Official Title: A Registered Observational Cohort Study of Facioscapulohumeral Muscular Dystrophy Type 1
NCT number: NCT04369209
Document date: 2022/2/1

Abstract

Background:

Facioscapulohumeral muscular dystrophy type 1 (FSHD1) is a rare, hereditary, and progressive muscular dystrophy characterized by weakness and wasting of the facial (facio-) and shoulder-upper arm (scapulo-humeral-). The clinical severity of varies from nonpenetrant/asymptomatic to disability with independent ambulation loss or even wheelchair dependence. Expansion of our knowledge on the genetic model -D4Z4 macrosatellite repeats contraction leading to inappropriate expression of *DUX4*- has led to advance in developing (targeted) therapeutic strategies. Thus, clinical trial readiness is now important, particularly patient registry. However, there is yet any FSHD registry in China, let alone epidemiology of FSHD), particularly in terms of prevalence and disease progression/disability.

Methods and analysis:

The China FSHD1 Registry (CF1R) is a nationwide, population-based, longitudinal, non-interventional, and a combination of retrospective and prospective observational cohort clinical study, collecting practice data, including demographic, disease progression history, clinical and genetic features in genetically-confirmed FSHD1 patients/families; this study is performed in Fujian Neuromedical Center (FNMC), where is a diagnosis center for clinical-genetic FSHD1 in China, the only one employing pulsed-field gel electrophoresis (PFGE)-based Southern blotting. Information regarding clinical suspected FSHD1 patients are obtained from three sources: FNMC, Genetic and Myopathy Group (branches of the Neurology Society of the Chinese Medical Association), and “FSHD-China” (an organization supported by FSHD patients). The study will firstly generate a data analyses concerning on estimated prevalence and disease progression/disability (with independent ambulation loss or wheelchair dependence); and some other several data analyses, nested as a sub-study within the CF1R, are expected to be generated via a subsequent protocol amendment.

Conclusion:

The data collected within this rare-disease patient registry (CF1R) will be utilized to synthesise real-world data on epidemiological information, disease progression/disability, and clinical and (epi)genetic information which are vital for adequate symptomatic management and clinical trial-readiness of FSHD1 in China.

Ethics and dissemination

The study was granted ethics committee approval by the First Affiliated Hospital of Fujian Medical University (MRCTA, ECFAH of FMU [2020]026) in 2020. Findings will be disseminated via presentations and peer-reviewed publications.

Trial registration: clinicaltrials.gov NCT04369209

Keywords: Facioscapulohumeral muscular dystrophy type 1 (FSHD1), Patient registry, Rare disease, PFGE-based Southern blotting, Prevalence, Disease progression

1. Introduction

Facioscapulohumeral muscular dystrophy type 1 (FSHD1) is one of rare diseases and is one of the most common inherited muscular dystrophies with autosomal dominance.[1] The core phenotype of FSHD is characterized by progressive weakness of the facial (facio-) and shoulder-upper arm (scapulo-humeral-). Muscle wasting often progresses slowly and typically in a descending pattern to include axial-, distal anterior leg, and hip girdle muscles.[2, 3] There is a large variability in the severity of symptoms, ranging from nonpenetrant/asymptomatic to disability with independent ambulation loss or even wheelchair dependence.[1, 4, 5]

FSHD1 has a unifying genetic model caused by a contraction to 1-10 D4Z4 macrosatellite repeats on a 4qA-specific FSHD1-permissive haplotype,[6-8] which is sensitively identified by pulsed-field gel electrophoresis (PFGE)-based Southern blotting.[9, 10] FSHD is linked to epigenetic derepression that lead to chromatin relaxation of the D4Z4 macrosatellite repeats to transcribe full-length *DUX4*, with transcripts stabilised by splicing to a poly(A) signal on specific 4qA haplotypes. Inappropriate expression of full-length *DUX4* and activation of its target genes is thought to be the major contributor to FSHD pathology, although the exact pathophysiological mechanism is still largely unknown.[7, 11] Expansion of our knowledge on the "DUX4opathy" models[12] has led to advance in developing (targeted) therapeutic strategies, thus, clinical trial readiness is now important. All of the aspects of patient registries, biomarkers, and clinical outcome measures are significant for clinical trial readiness in rare diseases such as FSHD, particularly patient registries,[13] which is one of the most effective methods of collecting all of the required large-scale patient data in rare diseases. Currently, there are 13 different countries established FSHD registries collecting data on over 3000 patients (clinical suspected and/or genetically confirmed patients),[13]such as Dutch[14] , UK[15] , US [16], Italian [17], French[18] , and etc. However, there is yet any reported cohort internationally of absolute genetically-confirmed FSHD1. In China, there is yet any FSHD registry, hence, there currently remains a dearth of data reporting on the epidemiology of FSHD (FSHD1), particularly in

terms of prevalence and disease progression/disability.

Here, we set up the China FSHD1 Registry (CF1R), which is a nationwide, population-based, longitudinal, and observational cohort clinical study focused on specific-disease of genetically-confirmed FSHD1 in all age groups, collecting data retrospectively at study entry and prospectively during follow up. The data to be collected is intended to help healthcare providers estimate the burden of and make important medical and financial decisions concerning FSHD in China, through an enhanced understanding of the prevalence, disease progression/disability, and natural history of genetically-confirmed FSHD1.

2. Objectives

2.1 Primary objectives

- To describe the methods of the China FSHD1 Registry to facilitate the development of other rare disease registries.
- To calculate true prevalence of genetically-confirmed FSHD1 in China basing on accurate diagnoses.
- To estimate disease progression of FSHD1 (resulting in disability with independent ambulation loss and/or wheelchair dependence).

2.2 Secondary objectives

- To establish a well-characterized baseline cohort for prospective follow-up and recruitment for future clinical trials.
- To collect long-term outcomes of FSHD1.
- To collect biological samples for future biomarker studies.

3. Methods

3.1 Registry identification

This rare- and inherited-disease patient registry (CF1R) has been registered with clinicaltrials.gov, under the

identifier NCT04369209.

3.2 Study design

The CF1R study is a nationwide, population-based, longitudinal, non-interventional, and a combination of retrospective and prospective observational cohort clinical study. The study is performed on genetically-confirmed FSHD1 patients or families (with at least 1 affected member) accrued through Fujian Neuromedical Center (FNMC); it is a diagnosis center for clinical-genetic FSHD1 in China, the only one employing PFGE-based Southern blotting, providing genetic tests for all suspected clinical FSHD1 patients/families.

3.3 Study population

To be eligible for **inclusion** in the CF1R study, participants must meet the following criteria:

- Male or female subjects of all ages at baseline living in China.
- Subjects with or without symptoms of FSHD, and genetic confirmation through PFGE-based Southern blotting: presenting at least one contracted D4Z4 macrosatellite repeats with 4qA-specific FSHD1-permissive haplotype.[6-8, 19]

Any patient meeting any of the following **exclusion** criteria at baseline cannot be included in the CF1R:

- Subjects or their legally designated representative are inability or unwillingness to provide informed consent.
- Subjects have severe complications.
- Subjects cannot cooperate with examination and follow-up.

3.4 Recruitment and screening

Participants will be recruited non-selectively and consecutively starting in January 2001. In order to

estimate the prevalence, we will try to reach all genetically-confirmed FSHD1 in China through a very intensive recruitment from three sources distributed across all six spatial zones in China. Genetic and Myopathy Group (branches of the Neurology Society of the Chinese Medical Association; https://www.cma.org.cn/) as well as "FSHD-China" (an organization supported by FSHD patients; http://www.fshd-china.org/), two of the three sources, will provide clinically suspected participants with FSHD as far as possible to FNMC, the third source, to access genetic test using PFGE-based Southern blotting; and FNMC will directly recruits participants through outpatients/inpatients to perform genetic test. For this study, we defined clinically suspected FSHD as: clinical weakness of the facial (facio-) and/or shoulder-upper arm (scapulo-humeral-) muscles for which the patient sought medical attention with exclusion of other diagnoses. If the diagnosis is genetically confirmed, the patients can be included in the study. All available data pertaining to demographics, clinical and (epi)genetic, and disease progression/disability will be collected within the CF1R Database at FNMC, and baseline data entries are mandated (**Fig. 1**).

3.5 Determination of sample size

As this is a combination of retrospective and prospective observational register of patients with genetically-confirmed FSHD1, there is no specific sample size based on statistical considerations. Since its initially performing PFGE-based Southern blotting in 2001, FNMC has completely finished 1,744 genetic tests (total number tests 1,802) for clinical suspected FSHD patients/families, and 997 patients were genetically confirmed, of whom were eligible participants appropriate to enable robust research and would be to be conducted on the data contained within the CF1R. The first results concerning prevalence and disease progression are expected at the end of 2021.

3.6 Patient and public involvement

Patients and the public are not directly involved in the development of the research question or in the design of the study.

3.7 Ethics

The CF1R is run following the recommendations from the Declaration of Helsinki and has received its Institutional Review Board approval in January 2020 (Ethics Committee of the First Affiliated Hospital of Fujian Medical University; approval No.: MRCTA, ECFAH of FMU 20201026). All eligible patients with genetically-confirmed FSHD1 at FNMC are provided with informed consent. Once a patient or their legal representative have agreed to participate and signed informed consent, they are enrolled in the CF1R study. The consent form allows for data and samples collection, future contact, for both communication and research, to be made and for the additional data to be entered by the data entry officers. Participants are able to withdraw their consent at any time. All data remain the property of the patient themselves and can be withdrawn at any time.

4. Biobank management

4.1 Determination of standard operating procedure (SOP)

An additional goal of this study is to obtain a well-documented cohort for long-term follow-up and future trial acquisition. Patients and their family members (if available) will be asked to participate in the CF1R study and for biomaterial storage. Hence, standardized written format of policies and procedures, namely, standard operating procedures (SOPs), must be made to ensure that all biological samples are collected, processed, and stored uniformly in FNMC Biobank.[20, 21]The SOPs in this CF1R study is developed strictly meeting the requirements of the Quality Assurance (QA) Program (also termed a Quality Management System, QMS) and uniquely depending on the characteristics of FSHD1. The original version of SOPs for CF1R study was drafted firstly by the Principal and Co-investigators; and then it was evaluated by all

members of the Research Committee (or Principal and Co-investigators), logistics company, and the Biobank of Neurogenctic/neuromuscular Diseases established by FNMC to determine the finally suitable SOPs.

4.2 Samples collection, processing, and storing

Samples of fresh peripheral blood cells (PBCs), serums, and muscle tissue (if muscle biopsy were performed) are directly obtained from eligible patients/families who sought medical attention at FNMC of the First Affiliated Hospital. To avoid degradation, fresh PBCs and serums are collected in vacutainer tube containing K2EDTA (Changgeng, China) and BD SST tube with silica clot activator, polymer gel, silicone-coated interior (Becton, Dickinson and Company, China), respectively.[22] 10.0ml PBCs will be drew from participants by specialized clinician according to SOP, which does not interfere routine diagnosis and treatment, and does not violate patient's rights and interests.[23] For the clinical suspected patients/families living in remote areas who choose to mail PBCs to FNMC for accessing genetic test, we will provide a specific vacutainer tube, namely, Cell-Free DNA BCT® RUO (Streck, USA), containing anticoagulant K3EDTA and cell preservative within liquid medium. This guarantees cellular genomic DNA preserving stability for up to 14 days at temperatures 6° C - 37° C,[24] and allows convenient transportation. Except those relatively remote provinces and municipalities located in west China (e.g., Tibet and Xinjiang), mailing samples (10.0ml PBCs) will arrive FNMC within 3 days.

To extracted high-quality DNA and gel DNA (high molecular weight DNA), these fresh PBCs (through direct drawing or mailing) will be either processed immediately or within 2 to 4 hours. The latter samples (delayed processing) must be preserved at relative low temperature at 4° C (not frozen) according to SOP. For the 10.0ml PBCs, 6.0ml will be extracted into gel DNA (embedded in low-melting agarose gel; SOP can be seen in **Appendix materials 1**); 1.0ml will be performed for routine DNA extraction (for Next Generation Sequencing or Sanger Sequencing, and Methylation test); and the remaining 3.0ml will be stored frozen at

-80° C for prospective collection. All of the PBCs, serums, muscle tissue, DNA, and gel DNA are proactively stored in the FNMC Biobank; the first four are stored frozen at -80° C and the last are stored refrigerated at 4℃.

4.3 Quality control and sample management

As a part of QA program, Quality Control (QC) is a system of technical activities measuring attributes and performance of a process or item, against defined standards.[21] Various forms of QC are applied to ensure that the both samples and related information meet requirements of investigators.[25, 26] In this study, QC of the routinely extracted DNA is concentration and purity test, measured by Nanodrop 2000 (Thermo-scientific, USA); while QC of gel DNA will be performed during PFGE-based Southern blotting. In addition, Research Committee (Principal and Co-investigators) will receive a QC report concerning nonconforming samples from the FNMC Biobank, and make a decision on sample destruction.

Sample storage in our study is managed by Biobank Information Management Systems (BIMS; Haier, China), basing on cloud platform. Accurate identification of samples is of critical importance to support research.[23] BIMS generate an unique code attaching to different samples for whole process. This code consists of sample type, diagnosis, and item category. Through clicking the link between sample and Biobank, we can obtain demographic and historical medical information of participants. For samples accessing and using, “Three Forms”, including EX-Biobank application form, IM-Biobank form of residual sample after test, and IM-Biobank form of test data, are required. Through this, we can trace samples trajectory with their unique code. BIMS can also records freezing and thawing time, samples numbers, and etc. In addition, an automatic temperature monitoring management system (ATMMS; Gsino Science&Technology, China) is utilized to continually monitor temperature of all equipment in FNMC Biobank.[21] Alarm system within ATMMS will give an alarm and activate staff “on call” when it detect abnormal temperature.

4.4 Genetic test (PFGE-based Southern blotting)

Laboratory SOP for PFGE-based Southern blotting are as followed.[27, 28] Five micrograms of high molecular weight DNA (embedded in gel) were digested with enzymes of *Eco*RI/*Hind*III (or *Eco*RI) and *Eco*RI/*Hind*III/*Bln*I (or *Eco*RI/*Bln*I) for array sizing; for analysis of the allelic variation on 4qter, DNA was digested with *Hind*III only (*EcoR*I and *Hind*III were purchased from New England Biolabs, and *Bln*I was purchased from Takara Bio). DNA was separated by PFGE. After blotting to an Amersham Hybond-N+ Membrane (GE Healthcare), the DNA was hybridised with probes p13E-11 (D4F104S1), 4qA (9B6A), and 4qB. After exposure, the *Eco*RI/*Hind*III (or *Eco*RI) array sizes were measured according to the MidRange PFG Marker (New England Biolabs) or CHEF DNA Size Standard (Bio-Rad Laboratories). The arrays were assigned to their respective chromosomes based on their *Bln*I sensitivity. The D4Z4 repeats number was calculated from the *Eco*RI fragment size using the formula: (D4Z4-containing fragment length (in kb, following *Eco*RI/*Hind*III)−4.7)/3.3 or (D4Z4-containing fragment length (in kb, following *Eco*RI digestion)−6.8)/3.3.[28]

4.5 Epigenetic test (Methylation)

We used the 4qA-allele-specific FasPAS primer (PAS-F 5'-GGATTTATAGGGAGGGGGTATTTTA-3'; PAS-R 5'-CTCCTAACRATCAAAAACATACCTCTATCTA-3') to investigate CpGs hypomethylation in a region distal to the DRA with sodium bisulfite sequencing (conducted by Genesky Biotechnologies, China).[29] For individuals carrying 4qA/B arrays, we specifically measured only one 4qA array; for individuals carrying 4qA/A arrays, we measured the average methylation on both 4qA chromosomes.[28]

5. Data collection

5.1 Data to be collected

When genetically-confirmed FSHD1 patients are included into the CF1R study, great care would be taken to

preserve the confidentiality of patient data, with patients being referred to by individual code numbers (e.g., FSHD_001_II1; 001 stands for family number and II1 stands for algebra). Both retrospective data [past medical history] and data available at the time of consent [baseline visit] are collected. Prospective data collected post-baseline at all routine clinical visits through outpatient service or telephone survey/remote video conferences by the same neurologists should be added periodically (at least annually) to the CF1R Database. As no investigations or data capture outside of routine clinical practice are mandated, only information which is available to the treating physician within this framework is recorded within the CF1R. A summary of the time-points at which specific data is collected is presented in **Table 1**, alongside the full list of agreed potential data fields and dictionaries contained within the CF1R (**Table 2**).

Briefly speaking, clinical assessments were structured according to our previous study. The clinical phenotypes were assessed by the 2016 Comprehensive Clinical Evaluation Form (CCEF) for FSHD.[30] Muscle strength was assessed bilaterally by manual muscle testing (MMT) in 14 muscle pairs (deltoid, triceps, biceps, wrist extensor, wrist flexor, finger extensor, finger flexor, gluteus maximus, gluteus medius, iliopsoas, hamstring, quadriceps femoris, tibialis anterior and gastrocnemius muscle); and we calculated the Medical Research Council (MRC) scale sum score by summing all of the average scores of each pair of muscles.[31] Muscle strength was also evaluated by FSHD clinical score.[32] Both the clinical severity score (CSS) and the age-corrected clinical severity score (ACSS) were adopted to determine disease severity; the ACSS is adjusted for the patient's age at diagnosis: ((CSS×2)/age at diagnosis)×1000.[33, 34] Motor function was assessed to record the maximum distance walked (in meters) in 6 minutes (6MWT).[35, 36]

Evaluation of disease progression/disability is mainly focusing on independent ambulation, which is measured by a simple 6-point assessment of modified Rankin Scale (mRS) due to its satisfactory reliability, only explicit criterion of walking, and idiosyncratic criteria to raters.[37] To date, longitudinal data has been up to 20 years (2001-2020) available, where the CF1R captured hard endpoints on onset age at first independent ambulation loss/or even first using of wheelchair. In addition, this CF1R Database can be

extended for future research.

5.2 Data collection tools and data management

The CF1R Database is a stand-alone and diseases-specific system, developed with Microsoft SQL Server database server. FNMC of the First Affiliated Hospital of Fujian Medical University manages the CF1R Database through a single online portal (https://168.2.5.202:9110) with an established Research Committee, including the Principal and Co-investigators, the data entry officers, the data custodian, the database program manager, and the data inspector. It approves and supervises projects that utilise the data held by the CF1R Database. The responsibilities include:

- Generate an investigation plan and ensure the accuracy, completeness, and legibility of paper case record form (CRF) (Principal and Co-investigators).
- Enter data from paper CRF into eCRF (the data entry officers).
- Oversee the confirmation, further investigation, and review of variables outliers, as required (the data inspector).
- Ensure timeliness of the data reported in the eCRF, post eCRF, and lock and export data (the database program manager).
- Data analysis and interpretation (the data custodian).

Patients' information is firstly collected in a purposely designed paper CRF. The CF1R Database can check data entered into eCRF automatically using logical checks (e.g., limits set within the database program). Additional requests for confirming or modifying questioned data may be generated through the eCRF, and the Principal investigator will be obliged to respond. Gene test report of PFGE-based Southern blotting is uploaded via a word/pdf secure file transfer process within the Database. Any corrections will overwrite the previous, initial information, and an audit trail allowing identification of any modifications will be maintained. A medical coding plan will specify the processes and the dictionaries used for coding. All

data coding will be done using internationally recognized and accepted dictionaries.

The web interface is housed on an IE Web Server at the First Affiliated Hospital of Fujian Medical University site. The CF1R Database is stored on file servers maintained by information department of the First Affiliated Hospital of Fujian Medical University. This department is responsible for the security of all hospital network servers, including firewalls, virus checking, network and workstation access passwords, and backup and disaster recovery. In addition to these security layers, the FNMC further safeguards data privacy by requiring individual application passwords and restricting access of confidential data to only those Research Committee members with a direct need. Individual identifiers are limited to a single data table with limited access by named registry personnel only.

The underlying source documents (paper CRF, informed consent, medical records, and etc.) are similarly protected. They are locked in file cabinets in secured areas at FNMC. File folders are stored in sample ID number order (rather than by name) to further protect them from unauthorized access.

5.3 Management and reporting of adverse events/adverse reactions

We will record details of adverse event reports associated with any treatment for FSHD. Adverse events will be coded with the Medical Dictionary for Regulatory Activities system and described by organ class. Cumulative adverse event information will be reported in the CF1R's interim and final analyses.

6. Statistical methods

All patients enrolled in the CF1R study are included in the data analysis set. Analysis as well as generation of plots and visualizations will be performed using SPSS software (version 25; IBM; USA) and GraphPad Prism (version 7; GraphPad Software, USA). Missing data will be addressed by deletion or through using of the last observation carried forward, determined on a case-specific basis. All continuous variables are described using standard statistical measures (e.g., number of observations, mean, standard deviation,

median, range, and minimum and maximum). All categorical data is summarized in frequency tables.

Patient demographic data and other baseline characteristics are all treated as described above. Since this study is population-based, one of primary objectives is to estimate prevalence that is calculated using the number of genetically-confirmed FSHD1 as the numerator, and the denominator is always set to the 2010 population census for China (1,332,810,869 persons; http://www.stats.gov.cn). Detailed prevalence data stratified by spatial zones, and provinces and municipalities are tabulated. The Kolmogorov-Smirnov test or Shapiro-Wilk's test is used to assess the normality of the variables. For normally distributed variables, group differences will be assessed by independent two-tailed Student's *t*-test or analysis of variance. For variables that are not normally distributed, the Mann-Whitney *U* or Kruskal-Wall *H* test will be used, as appropriate, or the data will be log-transformed for parametric testing. Categorical variables will be compared for the groups using the Chi-squared test (Fisher's exact test when the expected value is < 5). Single variable analysis is implemented to analyze the correlations between exposures, possible confounding conditions, and outcomes. Cumulative disability curves were used to assess disease progression with time span as independent ambulation loss since first-ever muscle weakness. A p value < 0.05 was considered statistically significant. Bonferroni correction will be applied to adjust for multiple comparisons.

7. Discussion

In the area of rare diseases, this China FSHD1 Registry (CF1R) is an invaluable source of collecting disease-specific, real-world data on epidemiological information, disease progression/disability, natural history, and clinical and (epi)genetic information which are vital for adequate symptomatic management and clinical trial-readiness.

To date, 1,744 genetic tests had been completely finished in FNMC (total tests number 1,802), and 997 (57.2%) patients were genetically confirmed with FSHD1. The data of this large number of enrolled patients

have been entered in the CF1R Database in a relatively short time (we have taken great care to preserve patient and data confidentiality) for prevalence and progression estimation, and that analysis, particularly in relation to that prediction of disease duration from first-ever muscle weakness to independent ambulation loss or wheelchair dependent may provide effective measures to delay the onset of disability. The CF1R allows for collection and analysis of longitudinal data, and next steps are to carry out an in depth analysis on this and extended data, in particular looking at the biomarkers of progression.

Some factors concerning the content of the CF1R Database have to be addressed. The present CF1R data are mostly retrospective because patients who were genetically confirmed with FSHD1 in the past and who are still attending FNMC have also been entered in the Database. The patients entered in the Database may be biased by regional distance from the clinic genetic center of FNMC or awareness of FSHD1. For example, CF1R includes more patients from Fujian province (location of FNMC) than other provinces and municipalities, especially those areas in west China. This bias will be resolved if more genetic test equipment set up, if more medical professionals in the area of FSHD1 train, and if stronger social awareness of FSHD1 develop. Despite these shortcomings, these data are important to alert health and political decision-makers to the possible underestimation of rare diseases, when emphasis is put on mortality, rather than on the prevalence of chronic diseases and disability rates.

In conclusion, CF1R Database is the first nationwide database for FSHD1 in China. It can serve several purposes, from patient recruitment for trials to estimation of the burden of FSHD1. The final goal is to collect uniformly information towards the global FSHD registry.

Declaration of competing interest

All authors report no conflict of interest.

Acknowledgment

We sincerely thank all participants for their contributions to this study and Professor Silvere Van der Maarel of Leiden University Medical Center for kindly providing the p13E-11, 4qA/4qB and B31 plasmids.

Figures

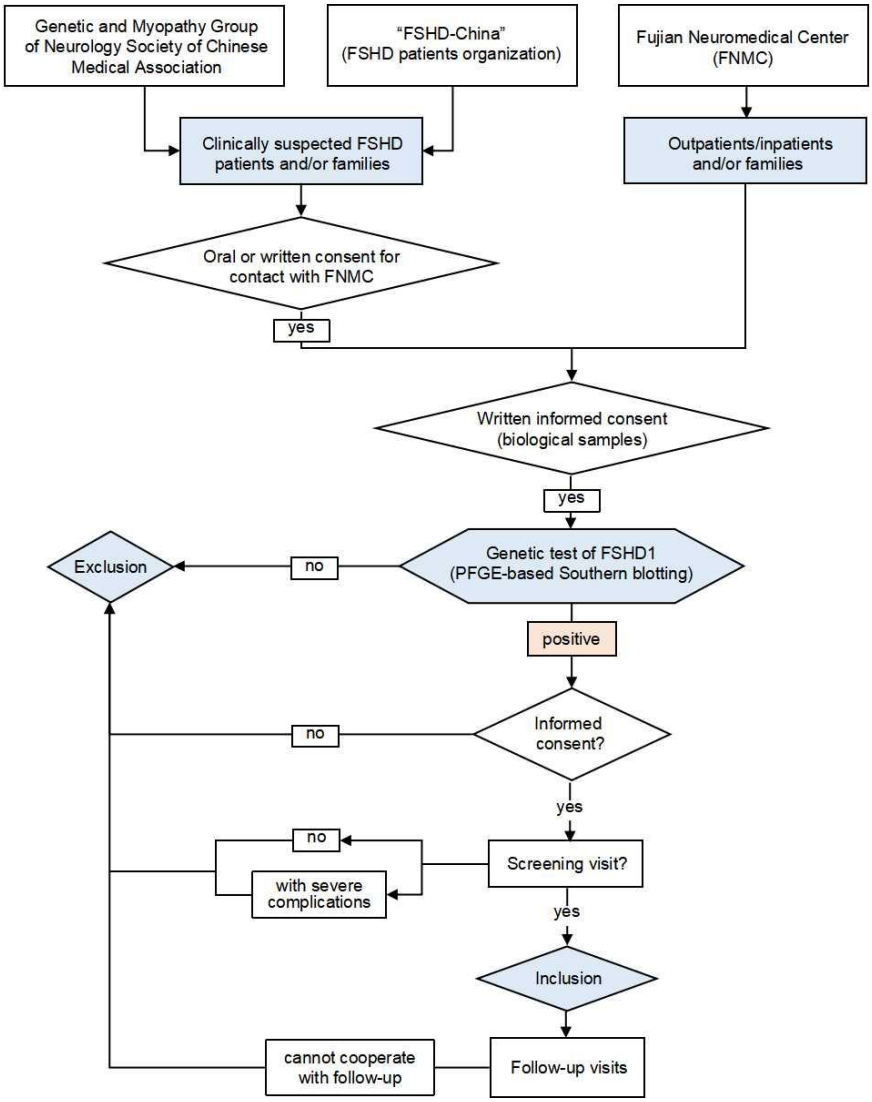

Fig. 1 Flow chart of inclusion

Tables

Table 1: Schedule of assessments for data recording (assessments in bold and italics are mandatory; all others are to be recorded on an if completed basis)			
Assessments	Study visit		
	Retrospective	***Baseline***	***Prospective***
Informed Consent(and/or assent, as applicable) [a]		×	×[a]
Confirm Eligibility		×	
Demographic information		×	
Medical history	×		×
(epi)Genetic test			
PFGE-based Southern blotting (D4Z4 repeats sizing)	×	×	×[b]
Sodium bisulfite sequencing (Methylation)	×	×	×[b]
Muscle function: MMT, FSHD CS, CSS/ACSS[c]	×	×	×
Motor function: 6MWT		×	×
Phenotypic classification: CCEF	×	×	×
Biological sample			
Peripheral blood cells		×	×[b]
Serums		×	×[b]
DNA		×	×[b]
RNA (in Trizol)		×	×[b]
Gel DNA (high molecular weight DNA)		×	×[b]
Muscle tissue[d]		×	×[b]
Disease progression			
Independent ambulation loss[e]		×	×
Wheelchair dependent		×	×

[a] Re-consent to adult registry consent when patient transitions from paediatric to adult patient

[b] Data/samples to be recorded during prospective visit if not available during retrospective or baseline visit

[c] Calculated as: ((CSS×2)/age at diagnosis)×1000

[d] Muscle tissuses would be preserved if muscle biopsy was performed

[e] Independent ambulation loss is measured by a simple 6-point assessment of modified Rankin Scale (mRS)

ACSS = Age-corrected CSS; CCEF = the 2016 Comprehensive Clinical Evaluation Form; CS = clinical score; CSS = Clinical severity scale; EMG = electromyography; FSHD = facioscapulohumeral dystrophy; MMT = manual muscle testing; PFGE = pulsed-field gel electrophoresis; 6MWT = 6-minutes walk test.

Note: Table 2 was located in the last of the manuscript.

References:

1. Wang, L.H. and R. Tawil, *Facioscapulohumeral Dystrophy.* Curr Neurol Neurosci Rep, 2016. **16**(7): p. 66.
2. Tawil, R. and S.M. Van Der Maarel, *Facioscapulohumeral muscular dystrophy.* Muscle Nerve, 2006. **34**(1): p. 1-15.
3. Mul, K., et al., *What's in a name? The clinical features of facioscapulohumeral muscular dystrophy.* Pract Neurol, 2016. **16**(3): p. 201-7.
4. Wohlgemuth, M., et al., *A family-based study into penetrance in facioscapulohumeral muscular dystrophy type 1.* Neurology, 2018. **91**(5): p. e444-e454.
5. Padberg, G.W.A.M., *Facioscapulohumeral disease*, in *Thesis Leiden, the Netherlands University of Leiden*. 1982.
6. Lemmers, R.J., et al., *Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere.* Nat Genet, 2002. **32**(2): p. 235-6.
7. Lemmers, R.J., et al., *A unifying genetic model for facioscapulohumeral muscular dystrophy.* Science, 2010. **329**(5999): p. 1650-3.
8. Wang, Z.Q., et al., *Distinguishing the 4qA and 4qB variants is essential for the diagnosis of facioscapulohumeral muscular dystrophy in the Chinese population.* Eur J Hum Genet, 2011. **19**(1): p. 64-9.
9. Tawil, R., et al., *Evidence-based guideline summary: Evaluation, diagnosis, and management of facioscapulohumeral muscular dystrophy: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology and the Practice Issues Review Panel of the American Association of Neuromuscular & Electrodiagnostic Medicine.* Neurology, 2015. **85**(4): p. 357-64.
10. DeSimone, A.M., et al., *Facioscapulohumeral Muscular Dystrophy.* Compr Physiol, 2017. **7**(4): p. 1229-1279.
11. Ricci, G., M. Zatz, and R. Tupler, *Facioscapulohumeral Muscular Dystrophy: More Complex than it Appears.* Curr Mol Med, 2014. **14**(8): p. 1052-1068.
12. Banerji, C.R.S. and P.S. Zammit, *Pathomechanisms and biomarkers in facioscapulohumeral muscular dystrophy: roles of DUX4 and PAX7.* EMBO Mol Med, 2021. **13**(8): p. e13695.
13. Tawil, R., et al., *Clinical trial preparedness in facioscapulohumeral muscular dystrophy: Clinical, tissue, and imaging outcome measures 29-30 May 2015, Rochester, New York.* Neuromuscul Disord, 2016. **26**(2): p. 181-6.
14. Mul, K., et al., *225th ENMC international workshop:: A global FSHD registry framework, 18-20 November 2016, Heemskerk, The Netherlands.* Neuromuscul Disord, 2017. **27**(8): p. 782-790.
15. Evangelista, T., et al., *Design, set-up and utility of the UK facioscapulohumeral muscular dystrophy patient registry.* J Neurol, 2016. **263**(7): p. 1401-8.
16. Hilbert, J.E., et al., *If you build a rare disease registry, will they enroll and will they use it? Methods and data from the National Registry of Myotonic Dystrophy (DM) and Facioscapulohumeral Muscular Dystrophy (FSHD).* Contemp Clin Trials, 2012. **33**(2): p. 302-11.
17. Nikolic, A., et al., *Clinical expression of facioscapulohumeral muscular dystrophy in carriers of 1-3 D4Z4 reduced alleles: experience of the FSHD Italian National Registry.* BMJ Open, 2016. **6**(1): p. e007798.http://www.fshd.it/.
18. Guien, C., et al., *The French National Registry of patients with Facioscapulohumeral muscular dystrophy.* Orphanet J Rare Dis, 2018. **13**(1): p. 218.
19. Lin, F., et al., *New Insights into Genotype-phenotype Correlations in Chinese Facioscapulohumeral Muscular Dystrophy: A Retrospective Analysis of 178 Patients.* Chin Med J (Engl), 2015. **128**(13): p. 1707-13.
20. Bell, W.C., K.C. Sexton, and W.E. Grizzle, *Organizational issues in providing high-quality human tissues and clinical information for the support of biomedical research.* Methods Mol Biol, 2010. **576**: p. 1-30.
21. *2012 best practices for repositories collection, storage, retrieval, and distribution of biological materials for research international society for biological and environmental repositories.* Biopreserv Biobank, 2012. **10**(2): p. 79-161.
22. Elliott, P. and T.C. Peakman, *The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine.* Int J Epidemiol, 2008. **37**(2): p. 234-44.
23. Kozlakidis, Z., C. Seiler, and D. Simeon-Dubach, *ISBER Best Practices Fourth Edition: A Success Story.* Biopreserv Biobank, 2018. **16**(3): p. 242-243.
24. Holland, N.T., et al., *Biological sample collection and processing for molecular epidemiological studies.* Mutat Res, 2003. **543**(3): p. 217-34.

25. Grizzle, W.E., et al., *Providing human tissues for research: how to establish a program.* Arch Pathol Lab Med, 1998. **122**(12): p. 1065-76.
26. Grizzle, W.E. and K.C. Sexton, *Development of a facility to supply human tissues to aid in medical research*. Molecular Pathology of Early Cancer. Vol. charpter 24. 1999: Amsterdam. p.371-383.
27. Wu, Z.Y., et al., *FSHD in Chinese population: characteristics of translocation and genotype-phenotype correlation.* Neurology, 2004. **63**(3): p. 581-3.
28. Qiu, L., et al., *Clinical and genetic features of somatic mosaicism in facioscapulohumeral dystrophy.* J Med Genet, 2020. **57**(11): p. 777-785.
29. Calandra, P., et al., *Allele-specific DNA hypomethylation characterises FSHD1 and FSHD2.* J Med Genet, 2016. **53**(5): p. 348-55.
30. Ricci, G., et al., *A novel clinical tool to classify facioscapulohumeral muscular dystrophy phenotypes.* J Neurol, 2016. **263**(6): p. 1204-14.
31. Mah, J.K., et al., *A multinational study on motor function in early-onset FSHD.* Neurology, 2018. **90**(15): p. e1333-e1338.
32. Lamperti, C., et al., *A standardized clinical evaluation of patients affected by facioscapulohumeral muscular dystrophy: The FSHD clinical score.* Muscle Nerve, 2010. **42**(2): p. 213-7.
33. Ricci, E., et al., *Progress in the molecular diagnosis of facioscapulohumeral muscular dystrophy and correlation between the number of KpnI repeats at the 4q35 locus and clinical phenotype.* Ann Neurol, 1999. **45**(6): p. 751-7.
34. van Overveld, P.G., et al., *Variable hypomethylation of D4Z4 in facioscapulohumeral muscular dystrophy.* Ann Neurol, 2005. **58**(4): p. 569-76.
35. *ATS statement: guidelines for the six-minute walk test.* Am J Respir Crit Care Med, 2002. **166**(1): p. 111-7.
36. McDonald, C.M., et al., *The 6-minute walk test as a new outcome measure in Duchenne muscular dystrophy.* Muscle Nerve, 2010. **41**(4): p. 500-10.
37. Wilson, J.T., et al., *Improving the assessment of outcomes in stroke: use of a structured interview to assign grades on the modified Rankin Scale.* Stroke, 2002. **33**(9): p. 2243-6.

Appendix materials

Appendix materials 1: SOP of gel DNA (high molecular weight DNA) extraction

1. 6ml fresh or refrigerated (within 6 hours, at 4° C) peripheral blood cells (PBCs) was put into a 25ml centrifuge tube.
2. Gently invert centrifuge tube upside down 10 times for mixxing.
3. To lyse red blood cells (RBCs), the PBCs was added 18ml (triple volume of PBCs) commercial 10% Red Cell Lysis Buffer (TIANGEN) and incubated on ice for 15 minutes.
4. Centrifuge the mixxed liquid at 3800rpm at 4° C for 10 min, and gently discard supernatant (RBCs).
5. Sediments (white blood cells, WBCs) was resuspend with 24ml (quadruple volume of PBCs) 10% RBC pyrolysis buffer (155mM NH_4Cl, 10mM $KHCO_3$, 1mM EDTA; pH 7.4), and was incubated on ice for 5 minutes.
6. Repeat step 4.
7. Resuspend sediments with 10ml 0.01M washing buffer (0.25M EDTA and 0.75M NaCl; pH 7.4).
8. Centrifuge the mixture at 3000rpm at 4° C for 10 minutes and gently discard the supernatant.
9. Resuspend sediments with 10ml 0.01M PBS (phosphate buffered saline) (0.14M NaCl, 30mM KCl, 10mM Na_2HPO_4 and 2mM KH_2PO_4; pH 7.4).
10. Repeat step 8.
11. Resuspend sediments with 1ml 0.01M PBS, and transfer the mixture into a 2.0ml centrifuge tube.
12. Centrifuge the mixture (in the 2.0ml centrifuge tube) at 2000rpm at 4° C for 10 minutes and gently discard the supernatant.
13. Melt 1.4% agarose (InCert agarose, FMC) on water bath at 60° C.
14. Resuspend final sediments with 600ul 0.01M PBS, then add 600ul of melted 1.4% agaros (10ml PBCs ≈ (14-28)×10^6 cells ≈ 100μg DNA).
15. Transfer the above mixture to a plastic mold to obtain gel DNA (0.1ml mixxted liquid ≈ a piece of gel.
16. Cooling and solidification of gel DNA at 4° C at least for 30 minutes.
17. Add 1ml ESP buffer (0.5M EDTA and 0.1M sarcosine [SIGMA]; pH 9.0) to a new 2.0ml EP tube.
18. Transfer solid gel DNA to the mixxted EP tube, and add proteinase K (SIGMA Aldrich) to a final concentration of 2 mg/ml, incubate on water bath at 51° C for 48-72 hours (4 pieces of gel ≈ 1 ml ESP ≈ 140 μl proteinase K).
19. Wah solid gel DNA with 10% TE buffer (10mM Tris-HCl, 0.1mM EDTA; pH 7.4) at 4° C three times, 30 minutes each time.
20. Digest residual proteinase K with adding 10ul PMSF (Beyotime) to TE buffer and incubating on water

bath at 51° C twice times, 30minutes each time.

21. Wah solid gel DNA with 10 TE buffer.
22. Store solide gel DNA in 0.5M EDTA buffer (pH 8.0) at 4° C.

Table 2 All data fields and dictionaries within the China FSHD1 Registry (data is grouped under headings, followed by a listing of all individual data variables collected)

Variable Group Heading	Variables	Item type within eCRF	Data type within eCRF	Definition	Range/option enumeration value	Units	Mandatory?
Group 1: Basic information	Code of subject	Gap filling	Character	Code of enrolled subjects. Subjects are numbered in three dimensions according to the order of enrollment: beginning with the disease abbreviation FSHD, followed by the patient's family number and generation number, connected by the following horizontal line "_", for example: FSHD_001_II1, FSHD is the abbreviation of the name of the disease, 001 stands for family number (data collection is conducted by family for hereditary diseases), and II1 stands for algebra	FSHD_001_II1, FSHD_002_I1	-	Yes
	Code of DNA	Gap filling	Numeric	Code of DNA extracted and stored in the FNC Biobank	Integer	-	Yes
	Subject name	Gap filling	Character	Subject name	-	-	Yes
	Contact telephone	Gap filling	Numeric	A phone number that the subject or family members are currently using to reach the subject. Leave at least two numbers, including the patient and a relative.	Integer	-	If available
	First visit date	Date	Date	Date of first time asking for PFGE-based Southern blotting at FNMC	YYYY, MM	-	Yes
Group 2: Filtering rule	Genetically confirmed FSHD1?	Multiple choice	Classification	Eligible subjects were genetically confirmed patients by PFGE-based Southern blotting who presented a contraction to 1-10 D4Z4 repeats with a 4qA-specific FSHD1-permissive haplotype.	1-Yes (geneticallu confirmed FSHD1); 0-No	-	Yes
	Gene report of PFGE-based Southern blotting	File	Pdf file	Gene analysis report by PFGE-based Southern blotting	-	-	Yes
Group 3: Demographic information[a]	Gender	Multiple choice	Classification	Gender of subjects	1-Male; 0-Female; 99-Unkown	-	Yes
	Native place	Multiple choice	Classification	Place of birth of subjects	See the sheet named "Option enumeration values of native place", mainly enumerates regions in mainland China; 99-Unkown	-	Yes
	Nation	Multiple choice	Classification	Nation of subjects	1-Han; 2-Other 55 nations besides Han; 99-Unkown	-	If available
	Date of birth	Date	Date	Date of birth of subjects	YYYY, MM	-	If available
	Patient source	Multiple choice	Classification	Sources of subjects with suspected clinical FSHD who came to our center for PFGE-based Southern blotting gene testing	1-Outpatients in our hospital; 2-Inpatients in our hospital; 3-Patients from hospitals networked within Genetic and Myopathy Group; 4-"FSHD-China" Organization	-	Yes
Group 4: (epi)genetic information[a]	Length of the contracted D4Z4 repeats array	Gap filling	Numeric	Length of the contracted D4Z4 repeats array with a 4qA-specific FSHD1-permissive haplotype; the major locus for FSHD1 has been mapped to the subtelomere of chromosome 4q35 where it is associated with contractions of a polymorphic repeat array consisting of 3.3kb repeated element, designated D4Z4.	One decimal point	kb	Yes
	The D4Z4 repeats number (units)	Gap filling	Numeric	The D4Z4 repeats number was calculated from the EcoRI fragment size using the formula: (D4Z4-containing fragment length (in kb, following EcoRI/HindIII)–4.7)/3.3 or (D4Z4-containing fragment length (in kb, following EcoRI digestion)–6.8)/3.3	One decimal point	units	Yes
	Stratification of the D4Z4 repeats number	Multiple choice	Classification	Stratification of the D4Z4 repeats number (units) (stratified to 3 levels: 1-3 units; 4-6 units; 7-9 units)	1- (1-3 units); 2-(4-6 units); 3-(7-10 units)	-	If available
	Inheritance patterns	Multiple choice	Classification	Inheritance patterns of the contracted D4Z4 repeats array with a 4qA-specific FSHD1-permissive haplotype.	1-Paternal; 2-Maternal; 3-De novo; 4-Unkown	-	If available
	Mosaic FSHD1?	Multiple choice	Classification	Whether the contracted D4Z4 repeats array with a 4qA-specific FSHD1-permissive haplotype is mosaic? Mosaic FSHD1 generally carry both a normally sized ancestral and a contracted D4Z4 repeats array.	1-Yes ; 0-No	-	If available
	Length of the second D4Z4 repeats array (< 38kb?)	Multiple choice	Classification	Whether the length of the second 4q fragment 2 is less than 38.0 kb? The fragment in which are located on the homologous chromosome of the fragment of contracted D4Z4 repeats array with a 4qA-specific FSHD1-permissive haplotype is defined as second 4q fragment 2.	1-Yes ; 0-No	-	If available
	Allelic sequence of the second 4q fragment	Multiple choice	Classification	Allelic sequence of the second 4q fragment.	1-A ; 0-B	-	If available
	Date of genetic confirmation	Date	Date	Date of genetic confirmation by PFGE-based Southern blotting.	YYYY, MM	-	If available
	Age of genetic confirmation	Formula	Formula	Age of genetic confirmation was derived by deduct the date of birth from the date of genetic confirmation.	Integer	years old	If available
	Detecting the 4qA-allele-specific methylation level?	Multiple choice	Classification	Whether to detect the 4qA-allele-specific methylation level? Using the 4qA-allele-specific FasPAS primer to investigate CpGs hypomethylation in a region distal to the DRA with sodium bisulfite sequencing.	1-Yes ; 0-No	-	If available
	The 4qA-allele-specific methylation value[b]	Gap filling	Numeric data	The 4qA-allele-specific methylation level.	Integer	%	If available
Group 5: Medical history[a]	Age of first visit	Formula	Formula	Age of first visit was derived by deduct the date of birth from the date of first visit.	Integer	years old	If available
	Is symptomatic at the first visiting?	Multiple choice	Classification	Whether there are symptoms at the first visit.	1-Yes (Symptomatic); 0-No (Asymptomatic/Nonpenetrant)	-	If available
	Onset age of first-ever muscle weakness[c]	Gap filling	Numeric	Onset age of first-ever muscle weakness	Integer	years old	If available
	Stratification of onset age of first-ever muscle weakness[c]	Multiple choice	Classification	Stratification of onset age of first-ever muscle weakness.	1-(0<X<=10 years old); 2-(10<X<=20 years old); 3-(>20 years old)	-	If available
	Proband?	Multiple choice	Classification	Whether a proband or not? The first patient in a FSHD family to seek medical treatment was defined as the proband.	1-Yes ; 0-No	-	If available
	Family history positive?	Multiple choice	Classification	Is there a relative within three generations with similar symptoms of muscle weakness?	1-Yes ; 0-No; 99-Unkown	-	If available

Continued

Group 6: Clinical information[a]	MRC score-MMT	Multiple choice	Classification	Muscle strength was assessed bilaterally by MMT in 14 muscle pairs: deltoid, triceps, biceps, wrist extensor, wrist flexor, finger extensor, finger flexor, gluteus maximus, gluteus medius, iliopsoas, hamstring, quadriceps femoris, tibialis anterior and gastrocnemius muscle. Finally, the results MMT were transfered to MRC score according to the established rules, and we calculated the MRC sum score by summing all of the average scores of each pair of muscles.	0-0; 1-1; 2-2; 3-3$^-$; 4-3; 5-3$^+$; 6-4$^-$; 7-4; 8-4$^+$; 9-5$^-$; 10-5	point	If available
	FSHD CS	Multiple choice	Classification	A protocol that quantifies muscle weakness by combining the functional evaluation of six muscle groups affected in FSHD.	-	point	If available
	I. Facial weakness	Multiple choice	Classification	A part of FSHD CS.	0-no weakness; 1-moderate weakness; 2-severe weakness	-	If available
	II. Scapular girdle involvement	Multiple choice	Classification	A part of FSHD CS.	0-no involvement; 1-mild involvement; 2-arm abduction > 45° but < 180°; 3-arm abduction ≦45°	-	If available
	III. Upper limbs involvement	Multiple choice	Classification	A part of FSHD CS.	0-no involvement; 1-at least two muscles affeged with 3<MRC≦5; 2-at least two muscles affeged with ≦3	-	If available
	IV. Legs involvement	Multiple choice	Classification	A part of FSHD CS.	0-no involvement; 1-unable to walk on tiptoes or hs (only one task impaired) ; 2-unable to walk on tiptoes and heels (two tasks impaired)	-	If available
	V. Pelvic girdle involvement	Multiple choice	Classification	A part of FSHD CS.	0-no involvement; 1-able to walk and to climb stairs without support but abnormally because of anterior leg muscle weakness; 2-able to walk unaided, to climb stairs or to stand up from a chair with support and/or tibialis anterior and quadriceps 3<MRC≦5; 3-able to walk unaided but unable to stand up from a chair or to climb stairs without support/more than 12 seconds and tibialis anterior and quadriceps MRC ≦3; 4-able to walk with support; 5-wheelchair bound	-	If available
	VI. Abdominal muscle involvement	Multiple choice	Classification	A part of FSHD CS.	0-no weakness; 1-presence of Beevor's sign	-	If available
	CSS	Multiple choice	Classification	A scale to assess muscle weakness in various muscular districts, considering the spread of symptoms to pelvic and leg muscles.	1-grade 0.5; 2-grade 1; 3-grade 1.5; 4-grade 2; 5-grade 2.5; 6-grade 3; 7-grade 3.5; 8-grade 4; 9-grade 4.5; 10-grade 5	-	If available
	ACSS	Formula	Formula	the ACSS is adjusted for the patient's age at diagnosis: ((CSS×2)/age at diagnosis)×1000	One decimal point	point	If available
	CCEF	Multiple choice	Classification	The CCEF is a clinical tool to capture various phenotypes from classic FSHD to individuals with incomplete phenotype, or asymptomatic carriers as well as subjects with atypical signs for which alternative diagnoses may be supposed.	1-Category A1; 2-Category A2; 3-Category A3; 4-Category B1; 5-Category B2; 6-Category C1; 7-Category C2; 8-Category D1; 9-Category D2	-	If available
	6MWT	Gap filling	Numeric	The maximum distance walked in meters in 6 minutes.	One decimal point	meter	If available
Group 7: Disease progression history[a]	To be penetrant?[d]	Multiple choice	Classification	Whether nonpenetrant patients developed symptoms during follow-up.	1-Yes ; 0-No	-	If available
	Onset age at first penetrance[e]	Gap filling	Numeric	Onset age at first penetrance	Integer	years old	If available
	Independent ambulation loss?	Multiple choice	Classification	During follow-up, we defined independent ambulation loss basing on the modified Rankin Scale (mRS), a simple 6-point assessment that included reference to limitation in activity, with a grade at 4-5 that individuals were unable to walk without assistance.	1-Yes ; 0-No	-	If available
	Onset age at onset of independent ambulation loss[f]	Gap filling	Numeric	Onset age at onset of independent ambulation loss	Integer	years old	If available
	Duration of independent ambulation loss[f]	Gap filling	Numeric	Duration of independent ambulation loss was derived by deduct the onset age at first-ever muscle weakness from the onset age at onset of independent ambulation loss	Integer	years	If available
	Wheelchair dependent[f]	Multiple choice	Classification	Whether wheelchair dependent?	1-Yes ; 0-No	-	If available
	Onset age at first using wheelchair[g]	Gap filling	Numeric	Onset age at first using wheelchair	Integer	years old	If available

[a] hidden logic, presentation conditions: "whether or not genetically-confirmed FSHD1" "equal to" "Yes"
[b] hidden logic, presentation conditions: "whether or not to detect the 4qA-allele-specific methylation level" "equal to" "Yes"
[c] hidden logic, presentation conditions: "whether or not symptoms at the first visit" "equal to" "Yes (symptomatic)"
[d] hidden logic, presentation conditions: "whether or not symptoms at the first visit" "equal to" "No (asymptomatic/nonpenetrant)"
[e] hidden logic, presentation conditions: "whether nonpenetrant patients (asymptomatic/nonpenetrant) developed symptoms during follow-up" "equal to" "Yes"
[f] hidden logic, presentation conditions: "whether or not independent ambulation" "equal to" "No"
[g] hidden logic, presentation conditions: "whether or not wheelchair dependent" "equal to" "Yes"
Abbreviation: ACSS = Age-corrected CSS; CCEF = the 2016 Comprehensive Clinical Evaluation Form; CRF = case record form; CS = clinical score; CSS = Clinical severity scale; FSHD1 = facioscapulohumeral muscular dystrophy type 1; PFGE = pulsed-field gel electrophoresis; MMT = manual muscle testing; MRC = the Medical Research Council scale; 6MWT = 6-minutes walk test.