10 November 2022
EMADOC-1700519818-955047
Executive Director

# Letter of Support of Model-based Clinical Trial Simulation Platform (CTSP) for Duchenne Muscular Dystrophy

On 04/12/2020, the Applicant Critical Path Institute Limited requested scientific advice for their model-based Clinical Trial Simulation Platform (CTSP) for Duchenne Muscular Dystrophy pursuant to Article 57(1)(n) of Regulation (EC) 726/2004 of the European Parliament and of the Council.

Model-based CTSP for Duchenne Muscular Dystrophy attempts to model disease progression with the aim to aid clinical trial design in Duchenne Muscular Dystrophy.

On 29/09/2022, the SAWP agreed on the Qualification Advice to be given to the Applicant.

On 13/10/2022, the CHMP adopted the advice to be given to the Applicant.

**Applicant's executive summary of development and validation**

The objective of the request

Critical Path Institute's (C-Path's) Duchenne Regulatory Science Consortium (D-RSC) seeks to obtain the European Medicines Agency's (EMA) feedback on a model-based clinical trial simulation platform aimed at optimizing clinical trial design of efficacy studies of potential therapies for Duchenne muscular dystrophy (DMD), through the Qualification of Novel Methodologies in Drug Development mechanism.

The proposed platform is based on a series of quantitative models describing the longitudinal progression of the disease across a series of clinically outcome measures. The D-RSC team developed models of the longitudinal changes in the velocity at which individuals can complete specified timed functional tests, frequently used as clinical trial efficacy endpoints (supine-stand, 4-stair climb and 10-meter walk/run test or 30-foot walk/run test), as well as the longitudinal changes in forced vital capacity (FVC) and North Star Ambulatory Assessment total score (NSAA). The longitudinal models incorporate relevant sources of variability, such as baseline severity at study start, age, steroid use (at baseline or naïve), genetic mutation, study type (clinical trial vs. observational), and race. Additional models were developed to predict dropout of individuals with DMD for each of the functional tests, and joint models of longitudinal change and dropout were developed. The trial simulation platform, which is based on the individual models for the five different endpoints, is intended to inform exploration of key design constructs including sample size, sampling scheme, patient enrollment criteria, dose and study design, and accelerate medical product development in DMD. The user of the tool chooses the endpoint of interest, proposed inclusion criteria for the trial, proposed length of the trial and estimated drug effects, and then can perform simulations. Each of the input parameters can be altered until an optimal trial design is determined. If desired, this can then be repeated for one or more of the other endpoints of

interest. By simulating outcomes of one or more endpoints in various potential trial populations, sponsors are able to ensure trials are designed as optimally as possible.

Characteristics of the Proposed Novel Methodology

- Intended Application: To help inform, through simulations, trial enrichment strategies, including the selection of inclusion/exclusion criteria, stratification approaches, timing and selection of clinical assessments, trial duration, and sample size for studies evaluating therapeutic candidates for DMD.
- General Area: Clinical trial simulation (CTS) for Duchenne muscular dystrophy (DMD) trials.
- General Description: A disease progression model-based CTS platform designed to optimize clinical trial enrichment and design of studies to investigate efficacy of potential therapies for DMD. Measurements of DMD disease progression will be based on changes in a series of endpoints:
- North Star Ambulatory Assessment
- Velocity of completion of supine-stand test Velocity of completion of 4-stair climb test
- Velocity of completion of 10-meter walk/run test or 30-foot walk/run test Forced vital capacity (FVC)
- Univariate models have been developed to understand the effects of known covariates (e.g., steroid use, mutation type, study type) on the progression of each proposed endpoint with age. Each univariate model may, on its own, be used to inform clinical trial design optimization decisions, for each endpoint.
- Target Population for Use: Individuals with DMD 4 years of age and older (endpoint-dependent), regardless of stage of disease.
- Stage of Drug Development for Use: All clinical efficacy evaluation stages of drug development in DMD, including early efficacy, proof-of-concept, dose-ranging, and registration studies.

Additional Clarifications

Clinical trial simulations based on this model should be considered in the context of additional aspects (e.g., characteristics of the individual population, clinical pharmacology and safety of potential therapies) and are not intended to replace the execution of actual clinical trials for the assessment of safety and efficacy.

The underlying individual-level data to develop each specific univariate model come from data sources in the D-RSC database that represent each specific endpoint and covariates, and thus may not come from all the data sources in the D-RSC database. Users should take this information into account when running simulations for trials intended to collect multiple endpoints.

Until relevant individual-level data from active arms of clinical trials are integrated into the D-RSC database, the drug effect components of the models to estimate potential symptomatic, disease-modifying, or combined effects remain theoretical.

By their nature, quantitative drug development tools, like the one proposed here, are dynamic and will be refined/improved as additional data become available. The regulatory qualification of the proposed CTS platform does not imply a regulatory qualification of the selected model endpoints.

The regulatory qualification of the proposed CTS platform does not imply that the models can predict a population that could be used as a virtual control dataset; such a discussion should occur between an individual drug sponsor and the Agency in the context of a drug development program.

Sources of Data

Longitudinal individual-level data from eleven studies in individuals diagnosed with DMD have been integrated to support development of the CTS platform.

An additional four studies have been integrated as an independent dataset for model validation.

Studies used for model development:

- University of California (UC) Davis test/re-test data for a clinical outcome assessment
- UC Davis natural history study
- Cincinnati Children's Hospital clinical data collection
- Cooperative International Neuromuscular Research Group (CINRG) Duchenne natural history study. (NCT00468832)
- Santhera Phase III idebenone placebo arm (NCT01027884)
- Eli Lilly Phase III tadalafil placebo arm (NCT01865084)
- Children's Hospital of Philadelphia (CHOP) natural history study
- Functional data from the Imaging DMD study of skeletal muscle MRI biomarkers (NCT01484678)
- PTC Therapeutics Phase II ataluren placebo arm (NCT0026488)
- PTC Therapeutics Phase III ataluren placebo arms (NCT00592553)
- CINRG Phase II steroid trial (NCT00110669)

Studies integrated as independent validation datasets:

- Biomarin Phase II drisapersen placebo arm (NCT01153932)
- Biomarin second Phase II drisapersen placebo arm (NCT01462292)
- Biomarin Phase III drisapersen placebo arm (NCT01254019)
- Biomarin natural history study to support drisapersen 2

Overview of the Completed Results and Key Findings

The completed results of this work include five individual disease progression models describing longitudinal changes in the velocity at which individuals can complete specified timed functional tests frequently used as efficacy endpoints for clinical trials (supine-stand, 4-stair climb, and 10-meter walk/run test or 30-foot walk/run test) and the longitudinal changes in forced vital capacity (FVC) and North Star Ambulatory Assessment total score (NSAA). Each model was developed using individual-level data from multiple integrated data sources, which included clinically relevant covariates based on data availability. For each disease progression model, a corresponding dropout model was developed to predict the time-varying probability of dropout, conditioned on individual-specific information, which could include the value of the functional endpoint. Each model was externally validated on multiple standardized and integrated datasets.

Key findings indicated that each of the models robustly describes population level disease progression

while quantifying relevant sources of variability. Sources of variability include observable features including race, mutation type, baseline steroid use, and study type. Unobservable features included between-subject variability and residual variability. The integrated dataset that supports this series of models has been explored intensively and final analysis datasets were extracted for each of the five endpoints. Separate but corresponding analysis datasets were extracted from the independent aggregated dataset for external validation. The data were carefully evaluated and datapoints missing due to the loss of ability to complete a test were labelled with an "inability" tag to distinguish them from other missing data. Inability to complete a test was indicated by the data contributor. The contributor either specified a reason that the test was not done 'due to disease progression' or entered an indicator value of the test not completed due to disease progression.

Data exploration showed that the datasets contained sufficient data across a suitable age range in order to develop models for the five endpoints. A rigorous process was performed to assess the most appropriate base model for each endpoint. Covariates including demographics, genetics and baseline severity were added to increase the predictive power of the disease progression models. Dropout models were developed to predict the time varying probability of dropout of individuals with DMD. For some endpoints, the longitudinal dynamics of the endpoint influenced dropout. A joint modelling approach was used to link the longitudinal trajectory of the clinical scores and baseline individual features to accurately predict dropout times.

Five disease progression models were developed: 1) NSAA, 2) velocity of completion of the supine-stand test, 3) velocity of completion of the 4-stair climb test, 4) velocity of completion of the 10-meter walk/run test or 30-foot walk/run test and 5) FVC.

The structural model that most appropriately described progression for each endpoint was the product of a Chapman-Richards growth function and a Sigmoid Emax model. The product of these two models accounts for both the incline phase due to growth and natural development and the decline due to disease progression. Based on data availability, covariates were included specific to each endpoint. Model diagnostics included examination of goodness-of-fit (GOF) plots and visual predictive checks (VPC). GOF plots were used to ensure that each model was accurate with no bias in the residuals. VPCs were used to ensure a good prediction accuracy. GOF and VPC plots were generated during both the base model and covariate model selection.

External model validation was performed by testing the predictive performance of each endpoint-specific model on an aggregated independent dataset that comprised three clinical trials and one observational study. VPCs were generated by simulating endpoint-specific trajectories using model estimates from the training data and the baseline values from the external datasets. The developed dropout models were used to predict dropout for the simulations in the external datasets. Results showed good predictive performance, indicating model robustness.

Overall, the modelling exercise and results comprise a comprehensive framework to optimize clinical trial design for efficacy studies in DMD. The collection of disease progression models is intended to offer sponsors several stage-specific endpoints to optimize trial design, including informing power and sample size calculations, trial duration, and entry/enrichment criteria.

**CHMP comments on CTSP development and validation**

- The utility of the platform as proposed in the Context-of-use statement is acceptable, although the understanding of the concept could have benefitted from some concrete examples on how the tool could be used/applied. Note that the stage of drug development for use includes all stages of drug development in DMD, including early efficacy, proof-of-concept, dose-ranging, and registration studies.

The prospect of moving from the crude criteria to a more informative model for the design of the clinical trial is appealing. However, it is important to elaborate on the use cases, in order to make sure that such a model will be used to optimize trial design, including enriching the populations included in the trials. The model may even have a potential to be used for interpolation to patient types that are difficult to participate in a trial. As described in the additional clarifications, simulations based on this CTS platform are not intended to replace the execution of actual clinical trials for the assessment of safety and efficacy of candidate medical products for DMD. In general, the development of a model-based clinical trial simulation platform in the setting of rare diseases with the purpose of improving the design of new clinical trials is appreciated. The utility of the multi-variate model (or multi-state model) is endorsed.

- During the discussion meeting the D-RSC team explained the COU in more detail. The general concept and principles of the COU are agreed upon, specifically around the use of the univariate models to inform stratification, selection of inclusion/exclusion criteria, and to help determine the optimal length and size of trials. However, this is considered of most value in early-stage trials to inform the RCT design for the assessment of safety and efficacy. Five endpoints are proposed to be included in the CTS platform to measure the DMD disease progression. It is important that the implication of this choice is not to reduce the inclusion of other endpoints in new clinical trials. It is important to consider whether the possibility exists to include more endpoints at this stage, which may be of added value in a wider population (younger ambulant and non-ambulant) or when considering potential extrapolation of efficacy or pharmacodynamics. Note that a future regulatory qualification of the proposed clinical trial simulation tool does not imply a regulatory qualification of the selected model endpoints.

The Applicant described very well how the univariate models were developed. However, more limited information was given on how these models will be combined.

Further general suggestions to improve the model strategy are given i.e. the Applicant is asked to consider:

(i) the inclusion of different subpopulations in the model (e.g. using $MIXTURE in NONMEM) representing subgroups of patients with a different disease progression. This is also referred to as a "latent class approach" in statistical literature

(ii) sharing of random effects between univariate models to account for the association between different endpoints (associations on an individual level vs. association on a population level via AGE)

(iii) using a latent variable representing "disease state" as the driver for the univariate models instead of AGE. This approach has the benefit that covariate inclusion will be more consistent as covariates that are introduced on "disease state" will affect all endpoints (to some extent).

(iv) the development of a mixed-effects multilevel item response theory model as previously used for e.g. Parkinson's disease1 or Alzheimer's disease2. Such a modelling framework would capture most of the comments mentioned above.

The tool will also help to inform the selection of inclusion/exclusion criteria. Targeted users of the CTS platform should consider the implications of leveraging the CTS platform in a manner that could exclude too many patients from the trial based on inclusion/exclusion criteria, limiting the generalizability of the trial results and/or ending up with a very small group of subjects. Future enhancements that could be considered, would include the development of a risk-based selection approach based on a

personalized prediction model. Note that a future regulatory qualification of the proposed clinical trial simulator does not imply that the models can predict a population that could be used as a virtual control dataset.

Furthermore, the new trials designed using this tool, should be valid independently from this tool, i.e. the clinical trial simulations should indeed not "replace the performance of actual clinical trials for the assessment of safety and efficacy". Results from the simulations should not be used in the primary analyses of the new designed trials, e.g., a Bayesian approach that will use prior results from this tool will not/hardly be endorsed. From a supportive perspective, re-estimation of model parameters after performance of the clinical trial may be considered valuable for model-refinement. Model validation using independent information should always be an integral part. In addition, it is envisioned that a model-based approach can be used to characterize variability in response between individual patients and patients' subgroups.

The overall dataset incorporates 1139 subjects with DMD over an age range of 4 to 34 years of age. The dataset is large and heterogeneous with respect to the time period the studies were performed, age categories, disease stage, duration of follow-up, number of longitudinal assessments, time interval between assessment, outcomes measured, concomitant use of corticosteroids. However, the adequacy of the database depends on what the most common / overlapping features are, so far it is largely dominated by the ambulant DMD population, the 6-12 years age category, modal FU, time function tests, etc. As an example, the NSAA and PUL are only used in the more recent datasets. So, the representativeness of the database for all DMD stages, coverage of endpoints, etc. needs additional data integration from further sources. Not all existing databases appear to have been shared with the Applicant. The overall dataset is largely dominated by the US region, data from Europe currently are scarce. Thus, the statement that the current database is representative of DMD patients and current DMD studies in general will need further justification.

The proposed endpoints are in general acceptable. However, it would be of interest to understand and adequately model the relationship between the different endpoints i.e., how predictive are initial changes in the timed functional tests endpoints for later changes in clinical outcome e.g., NSAA score. A similar trend among different endpoints is to be considered. Note that the timed-function tests on their own so far have not been accepted as decisive endpoints for decision making because of baseline variability aspects, small observed changes from baseline, which make the clinical relevance of results difficult to evaluate.

Note that the proposed endpoints also reflect data availability within the database. For instance, the revised PUL is nowadays considered as one of the most important endpoints in the non-ambulatory population. The current CTS platform does not include the PUL, so trials that aim to include this endpoint cannot be simulated with the current platform. Also, as a consequence, relatively newer endpoints (e.g., NSAA & PUL) or other novel digital endpoints (e.g., stride velocity, myogrip, Actimyo) are under-represented. Steps should be taken to increase awareness that the development of this CTS platform is not intended to affect the development of alternative and potentially better endpoints (e.g., MFM which can be used in ambulant and non-ambulant individuals and measures three muscular dimensions, currently used in some European centres for clinical management) or using more sensitive and innovative approaches with the potential for use at home (i.e., digital Actimyo).

The Applicant attempted to perform a sensitivity analysis using a model based on a literature review and meta-analysis of survival data in DMD, but the level of granularity in the published meta-data prevented this analysis from being feasible. In the briefing package the Applicant states that the estimated mortality rates stratified by age groups will be used as fixed parameters in the "multi-state

model". If future meta-data include the necessary details, the following considerations are recommended to be included in the analyses:

- how to assess the appropriateness of the literature model,
- how to set up the sensitivity analysis,
- how the dropout model will be integrated in the "multi-state model" or in the univariate models if developing the multi-state model is not pursued (some information is given for the Brooke score model only)
- how to account for dropout not attributable to death.

As mortality was a negligibly occurring event in the population represented in the datasets used to develop the CTS platform, population-wide mortality may be acceptable as a source of data for the mortality component to the CTS platform. However, since the mortality meta-data match the patient-level data across the dropout models lacked the necessary granularity in the publications, this analysis was not feasible. Wherever possible with the worldwide available data, the Applicant is advised to consider joint modelling of drop-out and response/endpoints/outcomes, Data quality is considered an important aspect and should be diligently considered before including worldwide literature data in the analysis.

The D-RSC team developed individual dropout models for each univariate model (submitted in the December 2020 Briefing Dossier) Models for dropout due to mortality were not feasible due to the extremely low frequency of lethality in integrated dataset. Dropout from study was similarly untenable to model due data paucity in the integrated data sources; in all of the clinical trial data incorporated into the CTS platform, dropout from study was a rare event. Due to the inconsistent definition of dropouts in the observational data sources, dropout models for loss of function were not informative.

The strategy to select the base model structure and to incorporate random effects follows general modelling best practices. The Applicant is however requested to be clear on the assumptions that will be part of the selected models. It is recommended to add an alternative model resembling the maturation and development – decline function that is used to describe changes in clearance with age consisting of two sigmoid Emax functions. This might add some more flexibility beyond the equations already considered and might help to alleviate some of the model misspecification that is seen in the GOF plots for the NSAA example. The mentioned cross-validation approach will improve model building but does not provide external validation. External validation is considered necessary. The Applicant may consider to involve groups that are not able to share data at this step and ask them to perform the external validation on their data. This should not be viewed as replacing the need to continuously expand the D-RSC database. The covariates selected are only generally sketched and need more explanation. E.g. the baseline functionality variables are not further specified where numerous choices can be made. In addition, the Applicant should consider (alternatively) estimating baseline variables instead of using them as covariate. Moreover, it is suggested that all potential covariates of interest be explored. The covariate model building approach is understood. However, the method has specific drawbacks, such as correlation between covariates and numerical instability with the number of covariates and bias due to the pre- defined covariate structure. It is therefore not completely understood why a forward-backward covariate selection approach was not chosen. If a full model approach is pursued, a recently published paper suggests using random effects instead of fixed effects in the full model approach (Nyberg et al, 2019.), although experience with this full model approach is minimal. It might be worth to explore. Selection of the covariate model will be based on the same model validation approaches which were described above as relates to base model selection, which is acceptable. If the full model approach is pursued, it should be verified that the number of covariates

will not be too large/the model too complex in relation to the number of observations included in the model (diagnostics such as the condition number should be reported). The graphical exploration of correlations between the different endpoints might not be very sensitive to detect non-linear correlations.

In addition, it is unclear how drug effects are going to be predicted based on the current model. Without extensive active arm patient-level data, specific disease modifying effects on a single or on multiple endpoints cannot be directly modelled. However, the used approach has not been described with sufficient details by the Applicant to have an informed opinion.

While it is agreed that the mentioned cross-validation approach will improve model building, this is not considered sufficient to provide external validation. Whereas the cross-validation is not the problem, the repeated evaluation of the validation cohort is violating the independence of the data e.g. "Summary measures of goodness-of-fit and VPC plots will be generated for each iteration". Therefore, the validation data will influence model building and hence cannot be used for the purpose of validation or the evaluation of predictive performance.

Note that it is also unclear whether the Applicant intends to validate all univariate model separately or validate the multivariate model. In case of the former the same problem as mentioned above applies. In summary, assessing model performance only with cross-validation is not considered sufficient. To improve the credibility of the model in adequately predicting disease progression, external validation will be needed for a qualification opinion.

Covariates were introduced in the model using a stepwise approach based on graphical evaluation. This approach is not in line with the proposed strategy for using the "full model" approach. Furthermore, inter-individual variability with eta shrinkage was above 35% were removed. As per EMA feedback, the Applicant incorporated a prior specification of covariates, testing them on all parameters. This improved model performance across all five endpoints. It is unclear from the documentation whether drop-out or lost-to-follow-up data is available in the provided datasets. As indicated above, these data are considered informative, and should ideally be used for developing a drop-out model. The imputation methods for height, weight and BMI is acceptable. However, the underlying assumptions (e.g. derivation of height from ulnar length) should be clearly stated in future documentation. The likelihood-based M3 method as proposed for missing observations due to the inability to perform a test because of disease progression (e.g., walk in a walk-run test) is a generally accepted method in PK/PD modelling. However, a drop-out model as mentioned above would be preferred.

The robustness of the preliminary results, in terms of the base model for the NSAA, cannot be assessed at the moment based on the information provided. The Chapman-Richards growth function x sigmoid Emax model as the base model structure to capture the dynamic changes of NSAA scores looks promising. Nonetheless, it is noted that this structural model also contains the most parameters and is therefore also most flexible in describing the data. It is unclear whether models included interindividual variability in the model parameter. Nonetheless, it is agreed that the selected base model provides clinical interpretability, given the nature of the parameters describing the maximum fractional decrease in NSAA score (DPmax) and the age at which NSAA score is half of its maximum decrease (DP50). The inclusion of covariates and random effects will improve the models' ability to capture the structural trend of the observed data. The population predictions appear to indicate that for almost all observations a value of approximately 30 is estimated. This is completely resolved by the inclusion of the ETAs. Therefore, the structural trend of the data is not adequately captured without the inclusion of covariates. In addition, it is also noted that in the VPC plots that the median, 5th and 95th percentiles of the observed data are not captured very well within the blue shaded areas (indicating the 90% confidence intervals) of the median, 5th and 95th percentiles (dashed blue lines) of the

predicted values. The addition of relevant covariates increased the model accuracy considerably as shown in the Goodness-of-Fit plots and VPCs in the briefing book. This may also support the need for a multivariate model, whenever the underlying data permit this.

As stated above, the detailed assessment of the model for NSAA and the relevance of the different parameter included is not considered possible due to the limited amount of data available to the Applicant, in particular from the European Union. Given the data driven approach taken by the Applicant, precision and plausibility of parameter estimates as well as overall model predictive performances should be taken into account to retain the different parameters in the different final models.

**Conclusion**

In conclusion, the EMA acknowledges the Applicant Critical Path Institute efforts in establishing their model-based Clinical Trial Simulation Platform (CTSP) for Duchenne Muscular Dystrophy treatments). The EMA has issued this Letter of Support to encourage the further development and validation of the CTSP, as well as encouraging sponsors to share patient-level data with the D-RSC team. A qualification opinion may be considered when the results of the final validation step are available, provided that the CHMP recommendations stated above are taken into account.

The letter of support is issued on the basis of this qualification advice.

Yours sincerely,

Emer Cooke
Executive Director