Study report

Deliverable 3

EMA/2020/46/TDA, Lot 5 (Pharmacoepidemiological research)

The Danish Medicines Agency, Data & Data Analytics Unit

"Description and assessment of fitness-for-purpose of real-world data (RWD) sources on Duchenne Muscular Dystrophy for regulatory decision-making"

Report information

Title	*"Description and assessment of fitness-for-purpose of real-world data (RWD) sources on Duchenne Muscular Dystrophy for regulatory decision-making"* Study report - Deliverable 3
Protocol version identifier	Version 3.11
Date of last version of protocol	09-04-2026
EU PAS register number	EUPAS1000000748
Active substance	Not applicable
Medicinal product	Not applicable
Product reference	Not applicable
Procedure number	EMA/2020/46/TDA, Lot 5 (Pharmacoepidemiological research)
Marketing authorization holder(s)	Not applicable
Framework contractor	The Danish Medicines Agency, Data Analytics Unit
Framework contractor contact person	Niels Henrik Meedom, NHEM@dkma.dk
Joint PASS	No
Research question and objectives	Feasibility assessment plan describing the methods to be used for addressing objective 2
Country(-ies) of study	Not applicable
Authors	Tine Iskov Kopp, Senior Advisor and Epidemiologist, DKMA Monika R K Handest, Academic Employee, DKMA Dino Masic, Scientific Research Advisor, TREAT-NMD Annie Poll, Research Communications Lead, TREAT-NMD

Table of contents

Abstract 4

List of abbreviations 5

1. Background and rationale 6

2. Objectives 6

3. Methods 7

3.1. Identification & shortlisting of data sources 8

3.1.1. Patient registries within the TREAT-NMD network 9

3.1.2. Patient registries outside the TREAT-NMD network 11

3.1.3. Non-patient registries 12

3.2. Data source characterisation of patient registries 12

3.2.1. Data Quality Tool 13

3.2.2. Data Quality Questionnaire 14

3.2.3. Governance & Ethics Questionnaire 14

3.2.4. Reporting of metadata from patient registries 14

3.3. Data source characterisation of non-patient registries 16

3.4. Fitness for use assessment 16

4. Timeline and communication 17

5. References 19

6. Appendices 20

Abstract

This report describes how Danish Medicines Agency (DKMA) plans to assess the fitness for use of Duchenne muscular dystrophy (DMD) data sources to address the four research questions identified in objective 1 of this European Medicines Agency (EMA) funded study.

In objective 1, a list of "core" or "need-to-have" variables - or data elements - were identified for each of the four research questions. These data elements will be thoroughly characterised and assessed for fitness for use in the identified DMD data sources. Additionally, a list of "non-core" or "nice-to-have" data elements were identified. Data elements on this list will be examined to identify whether they are collected in the data source in question or whether they can be found elsewhere.

The overall feasibility plan consists of 3 steps which will partly run in parallel:

1. **Identification & shortlisting of data sources.** Data sources will be identified through web searches on Google and PubMed.gov, as well as review of data sources for rare diseases listed on Orphanet and EMAs Catalogue of Real-world data (RWD) sources which include patients with neuromuscular diseases (NMD)/DMD. All data sources are categorised into:
    - Patient registries within the Translational Research in Europe for the Assessment and Treatment for Neuromuscular Disorders (TREAT-NMD) network,
    - Patient registries outside the TREAT-NMD network, and
    - Non-patient registries (i.e., all other types of data sources including natural history studies)
2. **Data source characterisation.** The DKMA has entered a collaboration with TREAT-NMD, which will manage and collect metadata from all patient registries whereas DKMA will collect information about non-patient registries. TREAT-NMD has developed an online Data Quality Tool, inspired by the EMA-HMA data Quality Framework, which is designed to describe quality of data including:
    - Structure of data submitted
    - Format of values (e.g. dates)
    - Coherence between data elements
    - Plausibility of values using dataset permitted values
    - Proportion of missing data (i.e., to calculate completeness)

    For patient registries, data elements will be systematically characterised based on either this Tool (only patient registries within the TREAT-NMD network) or an Excel-based Data Quality Questionnaire based on the same logics as the Data Quality Tool. Additionally, for all data sources, a separate Data Governance & Ethics Questionnaire will be completed.
    For non-patient registries, available information will be used to complete the Data Quality Questionnaire and Data Governance & Ethics Questionnaire; and if needed, data holders or principal investigators will be contacted to fill in remaining information and for validation purposes.
3. **Fitness for use assessment of core data elements and data sources**. DKMA will assess fitness for use of all data sources based on results from the data source characterisation in close collaboration with EMA and TREAT-NMD. A final feasibility report will include summary descriptions of all data sources. The Structured Process to Identify Fit-for-Purpose Study Design (SPIFD) will be used to assess the fitness for use of all core data elements based on results from the data source characterisation including completeness, coherence and feasibility in line with the EMA-HMA Data Quality framework as well as data governance and ethics.

List of abbreviations

CDA	Confidential disclosure agreement
DKMA	Danish Medicines Agency
DMD	Duchenne muscular dystrophy
DSA	Digital services act
EHR	Electronic health record
EMA	European Medicines Agency
FDA	Food and Drug Administration (US)
NMD	Neuromuscular disease
PMDA	Pharmaceuticals and Medical Devices Agency (Japan)
RWD	Real-world data
RWE	Real-world evidence
SOC	Standard of care
SPIFD	The Structured Process to Identify Fit-for-Purpose Study Design
TGDOC	TREAT-NMD Global Data Oversight Committee
TREAT-NMD	Translational Research in Europe for the Assessment and Treatment for Neuromuscular Disorders

1. Background and rationale

Duchenne muscular dystrophy (DMD) is a severe and rare progressive inherited neuromuscular disorder (NMD), leading to loss of ambulation and premature mortality (Duan et al., 2021). DMD is inherited in an X-linked recessive manner and therefore primarily affects boys. There is currently no medical cure for DMD; instead, treatments focus on managing symptoms and improving quality of life. Recently, several newer therapies such as gene therapy, antisense oligonucleotides and myostatin inhibitors have received restricted approval through the Accelerated Approval pathway (FDA), Conditional Early Approval System (PMDA) or as Conditional Marketing Authorization (EMA). However, uncertainty remains regarding their long-term safety and effectiveness - particularly in real-world settings and across the full disease trajectory. As such, there is a growing need to supplement clinical trial data with real-world evidence (RWE) to support regulatory decisions, including those related to marketing authorisations and post-authorisation requirements.

As a first step towards conducting an RWE study, the identification and assessment of the quality and fitness for use of data sources is essential. A list of core data elements covering four key regulatory research questions was identified as part of objective 1. These are the four research questions of interest:

1. **"How has** the <critical clinical outcome, e.g. NSAA total score, 6MWT, 4SC, respiratory function tests etc.> **evolved from the patient's date of birth or from the relevant index date to be defined** (e.g. age at symptoms onset, date of initial diagnosis) **over the disease lifespan"**?
   Stratified by genotype, disease severity (ambulant patients vs non-ambulant patients), proposed indications in MAA, past and current treatment (standalone vs. combination), age at disease onset / at baseline.
2. **"What is the impact of the evolving clinical guidelines on disease management"** *i.e. on the choice of SOC as well as other types of care such as physiotherapy and other care measures reported by caregivers?*
3. **"How has loss of ambulation in DMD patients evolved over time?"** *Is there a difference depending on genotypes; depending on standard of care (SOC) (e.g. steroids and/or other disease management options such as physiotherapy and other care measures reported by caregivers), depending on proposed indications in marketing authorisation applications; ongoing patient's treatment as well as past treatments (distinguishing between standalone combinations), and depending on age at disease onset or baseline?*
4. Is there an **increase in the occurrence of adverse events** of special interest observed over time in DMD treated patients?

Additionally, a list of data elements that were considered non-core to answer these questions was also identified. Next step is to identify data sources and assess to what extent they collect these data elements and their quality and fitness for use.

2. Objectives

To describe how the Danish Medicines Agency (DKMA) plan to assess the fitness for use of the data sources to address the four research questions identified in objective 1.

3. Methods

DKMA has established collaboration with TREAT-NMD for assisting on the data source characterisation on patient registries. The TREAT-NMD Global Registry Network[1] is a federated network of registries that collaborate across rare, genetic neuromuscular diseases to support clinical trials and drug development efforts globally. The role of the network is to bring together independent, primarily national, patient registries from across the world to collaborate on queries at a global level. Individual registries retain ownership and control of their data and choose whether to take part in global registry enquiries and projects. Governed by a Charter, the registry network is overseen by TGDOC – the TREAT-NMD Global Data Oversight Committee – which is led by a core committee of three chairs. TGDOC includes member registry curators, patient representatives, and representatives of the TREAT-NMD ethics board.

For efficiency and to produce the deliverables in high quality for the achieving of objective 2, the work will be divided into three major steps (Figure 1), which will partly run in parallel. This work plan will allow timely quality control and allow for iteration and/or timely mitigation of risks or barriers identified.

First, a list of potential data sources will be identified and categorised into either patient registries (within and outside the TREAT-NMD network) or non-patient registries (Figure 1). Second, information about data availability and quality will systematically be collected. And third, all information will be used to assess the fitness for use of all data sources in relation to the four research questions and related to the core data elements. We will also identify non-core data elements in all data sources by asking data holders whether they collect this data or if they can get this data elsewhere.

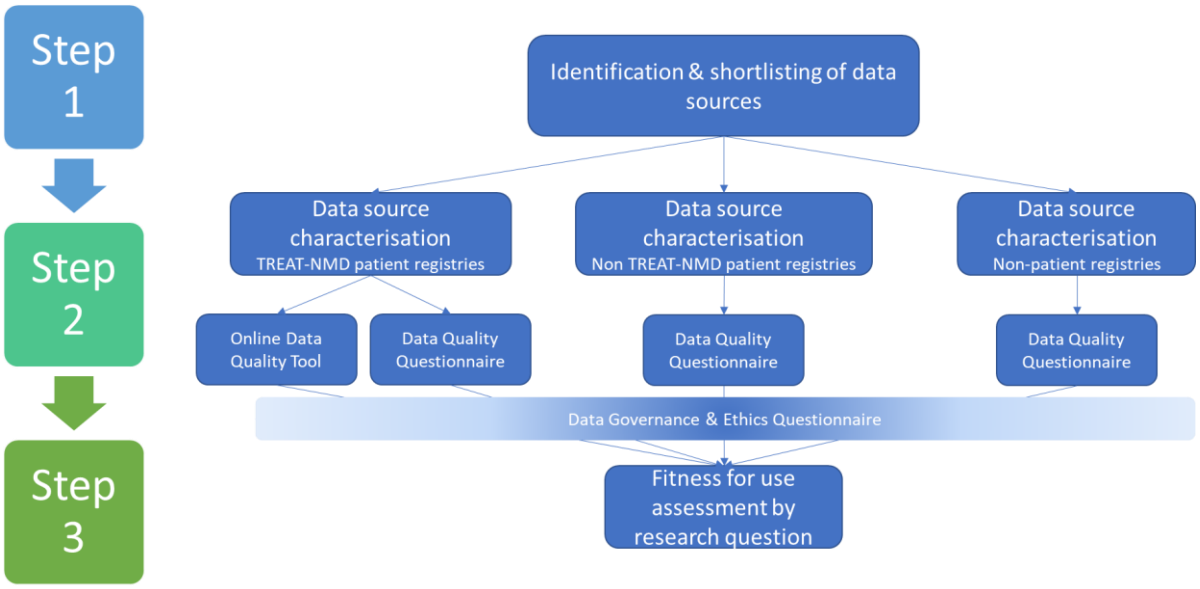

Figure 1. Illustration of the project's overall three processes.

[1] https://www.treat-nmd.org/

3.1. Identification & shortlisting of data sources

The initial list of data sources included with EMA's technical specifications for the study was further enriched and revised by several strategies:

- Two DKMA staff independently searched for combinations of [country] AND [neuromuscular disease] AND [registry OR register OR database OR cohort] and [duchenne muscular dystrophy] AND [registry OR register OR database OR cohort] via Google and PubMed.
- All data sources for rare diseases listed on Orphanet[2] and EMAs Catalogue of real-world data (RWD) sources[3] which included patients with NMD/DMD were reviewed. If a registry listed on one of these websites was not found by Google search in English, the local language was used instead.
- Experts within the field were consulted (including The Danish Rehabilitation Centre for Neuromuscular Diseases and the TREAT-NMD Network) with the aim of adding data sources which were not available through online searches. The experts reviewed the initial list and added data sources as well as comments about their status (i.e., inactive, closed etc). The registries that were considered inactive/closed were retained as we still wanted to contact these registries.

The identified data sources were then categorised into three main categories:

1) **Patient registries** within the TREAT-NMD network,
2) **Patient registries** outside the TREAT-NMD network, and
3) **Non-patient registries**

"Patient registries" are purpose-built data collection systems designed to capture clinical information on patients with a specific disease or condition - in this case, DMD[4,5]. Data are typically entered prospectively by clinicians, nurses or other healthcare professionals, patients themselves or their caregivers - or a combination - and the data elements collected are specifically chosen to reflect the natural history, burden, or management of the disease. Hence, the aim of data collection is not based on one narrowly defined research question, but data can be used for many research purposes.

"Non-patient registries" are all other types of data sources including cohorts based on administrative, population-based registries, electronic surveys and clinical longitudinal studies.

Studies that did not quite fit into any of these categories were initially assigned to the best-fitting category. However, data sources can change category if new information emerges during the data source characterisation process.

Shortlisting of data sources will in general be based on various criteria including the following:

1. Does the data source include DMD patients?
2. Is the data source holder willing to participate in this project (core criteria for patient registries only)?
3. Does the data source collect the core data elements identified in objective 1? And if yes, which ones and to which extent?

[2] https://www.orpha.net/pdfs/orphacom/cahiers/docs/GB/Registries.pdf
[3] https://catalogues.ema.europa.eu/catalogue-rwd-sources
[4] https://www.ncbi.nlm.nih.gov/books/NBK208643/
[5] https://www.ema.europa.eu/en/human-regulatory-overview/post-authorisation/patient-registries

A key criterion is the willingness and capability to participate considering that this work will require some resources from the patient registry holders. However, this also depends on the type of registry and their data structure. Some registries part of the TREAT-NMD network have adopted their data in a harmonized way which is easy to adapt by the Data Quality Tool (see section 3.2.1) whereas others may need to perform additional steps to manually restructure their data prior to uploading to the tool. Also, not all registries may be able to extract the data needed and/or have the resources needed to complete the Data Quality Questionnaire (or Tool).

3.1.1. Patient registries within the TREAT-NMD network

All patient registries within the TREAT-NMD network, which previously indicated that they collect data on patients with DMD (*n*=40, either by clinicians, nurses, patients/caregivers or a combination) have been contacted via the TREAT-NMD curator's portal with a summary of the project and invited to take part in a brief meeting organised by TREAT-NMD staff to explain the expected work and timelines.

The following inclusion criteria will be applied to ensure that a data source is appropriate for this study:

- Include DMD patient data
- Is willing and able to sign a confidential disclosure agreement (CDA) and a digital services act (DSA) with TREAT-NMD with regards to the project
- Has consent and ethical approval in place to collect patient data as an entity, and to share aggregate counts of data with TREAT-NMD and external sources
- Is willing and able to complete the project within the appropriate timelines

As of 11/3/26, 28 TREAT-NMD registries have indicated that they are able, and would like to take part in the project (Figure 2 and Table 1).

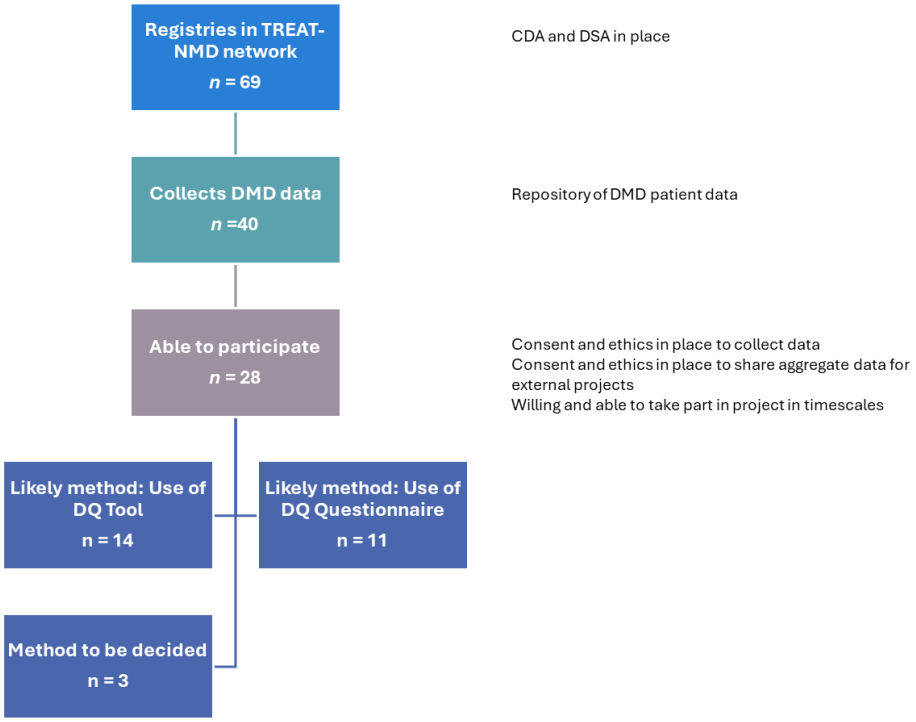

Figure 2. Flow-chart of registry participation within the TREAT-NMD network. CDA, confidential disclosure agreement; DSA, digital services act; DQ, data quality.

Table 1. Data sources within the TREAT-NMD network who have agreed to participate in the project.

Name of registry	Country
Australian Neuromuscular Disease Registry	Australia
The Belgian Neuromuscular Disease Registry (BNMDR)	Belgium
The Bulgarian DMD/BMD patient registry	Bulgaria
The Canadian Neuromuscular Disease Registry (CNDR)	Canada
The NRDRS-DMD/BMD database	China
REaDY Registry: Registry of muscular dystrophies	Czech Republic and Slovakia
[Unknown name]	Egypt
[Unknown name]	Egypt Paediatric
[Unknown name]	Georgia
German DMD- and SMA-patient registry	Germany and Austria
The Duchenne Registry	USA
[Unknown name]	Hong Kong
The Hungarian National Registry for DMD/BMD	Hungary
Indian Association of Muscular Dystrophy (IAMD)	India
The Italian neuromuscular registry	Italy
The Registry of Muscular Dystrophy (Remudy)	Japan
[Unknown name]	Latvia
[Unknown name]	Lebanon
Dutch Dystrophinopathy Database (DDD)	Netherlands
Pūnaha Io – the New Zealand Neuro-Genetic Registry & BioBank (formely the NZ NMD Registry)	New Zealand

[Unknown name]	Pakistan
The Polish Registry of Patients with Neuromuscular Diseases	Poland
[Unknown name]	Romania
[Unknown name]	Slovenia
Spanish DMD Patient Registry	Spain
The Swedish National Registry for Neuromuscular Disorders (NMiS)	Sweden
Swiss Registry for Neuromuscular Disorders - Swiss-Reg-NDM	Switzerland
Turkish NMD Registry - KUKAS	Turkey

3.1.2. Patient registries outside the TREAT-NMD network

This category consists of patient registries prospectively collecting data on DMD patients – either by clinicians/nurses during visits or by patients/care givers – or a combination and who are *not* part of the TREAT-NMD network. TREAT-NMD will also facilitate contact with these registries.

Potential patient registries identified outside the TREAT-NMD network are listed in Table 2. All registries are contacted by e-mail and invited to participate. A total of three e-mail inquiries/reminders will be sent out and if there's still no reply, we consider that these registries are either not interested in participating or do not have the resources to participate or that they are not actively collecting data.

Table 2. Identified potential data sources outside the TREAT-NMD network prior to shortlisting.

Name of registry	Country
The CHFU database for dystrophinopathy	China
French Dystrophinopathies Registry (DYS)	France
MDA-Hellas Registry	Greece
Dystrophy Annihilation Research Trust	India
ICMR Rare Diseases Registry	India
Ambispective, observational study of patients with duchenne muscular dystrophy in Saudi Arabia	Saudi Arabia
Serbian Neuromuscular Disease Network (NMD SerbNet)	Serbia
Spanish Registry of Neuromuscular Diseases (NMD-ES)	Spain
The NorthStar programme	UK
ImagingNMD database	USA
The Muscular Dystrophy Association's neuroMuscular ObserVational Research Data Hub (MOVR)	USA
CureDuchenne Link registry	USA
[Finnish DMD Registry]	Finland
UMD-DMD France	France
The Portuguese patient registry for Duchenne and Becker Muscular Dystrophy	Portugal
The Duchenne Regulatory Science Consortium (D-RSC) Database	Global (Netherlands)
EURO-NMD Registry	Europe
Duchenne Data Platform	Netherlands
Italian DMD Consortium	Global (Italy, UK, US)

3.1.3. Non-patient registries

This category includes various types of data sources such as cohort studies based on population-based administrative registries including claims data, electronic surveys, and clinical, longitudinal studies (Table 3). Randomized, clinical trials are *not* included.

Table 3. List of non-patient registries for assessment.

Name of data source or type of data	Country	Short description of data	Link to published article or clinical trial registration
[Brazilian data sources: EHR records and Rare Disease Registry]	Brazil	EHR records retrospectively collected (manually) on DMD patients and a rare disease registry	Article
Natural History Study In Chinese Male Patients With Duchenne Muscular Dystrophy	China	Multicenter, prospective study on DMD patients (trial setting – commercial data)	Clinicaltrials.gov
[Croatian Health registries]	Croatia	Prevalence study based on population-based administrative registries	Article
[Finnish Health Registries]	Finland	Population-based administrative registries	Article
[Danish Health Registries]	Denmark	Population-based administrative registries	*
Longitudinal Study of the Natural History of Duchenne Muscular Dystrophy (DMD)	Global	Multicenter, prospective study on DMD patients (trial setting - commercial)	Clinicaltrial.gov
Real DMD	USA	Electronic survey on DMD patients	Article
Medicaid Real-World Healthcare Data	USA	Insurance claims data	Article
The muscular Dystrophy Surveillance Tracking and Research Network (MD STARnet)	USA	Population-based surveillance of medical records	Article

**No published article available, but due to the rich health data available in Denmark and the possibility of combining administrative registry data with electronic health record (EHR) data, these data sources are of interest.*

3.2. Data source characterisation of patient registries

Registries within the TREAT-NMD network will be given two choices of participation method when assessing data availability and quality: either use of the Data Quality Tool (see section 3.2.1) or use a detailed questionnaire designed to collect information as closely aligned to the tool as possible, i.e., the Data Quality Questionnaire (see section 3.2.2).

For patient registries outside the TREAT-NMD network, the Data Quality Questionnaire will be used to collect information about data availability and quality in a similar manner as for TREAT-NMD registries. As the Data Quality Questionnaire is set up to accommodate the TREAT-NMD harmonised data set, it may need to be adapted for registries collecting core data elements which are not part of the TREAT-NMD harmonised data set. However, the overall questionnaire structure will be as identical as possible for all patient registries to ensure homogenous outcomes.

Additionally, all patient registries will complete the Data Governance & Ethics Questionnaire (see section 3.2.3).

A report will be delivered to DKMA covering key results from the data source characterisation.

3.2.1. Data Quality Tool

The EMA Data Quality framework[6] sets out key domains of data quality, and to support this effort, TREAT-NMD aimed to develop a tool that could be used by registries to establish both a current (baseline) state of data quality and to monitor progress against improvement plans across these domains. The Data Quality Tool is designed to work with the TREAT-NMD DMD core dataset v1.2 and in line with the EMA Data Quality framework.

In developing the Data Quality Tool, TREAT-NMD had several clear aims:

- To align data quality checks with the EMA Data Quality framework.
- To investigate consistency between all data elements within a registry, data completeness and longitudinally at a data element/year/patient level and to check that data is being formatted and structured in a harmonised way.
- To give control of the tool to registries so it can be used for internal (to the registry) monitoring.
- To develop a single tool that could provide both project-orientated insight and ongoing, independent monitoring in a single environment.
- Ensure no data is stored on the servers running analysis to ensure patient privacy and to abide by registry data sharing policies and agreements.

The tool was developed as an open source custom tool that requires no programming experience to accommodate the broad aims outlined above, and the technical limitations of registries (e.g. that curators could not use an R package). The tool, which was developed in 2024, is a cloud-based solution that requires data to be uploaded for analysis before immediate deletion. The tool was designed to describe (for all data elements):

2. Structure of data submitted
3. Format of values (e.g. dates)
4. Coherence between data elements
5. Plausibility of values using dataset permitted values
6. Proportion of missing data

Registries can use the tool independently to understand and improve their data quality, and results can be shared with TREAT-NMD.

The tool allows registries to keep control of their data and allows TREAT-NMD to run data quality queries against the data using custom scripts written in .net and Python. The data output report viewable by TREAT-NMD contains no patient-level data but offers a high-level overview of completeness of each data element and a list of quality control errors. Data are submitted to the tool in a standardised excel template. The tool itself is a web application, built by Waymark Ltd using the Microsoft App Service, and uses a secure User Authenticated environment. The tool checks quality across several EMA quality framework domains, in particular structure, format, coherence, reliability and completeness. Completeness is calculated per

[6] Data quality framework for medicines regulation | European Medicines Agency (EMA)

data element, whereas coherence and reliability are provided as averaged values per data source. For more details on tool operation, please see Supplementary File 1.

During this project, registries will work towards transforming their data into the DMD standardised excel template required for the tool, and TREAT-NMD will offer technical support to curators where appropriate. TREAT-NMD anticipates submissions to the tool for this project will begin in March 2026, with analyses on quality and availability performed on each registry at the point of submission (see also Figure 1).

3.2.2. Data Quality Questionnaire

For registries unable to use the Data Quality Tool at this point in time, an excel-based Data Quality Questionnaire based on the same logics as the Data Quality Toll will be offered as a work around solution. This questionnaire will be initially created by TREAT-NMD and will incorporate feedback and suggestions from the DKMA team. The questionnaire will be approved by EMA before disseminating. Registries will complete the questionnaire between March-May 2026, with analyses on quality and availability performed on each registry at the point of submission (see also Figure 6). For registries within the TREAT-NMD network, the Data Quality Questionnaire will be submitted via the registry portal to a secure Microsoft SharePoint area which can be accessed by the TREAT-NMD scientific team for analysis. For registries outside the TREAT-NMD network, the questionnaire will be sent by e-mail. The output from the Data Quality Questionnaire will be similar to the output from use of the Data Quality Tool.

3.2.3. Governance & Ethics Questionnaire

All participating registries will be invited to complete a questionnaire detailing information on governance and ethics. This questionnaire will be initially created by DKMA and will incorporate feedback and suggestions from the TREAT-NMD team. The questionnaire will be approved by EMA before disseminating. The Questionnaire will be conducted using Microsoft Forms, which will be accessed by the scientific team at TREAT-NMD.

Throughout the project, the registries will be supported by the TREAT-NMD registry network manager and its scientific team. Adherence to timelines and progress of the registry's efforts will be regularly checked by both the network manager and the project manager team.

3.2.4. Reporting of metadata from patient registries

TREAT-NMD will deliver an individual report for each participating registry to DKMA, covering the following information:

- Registry details including number of (genetically confirmed) DMD patients, diagnostic methods available in the registry and average length of follow-up.
- Data element collection in the form of presence/absence heatmaps, how often data elements are aimed to be collected, average of how often it is collected in practice, and if a data dictionary is used for collection (for example, the TREAT-NMD data dictionary).
- Number of violations of quality checks (list of checks provided in Supplementary File 1), completeness graphics and accompanying data (see below section for more information on calculating completeness statistics)
- Tabulated results of Data Governance & Ethics questionnaires

A template 'shell' report will be shared with DKMA ahead of time with subheadings for feedback to ensure that reports meet the project needs.

Note on 'Completeness'

There is a large amount of logic in-built to the Data Quality Tool (and Questionnaire) to determine completeness. All data elements collected by a registry are assumed to be required unless logic renders this assumption null and void. The logic relies on a specific value in a different data field to conclusively mark the original data element as Not Required for an individual patient in each time period. Some examples of the logic include:

- A date of death is not required if the patient has a status of `Alive` equals YES.
- A value for `Forced Vital Capacity` is Not Required IF a patient is known not to have had a lung function test (i.e. a patient has a value of NO for the data element `lung function test performed` in the same time period.
- A value for walk unaided is Not Required in a given time period IF a patient has a value of NO for the motor ability `able to stand without assistance` in the same time period.
- A start date (and other fields) of a wheelchair episode would be marked as Not Required if NEVER was selected for wheelchair usage for an individual patient.

In the Data Quality Tool, the calculation of the number of patients that required or did not require a value for each data element will be calculated programmatically using the set of rules and logic provided by TREAT-NMD. This will establish both whether missing values were expected to be missing (for example, no data provided for wheelchair start date is expected if wheelchair usage has been stated as Never), and whether a patient had provided data but was not expected to (for example, a patient has multiple entries for Sex even though this data element is only expected to be submitted once). Using these figures provided by the tool, the proportion of completeness for each data element is calculated as:

$$1 - \frac{Number\ of\ patients\ missing\ a\ value}{Number\ of\ total\ patients - Number\ of\ patients\ not\ requiring\ a\ value}$$

For the tool, required and not required data element proportions will be presented as bars, with complete and missing proportions highlighted for the required elements (Figure 3).

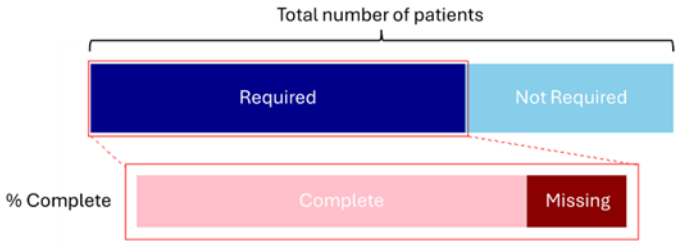

Figure 3. Illustration of how number of patients will be presented.

For the Data Quality Questionnaires, TREAT-NMD will attempt to capture as much information aligned to the tool, but with stated limitations as calculations for curators to perform manually can become quite complex. For example, it may not be possible to calculate Required and Not Required values with as much granularity as the tool. The DKMA and EMA will be consulted through design of the Data Quality Questionnaire to ensure that all parties are satisfied.

3.3. Data source characterisation of non-patient registries

For non-patient registries, DKMA will complete the two questionnaires (i.e., the Data Quality and Governance & Ethics Questionnaire as described in section 3.2.2 and 3.2.3, respectively) as best as possible based on available information on the data sources. As such information in the public domain might be limited, data holders and/or principal investigators in these data sources may be contacted by DKMA when possible, to get missing information and for validation purposes. To the extent possible, DKMA will complete the questionnaire with quantitative results on completeness. If missing values of a data element is reported, the inverse will be used to calculate completeness:

$$\frac{(Total\ records\ \div number\ of\ missing\ values)}{Total\ records} \times 100$$

If completeness cannot be calculated all available descriptions of quality will be added here.

3.4. Fitness for use assessment

A feasibility assessment report will include a summarised description of all included data sources including reasons for not participating. Summary tables or other appropriate graphics will illustrate results on data governance & ethics information as well as details about the study population including data collection periods, study size and duration of follow-up. The report will also include study limitations, lessons learned and recommendation for holders of RWD sources to consider if they want to make their data fit-for-purpose to address the research questions.

The Structured Process to Identify Fit-For-Purpose Data (SPIFD) framework is a structured tool developed to assist in conducting feasibility assessment to determine if a data source is fit for purpose (Gatto et al., 2022). A second version of the SPIFD tool was later developed incorporating the hypothetical target trial design, i.e., SPIFD2 (Gatto et al., 2023). However, to increase the usability of this project for a wider range of research opportunities, we will focus on the data source aspects in this project and thus use the first SPIFD version with minor modifications to assess the fitness for use of all included data sources.

The SPIFD template will be completed with results from the data sources characterisation with the aim of assessing each data element by data sources and by research question identified as part of objective 1. Thus, four SPIFD templates will be completed – one for each research question. Each of the four SPIFD templates will include all the data elements considered as core for the research question in question.

The design elements in the SPIFD template will be categorised into the five domains as was done during the virtual expert workshop in December 2025:

- Primary clinical outcomes (longitudinal)
- Key stratification data elements (to ensure clinically interpretable subgroup analyses)
- Safety and risk data elements (pharmacovigilance relevant)
- Care and standard of care (SOC) context (co interventions and management)
- Confounders (data elements whose presence affects the data elements being studied)

Each result on completeness will be ranked according to a pre-defined definition (Figure 4) resulting in a heatmap of all data elements per data source (Figure 5). This definition allows for both quantitative and qualitative assessment of completeness.

	5	Nearly all patients/items with a value recorded (95-100%)
	4	Majority of patients/items with a value recorded (75-95%)
	3	More than half of patients/items with a value recorded (50-75%)
	2	Less than half of patients/items with a value recorded (25-50%)
	1	Very few patients/items with a value recorded (<25%)

Figure 4. Ranking of completeness of data elements.

Study characteristics and considerations	Requested information	Data source #1	Data source #2	Data source #3
DESIGN ELEMENTS				
Primary clinical outcomes (longitudinal)	*[Enter text]*	4	3	3
Key stratification data elements (to ensure clinically interpretable subgroup analyses)	*[Enter text]*	5	5	5
Safety and risk data elements (pharmacovigilance relevant)	*[Enter text]*	4	1	4
Primary outcome(s) (definition & ascertainment)	*[Enter text]*	4	3	3
Care and standard of care (SOC) context (co interventions and management))	*[Enter text]*	4	3	3
Length of follow-up and data recency	*[Enter text]*	4	3	3
Confounding variable 1	*[Enter text]*	4	3	3
. . . Confounding variable N		3	4	2

Figure 5. Heatmap of completeness results of data elements by data source.

On a data source level, results on coherence and reliability as well as logistical aspects will also be ranked according to pre-defined definitions comparable to completeness (under construction). Logistical aspects include:

- Budgetary considerations
- Timeline considerations
- Other practical considerations (such as specific data access requirements)
- If data source has previously been used for peer-reviewed published research
- If data source has previously been used for research to support a regulatory decision?

The report and SPIFD heatmap will *not* include any sensitive information as only aggregate-level data will be shared with DKMA.

4. Timeline and communication

The below Figure 6 illustrates the tentative timelines of the feasibility assessment. As part of the study includes information from external parties (i.e., data holders), we rely on them to respond to our requests in a timely manner in order to deliver within deadlines. To avoid delays as much as possible, we will ensure

regular follow-up in communication with data holders. Regular meetings with EMA and TREAT-NMD are scheduled to ensure deadlines are met and to communicate if we are falling behind schedule.

	Main responsible party								
	TREAT-NMD	DKMA	JAN	FEB	MAR	APR	MAY	JUN	JUL
Identification and shortlisting of data sources									
TREAT-NMD sources identified	x								
Shortlisting of TREAT-NMD data sources based on availability and likely value	x								
Non-TREAT-NMD data sources identified	x	(x)							
Shortlisting of non-TREAT-NMD data sources based on availability and relevance	x	(x)							
Data source characterisation									
Preparation and verification of Data Quality Tool and questionnaires	x								
Questionnaires to EMA for review					x				
Confirmation of data source readiness and contribution	x								
Collecting of metadata via Data Quality Tool/Questionnaire									
TREAT-NMD Sources	x								
Non-TREAT-NMD sources	x								
Completed questionaires and/or data quality report	x								
Collection of documentation and completion of questionnaires from non-patient registries		x							
Fit-for-purpose assessment of data sources									
SPIFD template modification		x							
Feasibility assessement and report preparation		x							
***Deliverable 4** : Feasibility assessment report Objective 2*		x						D 26.	

Figure 6. Illustration of the tentative timeline for the feasibility assessment plan.

5. References

Duan, D., Goemans, N., Takeda, S., Mercuri, E., Aartsma-Rus, A., 2021. Duchenne muscular dystrophy. Nat. Rev. Dis. Primer 7, 13. https://doi.org/10.1038/s41572-021-00248-3

Gatto, N.M., Campbell, U.B., Rubinstein, E., Jaksa, A., Mattox, P., Mo, J., Reynolds, R.F., 2022. The Structured Process to Identify Fit-For-Purpose Data: A Data Feasibility Assessment Framework. Clin. Pharmacol. Ther. 111, 122–134. https://doi.org/10.1002/cpt.2466

Gatto, N.M., Vititoe, S.E., Rubinstein, E., Reynolds, R.F., Campbell, U.B., 2023. A Structured Process to Identify Fit-for-Purpose Study Design and Data to Generate Valid and Transparent Real-World Evidence for Regulatory Uses. Clin. Pharmacol. Ther. 113, 1235–1239. https://doi.org/10.1002/cpt.2883

6. Appendices

Supplementary File 1. TREAT-NMD Data Quality Requirements

TREAT-NMD Data Quality Requiremen