Stud. Health Technol. Inform.* **111**, 37–42 (2005).
9. An, G. In silico experiments of existing and hypothetical cytokine-directed clinical trials using agent-based modeling. *Crit. Care Med.* **32**, 2050–2060 (2004).
10. Kovatchev, B.P., Breton, M., Man, C.D. & Cobelli, C. In silico preclinical trials: a proof of concept in closed-loop control of type 1 diabetes. *J. Diabetes Sci. Technol.* **3**, 44–55 (2009).
11. Le Compte, A. *et al*. Blood glucose controller for neonatal intensive care: virtual trials development and first clinical trials. *J. Diabetes Sci. Technol.* **3**, 1066–1081 (2009).
12. Petersen, B.K., Ropella, G.E. & Hunt, C.A. Virtual experiments enable exploring and challenging explanatory mechanisms of immune-mediated P450 down-regulation. *PLoS One* **11**, e0155855 (2016).
13. Jamei, M., Dickinson, G.L. & Rostami-Hodjegan, A. A framework for assessing inter-individual variability in pharmacokinetics using virtual human populations and integrating general knowledge of physical chemistry, biology, anatomy, physiology and genetics: a tale of 'bottom-up' vs 'top-down' recognition of covariates. *Drug Metab. Pharmacokinet.* **24**, 53–75 (2009).
14. Smith, A.K. *et al*. Competing mechanistic hypotheses of acetaminophen-induced hepatotoxicity challenged by virtual experiments. *PLoS Comput. Biol.* **12**, e1005253 (2016).
15. van de Pas, N.C., Woutersen, R.A., van Ommen, B., Rietjens, I.M. & de Graaf, A.A. A physiologically based in silico kinetic model predicting plasma cholesterol concentrations in humans. *J. Lipid Res.* **53**, 2734–2746 (2012).
16. Lamba, M. *et al*. Model-informed development and registration of a once-daily regimen of extended-release tofacitinib. *Clin. Pharmacol. Ther.* **101**, 745–753 (2017).
17. Girard, P. Clinical trial simulation: a tool for understanding study failures and preventing them. *Basic Clin. Pharmacol. Toxicol.* **96**, 228–234 (2005).
18. Kimko, H.C. & Duffull, S.B. Simulation for designing clinical trials: a pharmacokinetic-pharmacodynamic modeling perspective. (CRC Press, New York, NY, 2003).
19. Fuertinger, D.H., Kappel, F., Thijssen, S., Levin, N.W. & Kotanko, P. A model of erythropoiesis in adults with sufficient iron availability. *J. Math. Biol.* **66**, 1209–1240 (2013).
20. Allen, R.J., Rieger, T.R. & Musante, C.J. Efficient generation and selection of virtual populations in quantitative systems pharmacology models. *CPT Pharmacometrics Syst. Pharmacol.* **5**, 140–146 (2016).
21. Cristofoletti, R., Patel, N. & Dressman, J.B. Assessment of bioequivalence of weak base formulations under various dosing conditions using physiologically based pharmacokinetic simulations in virtual populations. Case examples: ketoconazole and posaconazole. *J. Pharm. Sci.* **106**, 560–569 (2017).
22. Ghosh, S., Young, D.L., Gadkar, K.G., Wennerberg, L. & Basu, K. Towards optimal virtual patients: an online adaptive control approach. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2007**, 3292–3295 (2007).
23. Moss, R., Grosse, T., Marchant, I., Lassau, N., Gueyffier, F. & Thomas, S.R. Virtual patients and sensitivity analysis of the Guyton model of blood pressure regulation: towards individualized models of whole-body physiology. *PLoS Comput. Biol.* **8**, e1002571 (2012).
24. Palumbo, P., Pizzichelli, G., Panunzi, S., Pepe, P. & De Gaetano, A. Model-based control of plasma glycemia: tests on populations of virtual patients. *Math. Biosci.* **257**, 2–10 (2014).
25. Schmidt, B.J., Casey, F.P., Paterson, T. & Chan, J.R. Alternate virtual populations elucidate the type I interferon signature predictive of the response to rituximab in rheumatoid arthritis. *BMC Bioinformatics* **14**, 221 (2013).
26. Duncan, T.M., Reed, M.C. & Nijhout, H.F. A population model of folate-mediated one-carbon metabolism. *Nutrients* **5**, 2457–2474 (2013).

27. Chan, K.E., Thadhani, R.I. & Maddux, F.W. Adherence barriers to chronic dialysis in the United States. *J. Am. Soc. Nephrol.* **25**, 2642–2648 (2014).
28. Usvyat, L.A. *et al.* Dynamics of hospitalizations in hemodialysis patients: results from a large US provider. *Nephrol. Dial. Transplant.* **29**, 442–448 (2014).
29. Wood, S.N. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J. Am. Stat. Assoc.* **99**, 673–686 (2004).
30. Hastie, T. & Tibshirani, R. *Generalized Additive Models. Encyclopedia of Statistical Sciences* (John Wiley & Sons, Hoboken, NJ, 2004).
31. Okazaki, M., Komatsu, M., Kawaguchi, H., Tsuchiya, K. & Nitta, K. Erythropoietin resistance index and the all-cause mortality of chronic hemodialysis patients. *Blood Purif.* **37**, 106–112 (2014).
32. Dou, Y., Kruse, A., Kotanko, P., Rosen, H., Levin, N.W. & Thijssen, S. Red blood cell life span and 'erythropoietin resistance'. *Kidney Int.* **81**, 1275–1276 (2012).

Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (http://psp-journal.com)