Quantifying the Benefits of Digital Biomarkers and Technology-Based Study Endpoints in Clinical Trials: Project Moneyball

Hiromasa Mori[a]   Stig Johan Wiklund[b]   Jason Yuren Zhang[c]

[a]UCB, Brussels, Belgium; [b]Captario, Gothenburg, Sweden; [c]UCB, Slough, UK

Abstract

Introduction: Digital biomarkers have significant potential to transform drug development, but only a few have contributed meaningfully to bring new treatments to market. There are uncertainties in how they will generate quantifiable benefits in clinical trial performance and ultimately to the chances of phase 3 success. Here we have proposed a statistical framework and ran a proof-of-concept model with hypothetical digital biomarkers and visualized them in a familiar manner to study power calculation. ***Methods:*** A Monte Carlo simulation for Parkinson's disease (PD) was performed using the Captario SUM® platform and illustrative study technology impact calculations were generated. We took inspiration from the EMA-qualified wearable-derived digital endpoint stride velocity 95th centile (SV95C) for Duchenne muscular dystrophy, and we imagined a similar measurement for PD would be available in the future. DaTscan enrichment and "SV95C-like" endpoint biomarkers were assumed on a hypothetical disease-modifying drug pivotal trial aiming for an 80% probability of achieving a study *p* value of less than 0.05. ***Results:*** Four scenarios with different combinations of technologies were illustrated. The model illustrated a way to quantify the magnitude of the contributions that enrichment and endpoint technologies could make to drug development studies. ***Discussion/Conclusion:*** Quantitative models could be valuable not only for the study sponsors but also as an interactive and collaborative engagement tool for technology players and multi-stakeholder consortia. Establishing values of digital biomarkers could also facilitate business cases and financial investments.

© 2022 The Author(s).
Published by S. Karger AG, Basel

Introduction

Background

Digital biomarkers will transform drug development. The pharmaceutical industry and digital health companies have been evaluating and piloting various digitally enabled measurements in drug development trials. The Digital Medicines (DiMe) Society's library of digital endpoints captures 302 examples across the industry [1].

Correspondence to:
Jason Yuren Zhang, jason.zhang@ucb.com

Fig. 1. Schematic illustration of the contributions of digital technologies to improving clinical study success.
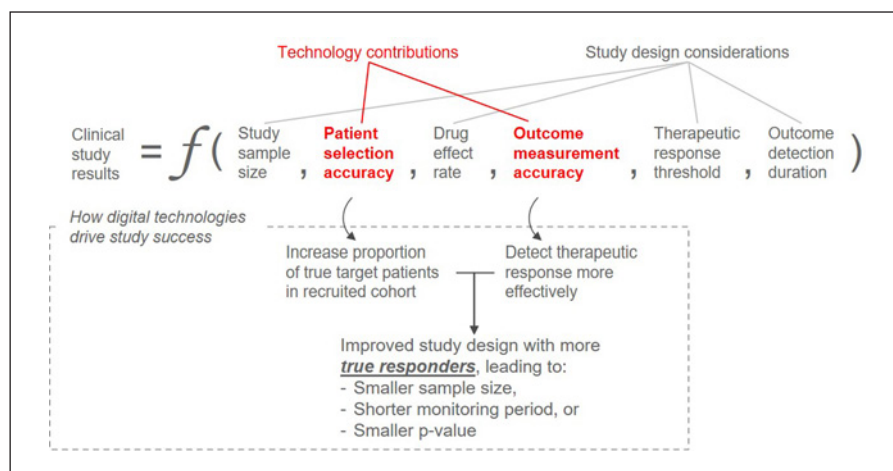
In 2019, stride velocity 95th centile (SV95C) received qualification from the European Medicines Agency (EMA) as the first *wearable-derived digital endpoint* for Duchenne muscular dystrophy [2]. Servais et al. [3] estimated the required pivotal trial sample size in the Duchenne study would be reduced by 70% compared to using the traditional 6-min walk test or North Star Ambulatory Assessment as the primary endpoint. This clearly illustrates the potential magnitude of impact that digital technologies can have on study performance. However, despite the numerous evaluation attempts, very few technologies have made a real-life impact on bringing new treatments to market by serving as pivotal study enrichment or endpoints.

Parkinson's disease (PD) is a neurodegenerative disorder characterized by motor impairments such as tremor, bradykinesia, dyskinesia, and gait abnormalities [4]. Disease onset is typically in late adulthood and progression takes place over decades. There is no approved disease-modifying drug treatment today [5].

Lack of precision in therapeutic outcome measures is a well-known problem in the treatment of PD [6] and many neurodegenerative diseases. Drug development for neurodegenerative disorders is difficult and large-scale phase 3 failures are common due to sparse, subjective data [7]. Dorsey highlights the variability of subjective or quasi-objective measures like finger-tapping in the Unified Parkinson's Disease Rating Scale (UPDRS) and argues that digital biomarkers could improve evaluation of new treatment in these therapeutic areas [7].

However, the pharmaceutical R&D community is cautious of adopting digital endpoints until they are fully proven [8]. Our project aims to develop a business tool to help close this gap, using PD as the exemplar disease with high unmet clinical need and significant technology potential.

Project Rationale

In Figure 1, we list six key factors that affect a drug development clinical study. Digital health technologies could improve accuracy in study patient selection (enrichment) and outcome measurement (endpoints). A study would be more effective if larger proportions of the study samples gave a true response signal, in other words, *correctly identified target disease patients* showing *disease-specific outcome improvement*. Other contributing factors are outside of a typical biomarker team's remit; drug effect rate is largely determined by the therapeutic compound in the context of disease biology. Clinical experts set the sample size, therapeutic response threshold, and outcome monitoring duration.

We have identified five gaps for pharmaceutical sponsors and technology providers to address in the ecosystem of digital biomarkers. The gaps 1, 2, and 3 occur as a result of poor mutual understanding between the pharmaceutical sponsors and technology providers. The direct aim of the *Moneyball* project was to address the gaps 4 and 5 below, and then we hope this will indirectly narrow the gaps 1, 2, and 3.

1. Gaps in evidence requirements: Digital biomarker development follows the V3 (verification, analytical validation, and then clinical validation) process [9]. Clinical validation is often the most resource-consuming step [9, 10] and needs drug development sponsor engagement. On the other hand, technologies can obtain regulatory approval with analytical validation. Explic-

it joint efforts are needed in advance to align biomarker measurements to therapeutic objectives [11].

2. Gaps in the economic models: Digital technology companies typically seek financial returns by selling the devices and/or services and often expect to make return on investment in a short-term. On the other hand, pharmaceutical companies would view digital clinical measurement technologies as study-long investments that will be rewarded upon drug product commercialization.

3. Digital innovation favours desirability over feasibility: Pharmaceutical companies often overestimate the technology benefits and underestimate the development and regulatory burdens. Due to a general shortage of technical expertise [12], mistakes are often made to assume that fast technology development cycles in the consumer IT world translate to regulated health devices.

4. Uncertainties in clinical study impact: A lack of quantified and agreed view on clinical study performance creates an imbalance between the benefits and burdens of the technologies. The discomfort and inconvenience to the patients are easier to imagine, but the values remain speculative in neurodegenerative diseases where there are few approved treatments and successful study templates. Many clinical development teams perceive technologies primarily as a potential study enrolment challenge and they seek positive proof of the benefits to justify such a burden.

5. Lack of a common value framework in multiparty technology collaborations: Digital biomarker development is often pursued in consortia comprised of not only technology providers and pharmaceutical companies but also academia and patient groups [2, 3, 13]. Quantification of the values should facilitate the attribution of resource contributions to the parties involved.

The Project Inspiration: Moneyball

Reading Michael Lewis' baseball novel *Moneyball*, we found a parallel between the sport and the pharma industry. For the last decade, the pharmaceutical industry has been attracted by promising but unproven trendy digital health wearables and artificial intelligence. In a similar way, rich major league baseball teams paid multimillion-dollar salaries to line up promising but unproven young players in their roster [14]. In the 2011 Bennett Miller film *Moneyball*, Peter Brand says the following to Oakland Athletics general manager Billy Beane [15]:

Your goal shouldn't be to buy players.
Your goal should be to buy wins.
In order to buy wins, you need to buy runs.
There is a championship team we can afford.

This called to mind how the pharmaceutical industry has been chasing novel digital technologies with very little clinical study performance benefits realized at the end. We reinterpreted the *Moneyball* lines to our business.

Our goal shouldn't be to buy technologies.
Our goal should be to buy clinical study successes. In order to buy clinical successes in terms of study p value, we need to buy true responders.
There is an optimized study design we can afford.

Billy himself was a high school baseball star who was scouted by the New York Mets, but he did not succeed as a big-league player. We have numerous cases of health technologies that were terminated after pilots. With this metaphor, we first asked what drove clinical study success, and then applied technology evaluation in a quantitative model. In the next section, we will present the quantitative model proposed for evaluation. This model will then be applied in an illustrative example in the design of a biomarker-enriched PD study.

Methods

General Modelling Concepts

Our model has been developed to reflect the probability of success of a drug development clinical trial in quantitative terms. Typically, this means the study achieving a *p* value of less than 0.05 for its primary endpoints and showing that the drug is more effective than the placebo. With Monte Carlo analysis output in a histogram format, we use the term probability of study success (PoSS) to indicate the probabilistic occurrence of $p < 0.05$. In Figure 2, we illustrate an example of 80% PoSS, which corresponds to 80% of the area under the histogram being to the left of the $p = 0.05$ line, and 20% of the area is to the right.

The structure of the model has similarities to the modelling framework proposed by Wiklund [16], with parts of the model being required for the assessment of the PoSS. Other parts of the model are used to capture project level implications, which are described in a later paragraph. In our model, we primarily focus on three components of the design of a clinical trial: the choice of endpoint, i; the choice of study population, j; and the sample size per treatment arm, n.

Treatment Effect

The observed treatment effect in a trial designed to use endpoint i, and targeted study population j, is denoted $\hat{E}_{ij}$ (where for ease of notation, we omit the fact that $\hat{E}_{ij}$ is also a function of n). The estimated treatment effect is assumed to reflect the observed difference between two treatment arms, e.g., between an active treatment group and a control group, each of size n. We model the observed value as the underlying true treatment effect, E_{ij}, plus random error, ε_{ij},

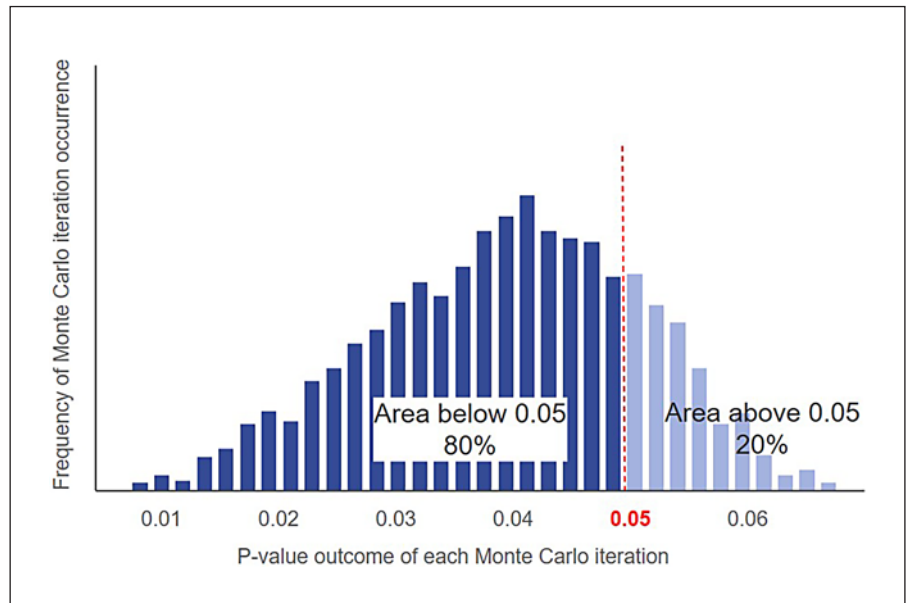$$\hat{E}_{ij} = E_{ij} + \varepsilon_{ij}.$$

Fig. 2. Illustration of PoSS of 80%.

The observational error, ε_{ij}, is approximated by a normal distribution with mean zero and variability given by the standard error, SE $(\hat{E}_{ij})$, i.e., $\varepsilon_{ij} \sim N [0, SE (\hat{E}_{ij})]$. For a continuous endpoint, the standard error would be calculated as

$$SE\left(\hat{E}_{ij}\right) = \sigma_{ij}\sqrt{2/n},$$

where we assume that the estimated treatment effect can be approximated by the difference between two group means. While the comparison of two group means is a simple approximation, the formulation is quite general, and the approximation is applicable to many different types of responses [17].

The true treatment effect, E_{ij}, is assumed to follow a stochastic distribution, representing the current belief and uncertainty regarding the effectiveness of the treatment under development. While the desired treatment effect is often specified as a single value in a target product profile (or similar document), we argue that a more realistic model should acknowledge the fact that the true treatment effect of an investigational treatment is unknown [18].

Criteria for Study Success

A requirement for considering a clinical trial to be successful is that its results show sufficient estimated efficacy. A common criterion to declare success is based on showing a statistically significant difference between the treatment groups. Success is then declared if the observed p value from the trial is lower than what is required for a given level of significance, α, i.e., success is declared if $\hat{p}_{ij} < \alpha/2$ (assuming a two-sided significance level is specified). Let I define an indicator function representing the outcome that the trial is successful:

$$I = \begin{cases} 1 & if \ \hat{p}_{ij} < \alpha/2 \\ 0 & if \ \hat{p}_{ij} \geq \alpha/2 \end{cases}.$$

We assume that the test statistic used for the evaluation of the trial can be approximated by the ratio of the estimated treatment effect and its standard error:

$$\hat{Z}_{ij} = \frac{\hat{E}_{ij}}{SE\left(\hat{E}_{ij}\right)},$$

which is consistent with the assumption that the analysis is approximated by the difference between two group means. With the test statistic, $\hat{Z}_{ij}$, being normally distributed, the one-sided p value is given by $\hat{p}_{ij} = 1 - \phi (\hat{Z}_{ij})$, where ϕ denotes the normal distribution function.

Choice of Study Population

The model described so far is generic and allows for the assessment of any choice of study population. We will illustrate, in this paragraph, how the model is adapted to the case where there is an option whether to use a technology (e.g., a biomarker) for population enrichment. Assume that there are two subgroups of patients; a positive subgroup that is expected to benefit from the treatment under development, and a negative subgroup that is expected to experience less benefit from the treatment. The treatment effects in the two subgroups will be denoted as E_{+j} and E_{-j}, respectively. Let $i = 1$ denote the strategy where a biomarker is used to screen and select the recruited patients, enrolling only the subset of patients categorized into the positive subgroup, and let $i = 2$ denote the strategy where the enrichment biomarker is not used. We will refer to the two study population strategies as "specific" (where patients are selected for enrolment, $i = 1$) and "nonspecific" (where patients are not biomarker-selected for enrolment, $i = 2$).

The specific strategy is applied with the intention to capture the positive subgroup, i.e., to have the treatment effect, E_{+j}. However, since the biomarker used for selection cannot be expected to have perfect sensitivity and specificity, the selection procedure will generally, unintentionally, include some patients from the negative subgroup. The probability of inclusion from the two subgroups is

given by the *positive predictive value (PPV)* of the biomarker, which is calculated as follows:

$$PPV = \frac{Sens \times Prev_+}{Sens \times Prev_+ + (1 - Spec) \times (1 - Prev_+)}.$$

As seen from this formula, the *PPV* is a function of the prevalence of the positive subgroup, and of the sensitivity and specificity of the biomarker used for patient selection. The treatment effect with the specific strategy is then

$$E_{1j} = PPV \times E_{+j} + (1 - PPV) \times E_{-j}.$$

For the nonspecific strategy, the two subgroups will be enrolled in proportions given by the prevalence of the subgroups:

$$E_{2j} = Prev_+ \times E_{+j} + (1 - Prev_+) \times E_{-j}.$$

The treatment effect anticipated for the two subgroups, E_{+j} and E_{-j}, are key inputs when comparing the two population selection strategies. We propose using a factor, F, to represent the relation between the subgroups, i.e., $E_{-j} = F \times E_{+j}$. This implies that an explicit assumption regarding the size and distribution is only required for the positive subgroup, E_{+j}.

Choice of Endpoint

The model allows for the assessment and comparison of any selection of feasible endpoints for the clinical trial. In the case of evaluating a digital technology, we assign assumptions regarding the treatment effect distribution for both a digitally enhanced endpoint and a standard endpoint in the indication of interest, e.g., the outcome of a rating scale. We will denote the two endpoints as $j = A$ and $j = B$, respectively, and the corresponding treatment effects are consequently denoted E_{iA} and E_{iB}.

Monte Carlo Simulation

We will utilize Monte Carlo simulations when evaluating the performance of various design strategies and, in particular, when assessing the value of digital technologies. A simulation will include K iterations, and in each iteration, k, a random number is drawn from the stochastic distribution assigned to each of the parameters in the model. For example, this implies drawing a new value for the true treatment effect, E_{ij}^k, and the random observational error, ε_{ij}^k, in each iteration. Based on these input values, other components and performance metrics can be calculated. In particular, the probability of study success can be calculated as follows:

$$PoSS = K^{-1} \times \left(\sum_{k=1}^{K} I^k \right),$$

i.e., the proportion of iterations representing a successful outcome.

Extensions to Other Data Types

We have previously described the model for the situation where the endpoint of interest is measured as a continuous variable. Our model can of course be adapted to other situations, and we will now illustrate an adaption of the model where the analysis is based on response rate differences.

As illustrated above, for the continuous endpoint, the observed treatment effect is obtained as the underlying true treatment effect plus a random error, $\hat{E}_{ij} = E_{ij} + \varepsilon_{ij}$, where E_{ij} represents the true difference between the mean of two treatment groups. This representation might also be used for response rates, using a normal distri-

bution approximation of the binomial distribution. A more accurate adaption to the response rate situation would use the underlying binomial distribution of the response rates. In this case, the observed treatment effect would be the difference between the observed response rates in the two treatment arms, i.e.,

$$\hat{E}_{ij} = \frac{\hat{R}_{ij}^A}{n} - \frac{\hat{R}_{ij}^C}{n}.$$

The observed number of responders in the control group is given by a binomial distribution, $\hat{R}_{ij}^C \sim$ binomial (n, P_{ij}^C), and for the number of responders in the active treatment arm, $\hat{R}_{ij}^A \sim$ binomial $(n, P_{ij}^C + E_{ij})$. The key input to the model is then the assumptions regarding the probabilities of responders in the control group, P_{ij}^C, and the improvement in response rate achieved by the active treatment, E_{ij}.

Time-Dependent Treatment Effect

The model described in previous paragraphs implicitly assumed that the duration of treatment or duration of follow-up was fixed, or that the treatment effect was not impacted by treatment duration. In many situations, however, the treatment effect will depend on the duration of treatment. The choice of follow-up time may, in these cases, be an important aspect of the design of the trial and, consequently, a key aspect in evaluating the merits of a digital technology's implementation. Our model would then be adapted to let the treatment effect be a function of time, $E_{ij}(t)$. If the underlying science suggests that the treatment effect of the drug would approximately follow an S-shaped increase and eventually approach a full effect, the logistic function may be used as a model:

$$E_{ij}(t) = \frac{E_{ij}^{\max}}{1 + \exp\left[-h(t - \tau)\right]}.$$

The input parameters to this model would be the maximal treatment effect eventually obtained after a long follow-up, $E_{ij}^{\max}$, the time at which half of the maximal effect is obtained, τ, and the slope of the treatment effect increase, h. Another alternative for a time-dependent treatment effect might occur when the underlying disease is continuously deteriorating, e.g., following an approximately linear decline. If the treatment is disease modifying, and thereby is reducing the slope of decline, an adaption of the model might be to assume that treatment effect is proportional to time, i.e., $E_{ij}(t) = m \times t$.

Project Level Extensions

In a previous paragraph, we introduced the PoSS as a key performance metric by which the use of digital technologies in a clinical trial could be evaluated. It should be noted, however, that the proposed quantitative modelling and simulation approach could be expan1jded to assess, from a holistic perspective, the impacts for the development project as a whole. For the example of a digital screening enrichment tool, such an end-to-end project evaluation would account for several aspects that might negatively impact the eventual value of using the tool. These include the following:

- Increased cost for performing the stratification.
- Longer time to recruit patients (due to a lower screening to enrolment ratio).
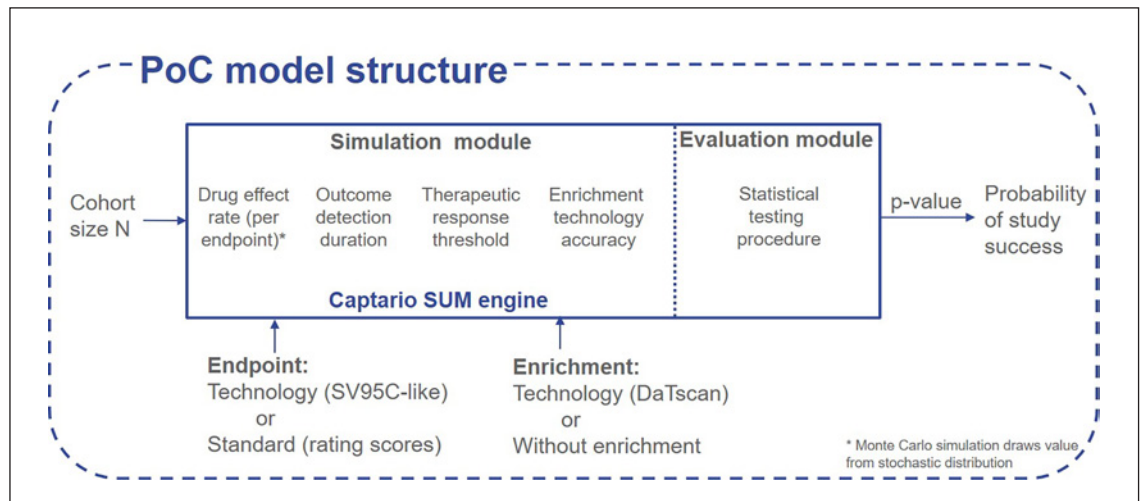- Lower market size (since only a subset of market is targeted).

Fig. 3. Schematic illustration of the quantitative model, the Monte Carlo simulation, and its application to the DaTscan and SV95C example.

These aspects of the tool should be balanced against the potential for positive impact, e.g.,
- Increase in the probability of success.
- Fewer patients required in the trial (due to a higher treatment benefit in the targeted subgroup).
- Potential for premium pricing in a specific targeted patient population.

Following the framework laid out in Wiklund [16], a multitude of project-level metrics could be obtained to inform the assessment of digital technology strategies. With a model including the downstream impacts on success probabilities in subsequent phases, as well as anticipated consequences on market and sales, key performance measures like the expected net present value, return on investment, and probability of technical and regulatory success, for example, could be obtained.

Technologies Providing Example Background

We built our proof-of-concept simulation model inspired by the qualified neuroimaging biomarker of dopamine transporter (DAT) developed by the Critical Path for Parkinson's (CPP). In 2018, the EMA issued a qualification opinion to use DAT imaging to enrich PD clinical trials. CPP's submission dataset (including their power calculation) was made public. CPP's analysis concluded, and subsequently convinced the authorities, that exclusion of subjects who had "scan without evidence of dopaminergic deficit" (SWEDD) could reduce the study sample size by 24% in placebo-controlled DAT-imaging enriched trials with a drug effect of 50% reduction in the progression rate [19].

Since we did not find digital endpoints for PD with the same evidence level as DAT-imaging, we referred to SV95C and its EMA biomarker qualification as if it were for PD. The EMA qualified SV95C as a secondary endpoint in Duchenne MD in 2019, and with a valid and suitable wearable device worn at the ankle, it would quantify a patient's ambulation ability directly and reliably [3, 20]. We must remind the reader that SV95C was developed for and is qualified for Duchenne muscular dystrophy, and we are not suggesting it could be used in PD. Rather, those evaluating our proof-of-concept simulation model should conclude that if in future a similarly evidenced outcome monitoring technology emerges for one of the treatable PD symptoms, then the improvement of the study performance may be quantified as illustrated in this paper. We did not attempt to replicate the exact scientific evidence of the SV95C biomarker into our PD model, but instead we assumed that improvement in signal objectivity and continuous data collection [3] could be replicated.

As illustrated in Figure 3 below, we designed a model and performed Monte Carlo simulations with patient selection and outcome signal detection input parameters, using Captario SUM® as the analytical engine. We illustrated four strategies for the use of digital technologies with the model:
1. Both DAT enrichment (exclusion of SWEDD) and SV95C-like digital endpoint
2. Without DAT enrichment but with SV95C-like digital endpoint
3. DAT enrichment with non-digital endpoint, e.g., UPDRS
4. Neither DAT enrichment nor digital endpoint.

The values assigned to the input parameters of the model, to reflect the four strategies above, are given in the Table 1 below. The model was equipped with a simple user-interface to input assumption parameters, as shown in Figure 4. The implementation of the model also included graphical capabilities to show study technology impact calculations for the scenarios in terms of PoSS, sample size required, and signal detection timeframe.

Results

Based on the *Moneyball* PoC model and the assigned illustrative input parameters, we present two primary graphical outputs for study technology impact calculations. In Figure 5, we illustrate how the PoSS of the different design strategies may depend on the sample size of the study. With the input parame-
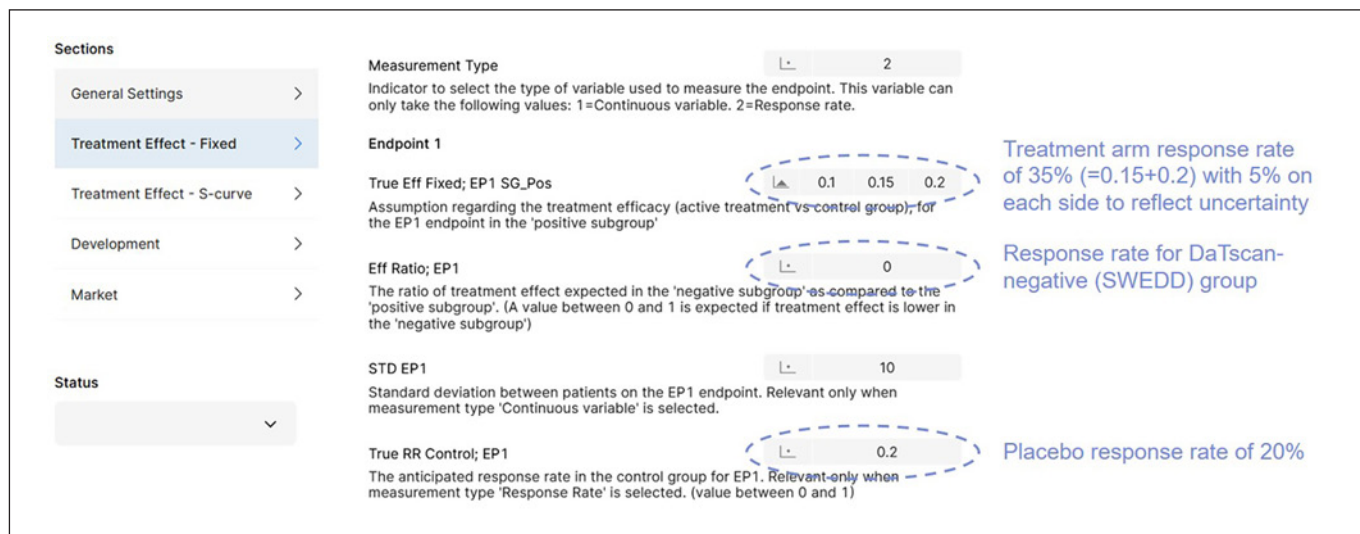
Fig. 4. Screen shot of the parameter input view of the PoC model in the Captario SUM® platform.

Table 1. Parameter values of the quantitative model used in the illustrating example

Parameter	Notation	Value used in simulation
Response rate in control group	P_{ij}^C	20%
Treatment effect (response rate difference between treatment arms)	(E_{ij})	
Positive subgroup, with SV95C-like	E_{+A}	Mean: 15% Range: 10–20%
Positive subgroup, without SV95C-like	E_{+B}	Mean: 15% Range: 0–30%
Factor of negative subgroup treatment effect	F	0
Sample size (per treatment arm)	n	Evaluated over a range 200–700
Significance level to declare study success	α	5%
Biomarker (DaTscan) sensitivity	$Sens$	80%
Biomarker (DaTscan) specificity	$Spec$	80%
Subgroup prevalence	$Prev$	Mean: 80% Range: 70–90%
Model for time-dependent treatment effect		
Maximal treatment effect	(E_{ij}^{max})	
Positive subgroup, with SV95C-like	E_{+A}^{max}	Mean: 15% Range: 10–20%
Positive subgroup, without SV95C-like	E_{+B}^{max}	Mean: 15% Range: 0–30%
Half-time for treatment effect	τ	
With SV95C-like		0.25 years
Without SV95C-like		0.5 years
Slope of logistic function	h	
With SV95C-like		5
Without SV95C-like		3
Number of iterations	K	10,000

Mori/Wiklund/Zhang

Fig. 5. An illustration, based on the Moneyball PoC model, of PoSS and its dependence of sample size for the four design strategies. Figure annotation illustrates the quantification of potential benefits in sample size reduction.
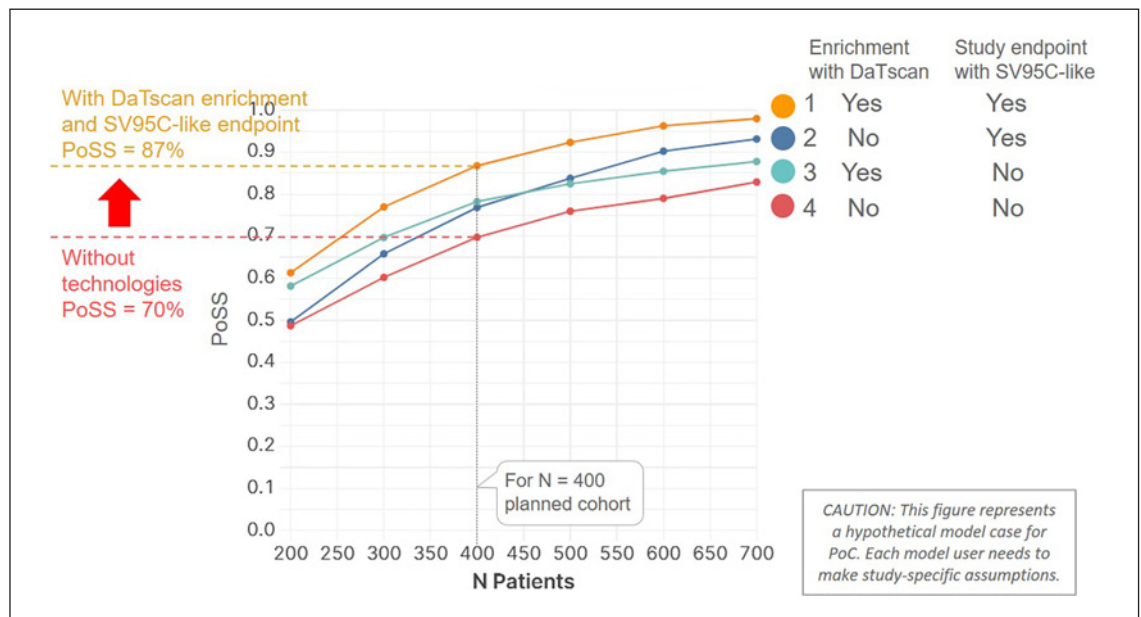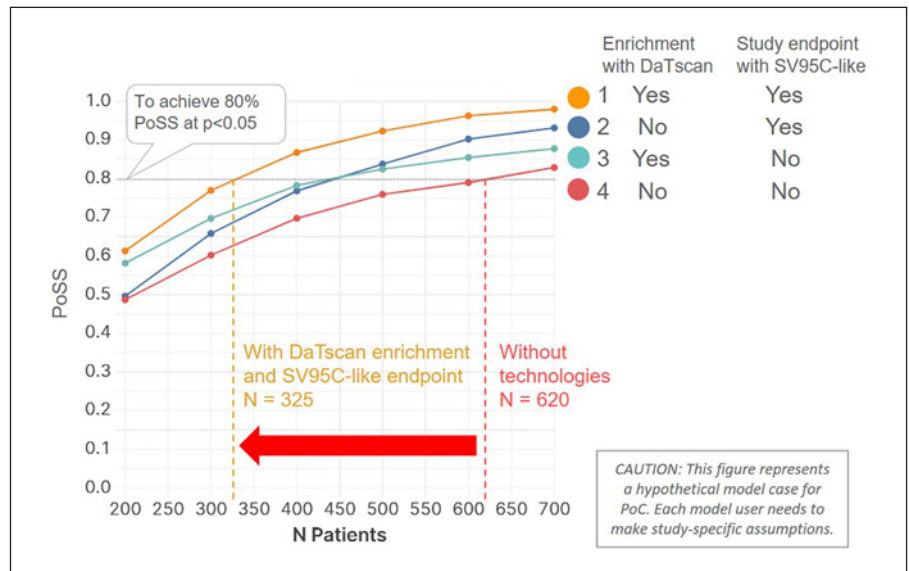

Fig. 6. An illustration, based on the Moneyball PoC model, of PoSS and its dependence of sample size for the four design strategies. Figure annotation illustrates the quantification of potential benefits in improving PoSS.

ters assigned for these illustrations, the difference between the design strategies could be quantified by reading off the sample size required to achieve a desired PoSS. Sample size reduction of nearly 50% reduction may appear drastic, but we reiterate that this should be attributed to the assumptions (and the previously established potential of DAT imaging and SV95C) rather than to the model.

The same type of graph could be used to quantify the difference in PoSS. As illustrated in Figure 6, considering a sample size of $n = 400$, these results would correspond to a PoSS improvement from 70% to 87% when both technologies are applied.

The model is also capable of running similar calculations for study duration impact (signal detection duration from the start of treatment, to be more precise). If we ap-

Fig. 7. An illustration, based on the *Moneyball* PoC model, of the PoSS and its dependence of treatment time for the four design strategies. Figure annotation illustrates the quantification of potential benefits in reducing treatment time and study duration.
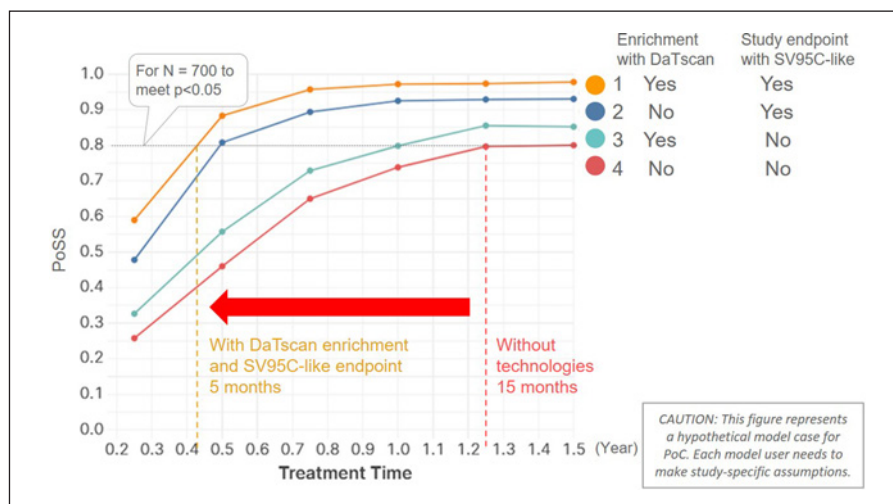
plied the approach outlined in the *Time-dependent treatment effect* paragraph above, then the results of Figure 7 would correspond to a substantial reduction in the study duration. Earlier drug launches lead to higher asset lifecycle values.

Discussion/Conclusion

The *Moneyball* project was undertaken not only to assess the feasibility of such a quantification tool but also to discuss the technology inclusion process with clinical development teams. We received largely positive feedback on our approach of tying technology-enabled measurements to study performance. However, many highlighted the challenge that, unlike with baseball players, technology performance statistics were often unavailable. Initial stakeholder insights can be summarized in the following four points. A *Moneyball* model could support meaningful business activities in

1. Identifying technology-enabled measurements with meaningful impact, and simulating their potential interactively and in real time,
2. Starting biomarker evaluations by thinking what drives clinical study performance,
3. Quantifying the benefits and costs of clinical measurement technologies ahead of time (and writing concrete business cases for investments in technology), and
4. Focussing on and allocating resources to enable technology-inclusive study designs several years before pivotal study initiation.

Our *Moneyball* proof-of-concept model was built to incorporate the functions that were required to illustrate the points described in the Project Rationale section of this paper. We used the existing modelling platform in Captario SUM® and made the customizations necessary for live demonstration and small group discussions. As stated before, the model neither reflected any actual drug in development nor was designed to be used immediately for on-going clinical development programmes. The PD disease model was deliberately over-simplified to limit the project scope.

Further development of the model and user interface are desired, in particular:

- Distinguishing different biomarker types. This first version of the model does not account for the nuances between predictive and prognostic biomarkers. The authors recognize this as a limitation to the model. In real study design simulations, the impact of each biomarker needs to be assessed in the context of the patient population and treatment intent. The model should address this need in future development.
- Reflecting the heterogeneity of symptomatic presentations between patients, while maintaining the relevance of technology-enabled measurements. It is critical to have early guidance from clinical study teams on the expected treatment response signals and their minimal clinically important difference.
- Integrating multiple biomarker technologies within patient selection or outcome measurements. We imagine some clinical studies consider using a combination of genotype-based disease-risk stratification and neuroimaging phenotypes like DaTscan for patient selection.
- Incorporating other types of study endpoint. We illustrated the quantitative model for the cases where the endpoint of interest was either a continuous endpoint

or a response rate. Other types of endpoint are, however, often used for the analysis of clinical trials, e.g., odds ratios, survival times, and hazard ratios. The Captario SUM® platform can be adapted to accommodate these situations:

- Adopting realistic disease progression and treatment effect curves. PD and other chronic diseases of future interest have a disease progression and treatment period of ten or more years. Some therapies require lifelong follow-up.
- Sharpening the digital biomarker contribution dialogue between pharma sponsors and technology partners by speaking the same language on study performance improvements. A discussion guide document listing key questions between the parties should be developed in future.
- Aligning and integrating with the study power calculation methodologies so that drug development strategy and novel measurement technologies can be evaluated concurrently.
- Making the model broadly available to the pharmaceutical R&D community, technology companies, and the ecosystem. We would like to pursue a collaborative and open-platform approach to make improvements to the toolkit.

We conducted the *Moneyball* proof-of-concept project as an illustration of quantitative modelling that could serve a broad set of stakeholders in the drug development technology ecosystem. The underlying framework is disease-agnostic, and with simple modifications to the assumptions, it could be adopted for therapeutic areas outside of PD. The model should also be applicable to other biomarker modalities, such as in vitro diagnostics.

Digital biomarkers will help novel PD therapies and drive the values of drug assets. The pharmaceutical industry must continue the journey. We recommend this type of integrated thinking process is incorporated into key portfolio management decisions of pharmaceutical companies. We believe this model is useful as a collaborative engagement tool with clinical development teams within pharmaceutical companies or technology providers seeking to confirm the value of their offering. Lastly, we caution the potential users that this model should be considered as a compass to set the general direction, rather than a map to make precise study protocol decisions.

Moneyball inspired us with innovative use of statistical modelling to win baseball games. We applied the spirit to address the uncertainties in drug development and newly emerging digital biomarkers. Our model could help identify the most valuable measures and technology players.

However, there is a difference – all stakeholders, including awaiting patients, can win if we can bring novel treatments to market. The authors sincerely hope this article stimulates broad collaborations in the digital biomarker ecosystem.

Statement of Ethics

This project and modelling did not use any data derived by patients, and therefore no consent or ethical approval was sought.

Conflict of Interest Statement

Hiromasa Mori was an employee of UCB, Belgium. Stig Johan Wiklund is an employee and a shareholder of Captario. Captario develops the software, Captario SUM®, in which the Monte Carlo simulations and numerical results of the paper were produced. Jason Zhang is an employee of UCB, UK.

Author Contributions

Hiromasa Mori conceived the rationale and structure for the research project. Stig Johan Wiklund developed the quantitative model used for evaluation, performed the Monte Carlo simulations, and generated the numerical results. Jason Zhang assured the quality of the model and assisted in the communication of expectations. All the authors contributed to the writing, editing, and approval of the manuscript.

Data Availability Statement

No empirical data have been used in the preparation of this article. Input values, used to produce results for the numerical illustration sections, are given in Table 1.

References

1 Digital Medicines (DiMe) Society. Library of digital endpoints. Available from: https://www.dimesociety.org/communication-education/library-of-digital-endpoints/ (accessed March 8, 2022).

2 Servais L, Yen K, Guridi M, Lukawy J, Vissière D, Strijbos P. Stride velocity 95th centile: insights into gaining regulatory qualification of the first wearable-derived digital endpoint for use in Duchenne muscular dystrophy trials. J Neuromuscul Dis. 2022 Mar;9(2):335–46.

3 Servais L, Camino E, Clement A, McDonald CM, Lukawy J, Lowes LP, et al. First regulatory qualification of a novel digital endpoint in Duchenne muscular dystrophy: a multistakeholder perspective on the impact for patients and for drug development in neuromuscular diseases. Digit Biomark. 2021 May–Aug;5(2):183–90.

4 Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, et al. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord. 2015 Oct;30(12):1591–601.

5 Armstrong MJ, Okun MS. Diagnosis and treatment of Parkinson disease: a review. JAMA. 2020;323(6):548–60.

6 Lang AE, Espay AJ. Disease modification in Parkinson's disease: current approaches, challenges, and future considerations. Mov Disord. 2018 May;33(5):660–77.

7 Dorsey ER, Papapetropoulos S, Xiong M, Kieburtz K. The first frontier: digital biomarkers for neurodegenerative disorders. Digit Biomark. 2017;1:6–13.

8 Landers M, Dorsey R, Saria S. Digital endpoints: definition, benefits, and current barriers in accelerating development and adoption. Digit Biomark. 2021 Sep–Dec;5(3):216–23.

9 Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling A, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). NPJ Digit Med. 2020;3(55):1–15.

10 Booz Allen Hamilton. Cost drivers in the development and validation of biomarkers used in drug development. Office of the Assistant Secretary of Planning and Evaluation, U.S. Department of Health and Human Services. 2018 Jul 20. Available from: https://aspe.hhs.gov/sites/default/files/migrated_legacy_files//184796/FinalBiomarkersReport.pdf (accessed January 14, 2020).

11 Manta C, Patrick-Lake B, Goldsack JC. Digital measures that matter to patients: a framework to guide the selection and development of digital measures of health. Digit Biomark. 2020;4:69–77.

12 Goldsack JC, Zanetti CA. Defining and developing the workforce needed for success in the digital era of medicine. Digit Biomark. 2020;4(Suppl 1):136–42.

13 Stephenson D, Alexander R, Aggarwal V, Badawy R, Bain L, Bhatnagar R, et al. Precompetitive consensus building to facilitate the use of digital health technologies to support Parkinson disease drug development through regulatory science. Digit Biomark. 2020;4(Suppl 1):28–49.

14 Lewis M. Moneyball: the art of winning an unfair game. New York: Norton; 2004.

15 Miller B. Moneyball (film). USA: Columbia Pictures; 2011.

16 Wiklund SJ. A modelling framework for improved design and decision-making in drug development. PLoS One. 2019;14(8):e0220812.

17 Miller F, Burman CF. A decision theoretical modeling for phase III investments and drug licensing. J Biopharm Stat. 2018;28(4):698–721.

18 Wiklund SJ, Burman CF. Selection bias, investment decisions and treatment effect distributions. Pharm Stat. 2021;20(6):1168–82.

19 Committee for Medicinal Products for Human Use. Qualification opinion on dopamine transporter imaging as an enrichment biomarker for Parkinson's disease clicnical trials in patients with early Parkinsonian symptoms. European Medicines Agency; 2018 May 29. EMA/CHMP/SAWP/765041/2017.

20 Committee for Medicinal Products for Human Use. Qualification opinion on stride velocity 95th centile as a secondary endpoint in Duchenne muscular dystrophy measured by a valid and suitable wearable device. European Medicines Agency; 2019 Apr 26. EMA/CHMP/SAWP/178058/2019.