Internal Medicine
The Japanese Society of Internal Medicine
doi: 10.2169/internalmedicine.8633-21
Intern Med 61: 2281-2286, 2022
http://internmed.jp

【 ORIGINAL ARTICLE 】

# Predicting the CTG Repeat Size from a Single Spirometry Test Performed at Any Time during the Disease Course of Myotonic Dystrophy Type 1

Kazuto Katsuse[1,2], Kenichiro Sato[2,3], Nobuyuki Tanaka[1], Idai Uchida[1], Tatsushi Toda[2], Takashi Mikata[1] and Yasufumi Motoyoshi[1]

**Abstract:**

**Objective** In myotonic dystrophy type 1 (DM1), the CTG repeat size in the dystrophia myotonica protein kinase gene has been shown to correlate with disease severity and is a potential predictive marker for respiratory decline. However, genetic testing can be challenging in some clinical situations. We developed a simple formula for estimating the CTG repeat size using a single spirometry test in patients with DM1.

**Methods** In this single-center retrospective study, we reviewed 50 consecutive patients with genetically confirmed DM1 whose follow-up visits were at our hospital. The patients were randomly assigned to training and test analysis subsets. By applying a linear mixed model to the longitudinal spirometry results of the training set, we calculated the fixed effects on the annual respiratory decline. Subsequently, we derived a prediction formula to calculate the repeat size that incorporated %vital capacity (%VC) and the patient's age at the time of the spirometry evaluation; the results were validated by the test set.

**Results** A total of 157 spirometry tests were recorded. The fixed effects on the annual %VC decline were $\hat{\beta}$=-0.90. The derived formula [repeat size=-16.8×(age+%VC/0.90)+2663] had a moderate predictive performance with a mean coefficient of determination $\bar{R}^2$ of 0.41.

**Conclusion** The CTG repeat size in patients with DM1 can be potentially predicted using a simple formula based on a single spirometry test conducted at any time over the disease course. It can be useful as a supportive tool for advance care planning when genetic testing is not available.

**Key words:** myotonic dystrophy type 1, linear mixed model, CTG repeat, respiratory decline, prediction

**(Intern Med 61: 2281-2286, 2022)**
**(DOI: 10.2169/internalmedicine.8633-21)**

## Introduction

Myotonic dystrophy type 1 (DM1), caused by the CTG triplet repeat expansion on chromosome 19q13.3, is the most common form of adult-onset muscular dystrophy (1, 2). DM1 is characterized by a slowly progressive muscle weakness involving multiple organs, including the respiratory, cardiac, ocular, endocrine, and central nervous systems (3). Respiratory dysfunction is the leading cause of death in patients with DM1 (≥30-40%); therefore, regular monitoring of the respiratory function is essential in the management of these patients (4-6). The prediction of respiratory decline would be beneficial for advance care planning and decision-making regarding the timing of introduction of non-invasive ventilation.

Larger CTG repeats have been associated with a younger age at the disease onset and greater clinical severity (7, 8). Although longitudinal respiratory decline in patients with DM1 is not well-documented, a significant inverse correlation has been found between the repeat size and respiratory function (9-12), indicating that the former is a potential predictive marker for respiratory decline.

Although genetic testing for CTG repeat in DM1 cases is

[1]Department of Neurology, National Hospital Organization Shimoshizu National Hospital, Japan, [2]Department of Neurology, Graduate School of Medicine, The University of Tokyo, Japan and [3]Department of Neuropathology, Graduate School of Medicine, The University of Tokyo, Japan
Received for publication September 6, 2021; Accepted for publication November 23, 2021
Correspondence to Kazuto Katsuse, katsuse-tky@umin.ac.jp

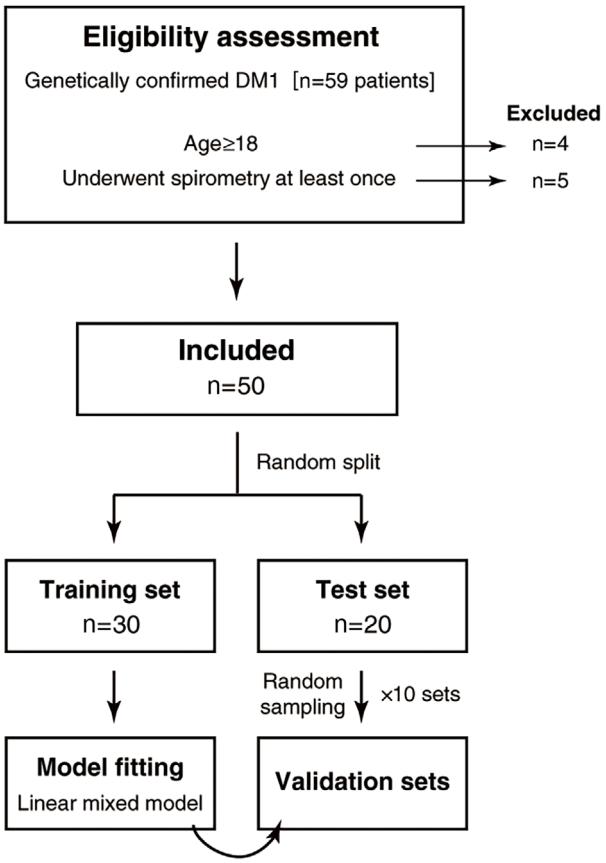


**Figure 1. Study flowchart.**

routine in developed nations, that service may not always be available in other areas (13). For example, a recent study from Malaysia mentioned inaccessibility of genetic testing in many Asian developing countries, which may cause a large number of diagnostic delays or underdiagnoses for DM1 (14). Even in developed countries where genetic testing is economically feasible, its usage is precluded by the stigma associated with incurable hereditary diseases (15). In fact, according to a survey conducted by the Japanese Society of Neurology in 2011, 43% of Japanese neurologists reported a negative attitude toward genetic testing for DM1, including reasons such as 'difficulty in providing psychological support to patients and their families after disclosure of the results' (16). In addition, the diagnosis of a typical case of DM1 is established predominantly through neurological and physiological examinations, with genetic testing being frequently omitted.

As an alternative to genetic testing, especially when it is not available, we aimed to derive a simple formula to predict the CTG repeat size based on the results of a single spirometry test and the patient's age at the time of the test. Given individual differences in the rate of respiratory decline, we employed a linear mixed model to analyze the longitudinal data of respiratory parameters and subsequently attempted to estimate the CTG repeat size using data from a single spirometry test performed at an arbitrary time during the disease course.

## Materials and Methods

### 2.1. Ethical approval

This study was approved by the Ethics Committee of Shimoshizu National Hospital in March 2021.

### 2.2. Study population

We retrospectively reviewed the medical records of all patients with genetically proven DM1, who presented for follow-up at a single center (The Neuromuscular Disease Center of National Hospital Organization Shimoshizu National Hospital, Japan) between January 2000 and July 2020. The requirement for informed consent was waived by the institutional review board because of the retrospective study design.

We included a total of 50 consecutive adult patients (≥18 years old) who had undergone at least 1 spirometry test (Fig. 1). We recorded data on the patients' age, sex, height, body mass index, brain natriuretic peptide level, and spirometry test results at follow-up visits.

### 2.3. The DM1 diagnosis

The DM1 diagnosis was based on a clinical evaluation and genetic confirmation of the CTG repeat expansion (≥50 repeats) in the 3' untranslated region of the dystrophia myotonica protein kinase (*DMPK*) gene, as assessed by Southern blotting or polymerase chain reaction (17). In cases with variations in the CTG repeat length (e.g., m≤repeat size≤n), the lower value (m) was used.

### 2.4. Spirometer test parameters

All spirometry tests were performed by a trained technician in a single session, with the patient in a seated position, according to the standards of the Japanese Respiratory Society (18). The slow vital capacity (VC) and forced vital capacity (FVC) were measured. We then calculated the percentage of VC (%VC) and FVC (%FVC) by comparing the measured values with the predicted ones obtained from the Japanese Respiratory Society.

### 2.5. Data analyses

#### 2.5.1. Data splitting

To determine whether or not our prediction formula could be generalized to new data, we randomly split the data from the 50 patients into 2 subsets: a training set (60%, n=30) and a test set (40%, n=20) (Fig. 1). For randomization, the "random.sample()" function from the Python (3.9.2) Standard Library was used.

#### 2.5.2. Model fitting

We used the training set to analyze changes over time in %VC and %FVC. Since the %VC and %FVC declined with age in patients with DM1 (Fig. 2), we used a linear mixed model to evaluate the longitudinal decline as fixed effects and incorporate individual differences in the rate of decline

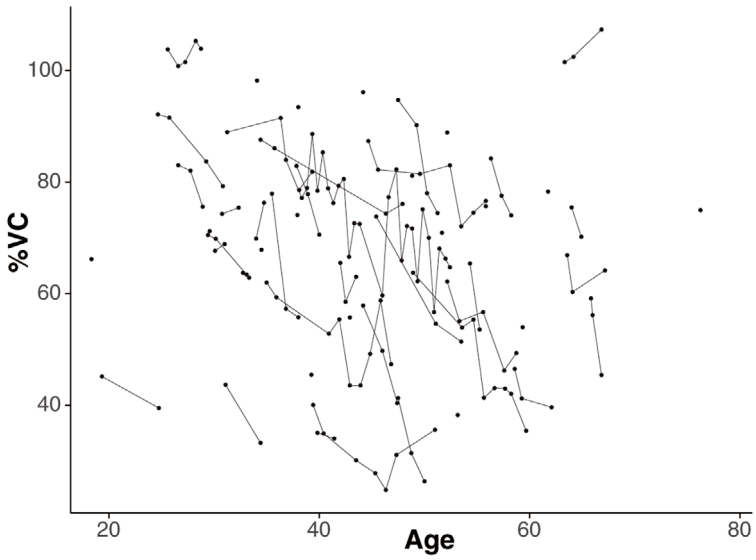


**Figure 2. The time series of %VC against age. Each line represents the longitudinal respiratory decline of each patient. Single points represent data of patients who underwent spirometry only once. A general trend of a decline in %VC with age can be observed.**

as random effects. Assuming that the $i^{th}$ patient completed an $s_i$ number of spirometry tests in total during the entire observation, the %VC or %FVC value of the $j^{th}$ ($j$=1, …, $s_i$) spirometry test of the $i^{th}$ ($i$=1, 2, …, 30) patient was denoted by $y_{ij}$, and the patient's age at the time of the $j^{th}$ test was denoted by as $x_{ij}$. Prior to fitting, the model was represented by $y_{ij} = (\beta + \varepsilon_i)x_{ij} + \delta_i$, where $\beta$ is a fixed effects parameter, and $\varepsilon$ and $\delta$ are random effects parameters reflecting individual differences. To estimate the repeat size $z_i$ for each patient $i$, we used the value of $\hat{\beta}$ calculated from the linear mixed model and the baseline ($j$=1) spirometry result $y_{i1}$ at an age of $x_{i1}$. Therefore, the linear regression model to be fitted was $z_i = \gamma(x_{i1} - y_i/\hat{\beta}) + \alpha$.

### 2.5.3. Validation of the model

We used the test set to validate the prediction formula $z = \hat{\gamma}(x - y/\hat{\beta}) + \hat{\alpha}$, where $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are fitted parameters from the training set (section *2.5.2.*). To examine whether or not this formula could be applied to spirometry results at an arbitrary time during the disease course, we randomly sampled a single result $y_{ij}$ ($j \in (1, \ldots s_i)$) from the longitudinal spirometry data for each patient $i$ ($i$=1, 2, …, 20) in the test set. This process was repeated 10 times at various time points to generate 10 different sets of randomly-sampled validation data. The accuracy of the model was determined using $R^2$, the coefficient of determination, which is a statistical measure of how well the data fit the regression line. We calculated $\bar{R}^2$, the mean value of $R^2$s obtained from 10 sets of the above validation data.

Statistical analyses were performed using the R software program, version 4.0.3. The "lmer()" function from lme4 package was used to fit the linear mixed model (19).

## Results

The demographic characteristics of the patients are shown in Table 1.

A total of 157 spirometry tests were recorded for the 50 enrolled patients (27 men; 54%). The average age at the first spirometry test (baseline) was 43.0±13.1 years old. Applying a linear mixed model to the longitudinal spirometry data of the training set, the fixed effects on the decline in %VC and %FVC were determined to be $\hat{\beta}_{\%VC}$=-0.90 (t=-3.28) and $\hat{\beta}_{\%FVC}$=-0.81 (t=-2.98), respectively. Using these values, we fitted the data of the training set to the linear regression model. The fitted parameters were $\hat{\gamma}_{\%VC}$=-16.8 [95% confidence interval (CI)=-20.4 to -13.3], $\hat{\alpha}_{\%VC}$=2664 (CI=2119 to 3118), $\hat{\gamma}_{\%FVC}$=-14.7 (CI=-18.2 to -11.2), and $\hat{\alpha}_{\%FVC}$=2517 (CI=2058 to 2976) (Table 2). The prediction formulas for the CTG repeat size $z$ based on %VC and %FVC were thus determined as $z = -16.8 \times (\text{age} + \%VC/0.90) + 2663$ and $z = -14.7 \times (\text{age} + \%FVC/0.81) + 2517$, respectively. These formulas were validated by the test set (Fig. 3). The $\bar{R}^2$ value for both formulae was 0.41.

## Discussion

We applied a linear mixed model to the longitudinal spirometry data of the training set and derived a formula to estimate the CTG repeat size using a respiratory parameter (% VC or %FVC) obtained from a single spirometry test and the patient's age at the time of the test. This estimation was validated by the spirometry results in the test set at randomly chosen time points over the follow-up period. To our knowledge, this is the first study to estimate the genotype of a patient with DM1 using a single spirometry test.

A linear mixed model framework is often used to model the longitudinal data of patients with individual differences (20). The fixed effects on the annual decline in %VC and %FVC were determined as $\hat{\beta}_{\%VC}$=-0.90 and $\hat{\beta}_{\%FVC}$=-0.81 (%/year), respectively. Our results showed a similar

**Table 1. Demographic Characteristics of Patients within the Training and Test Sets.**

| | | Training set | Test set |
|---|---|---|---|
| n | 50 | 30 (60%) | 20 (40%) |
| Sex (male) | 27 (54%) | 16 (53%) | 11 (55%) |
| Age at baseline [y/o] | 43.0±13.1 | 41.5±12.8 | 45.4±13.6 |
| Body height [cm] | 161.9±9.1 | 162.1±9.7 | 161.6±7.9 |
| Body mass index [kg/m$^2$] | 22.4±4.8 | 22.5±4.9 | 22.3±4.7 |
| CTG repeat size ($z$) | 610.8±555.6 | 657.3±612.6 | 541.0±463.3 |
| 50≤$z$<200 | 16 | 9 | 7 |
| 200≤$z$<500 | 6 | 3 | 3 |
| 500≤$z$<1,000 | 18 | 12 | 6 |
| 1,000≤$z$ | 10 | 6 | 4 |
| Follow up period [year] | 5.64±4.34 | 5.49±4.77 | 5.87±3.71 |
| Total number of spirometry completed in each patient [times/person] | 3.14±4.02 | 3.13±4.92 | 3.15±2.06 |
| %VC at baseline | 70.9±17.5 | 70.0±18.0 | 72.2±17.0 |
| %FVC at baseline | 70.0±19.0 | 69.1±18.1 | 71.3±20.6 |
| Invasive or non-invasive ventilator use initiated during follow up period | 7 | | |
| Death during follow up period | 2 | | |
| BNP [pg/mL] | 32.5±105.4 | | |

BNP: brain natriuretic peptide, %FVC: percentage of forced vital capacity, %VC: percentage of vital capacity
The data are expressed as the mean±standard deviation. The mean BNP was calculated using initial value for each patient.

**Table 2. Fitted Parameters for the Equation $z = \gamma(x_{i1} - y_i/\beta) + \alpha$ and Their Validation.**

| | $\hat{\beta}$ | $\hat{\alpha}$ | $\hat{\gamma}$ | $\bar{R}^2$ |
|---|---|---|---|---|
| %VC | -0.90 | 2,664 [2,119, 3,118] | -16.8 [-20.4, -13.3] | 0.41 |
| %FVC | -0.81 | 2,517 [2,058, 2,976] | -14.7 [-18.2, -11.2] | 0.41 |

%FVC: percentage of forced vital capacity, %VC: percentage of vital capacity
Square brackets indicate 95% confidence interval.

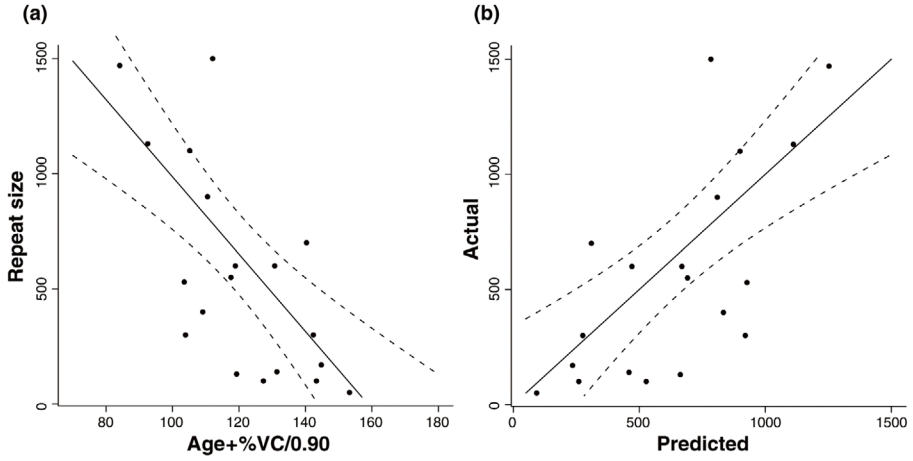


**Figure 3. (a) An example of the validation process. Each point represents the randomly sampled longitudinal spirometry data for each patient in the test set. The solid line represents the fitted line from the training set, (repeat size, $z$) = −16.8 × (age + %VC/0.90) + 2663. The dotted lines represent the limits of the 95% confidence interval. (b) Actual versus Predicted plot.**

trend to those of previous longitudinal studies, wherein the annual decline in %VC was calculated as -1.57 (9) or % FVC as -0.72 (21). We consider this study to be unique, as the repeat size was shown to be linked to the subtraction of the fixed effects [age + %VC/0.90 ($\propto y - \hat{\beta}x$)]. The repeat size was positively correlated with a younger age at the disease

onset (smaller $x$) and severe decline in %VC (smaller $y$), which is consistent with the findings from previous studies (7-12).

The linear mixed model has enabled the prediction of the CTG repeat size from a single respiratory parameter measured at an arbitrary time, not limited to the time of the initial visit or the diagnosis. Previous studies have suggested that the respiratory decline in DM1 is associated not only with the repeat size but also with the severity of muscle (10) and cardiac involvement (22). Although we were unable to incorporate these features into a mixed model simultaneously due to the retrospective design of our study, our simple model achieved moderate fitting, with a mean $\bar{R}^2$ value of 0.41 for the test set. This level of fitting may not be sufficient for an accurate estimation required for diagnostic purposes, and it does not rule out the clinical importance of individualized care (23); nevertheless, our study suggests that the prediction formula using a single spirometry test may be a viable alternative tactic in terms of its simplicity and applicability at any time, as genetic testing is sometimes difficult to perform due to social, ethical, or economic concerns.

The limitations of this study are largely attributed to its single-center, retrospective design. The distribution of repeat lengths was biased, since our center catered mainly to an adult population. Consequently, congenital cases with very large repeat sizes (>3,000) were not included. Despite the simplicity of the linear formula that has been derived, a Poisson regression model with a log link function might be more suitable. Bias may also have been introduced by the number of follow-ups, as the fixed effects may have been heightened by cases with a greater number of follow-ups. In our retrospective study, data on the family history, limb muscle weakness, dysphagia, sleep apnea, cognitive impairment, and other potential complications of DM1 were not available due to missing or otherwise insufficient data, making the generalizability of the results uncertain. As spirometry tests were performed in a single session at each time point, there remains some concern about data reproducibility particularly in patients with potential cognitive dysfunction. Simultaneous arterial blood gas analyses, diaphragm ultrasonography, and phrenic nerve conduction studies may overcome this reproducibility concern regarding the spirometry test and thus increase the accuracy of the prediction. However, it may safely be said that our simple protocol is feasible, as it is often difficult to conduct multiple or invasive tests in clinical settings. Further studies with a larger patient cohort and longer follow-up period are required to determine whether or not our findings can be universally applied. A prospective design is also preferable to control for confounding factors that may affect the respiratory function, such as by simultaneously assessing the limb muscle and cardiac involvement.

In conclusion, we derived a formula that can potentially predict the repeat size of a patient with DM1 using a single spirometry test performed over the disease course. This prediction formula may be a feasible, simple alternative to genetic testing when such testing is not available.

**The authors state that they have no Conflict of Interest (COI).**

Kazuto Katsuse and Kenichiro Sato contributed equally to this work.

## References


1. Brook JD, McCurrach ME, Harley HG, et al. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. Cell **68**: 799-808, 1992.
2. Emery AE. Population frequencies of inherited neuromuscular diseases - a world survey. Neuromuscul Disord **1**: 19-29, 1991.
3. Udd B, Krahe R. The myotonic dystrophies: molecular, clinical, and therapeutic challenges. Lancet Neurol **11**: 891-905, 2012.
4. Mathieu J, Allard P, Potvin L, Prévost C, Bégin P. A 10-year study of mortality in a cohort of patients with myotonic dystrophy. Neurology **52**: 1658-1662, 1999.
5. Wahbi K, Porcher R, Laforêt P, et al. Development and validation of a new scoring system to predict survival in patients with myotonic dystrophy type 1. JAMA Neurol **75**: 573-581, 2018.
6. Hawkins AM, Hawkins CL, Razak KA, Khoo TK, Tran K, Jackson RV. Respiratory dysfunction in myotonic dystrophy type 1: a systematic review. Neuromuscul Disord **29**: 198-212, 2019.
7. Marchini C, Lonigro R, Verriello L, Pellizzari L, Bergonzi P, Damante G. Correlations between individual clinical manifestations and CTG repeat amplification in myotonic dystrophy. Clin Genet **57**: 74-82, 2000.
8. Groh WJ, Groh MR, Shen C, Monckton DG, Bodkin CL, Pascuzzi RM. Survival and CTG repeat expansion in adults with myotonic dystrophy type 1. Muscle Nerve **43**: 648-651, 2011.
9. Boussaïd G, Wahbi K, Laforet P, et al. Genotype and other determinants of respiratory function in myotonic dystrophy type 1. Neuromuscul Disord **28**: 222-228, 2018.
10. Rossi S, Marca GD, Ricci M, et al. Prevalence and predictor factors of respiratory impairment in a large cohort of patients with myotonic dystrophy type 1 (DM1): a retrospective, cross sectional study. J Neurol Sci **399**: 118-124, 2019.
11. Vivekananda U, Turner C. A model to predict ventilator requirement in myotonic dystrophy type 1. Muscle Nerve **59**: 683-687, 2019.
12. Mazzoli M, Ariatti A, Garuti GC, et al. Predictors of prognosis in type 1 myotonic dystrophy (DM1): longitudinal 18-years experience from a single center. Acta Myol **39**: 109-120, 2020.
13. Zhong A, Darren B, Loiseau B, et al. Ethical, social, and cultural issues related to clinical genetic testing and counseling in low- and middle-income countries: a systematic review. Genet Med **23**: 2270-2280, 2021.
14. Ambrose KK, Ishak T, Lian LH, et al. Analysis of CTG repeat length variation in the *DMPK* gene in the general population and the molecular diagnosis of myotonic dystrophy type 1 in Malaysia. BMJ Open **7**: e010711, 2017.
15. Uhlmann WR, Roberts JS. Ethical issues in neurogenetics. Handb Clin Neurol **147**: 23-36, 2018.
16. Yoshida K, Ohata T, Muto K, et al. Survey on the attitude toward genetic testing of neurologists certified by the Japanese Society of Neurology. Rinsho Shinkeigaku (Clin Neurol) **53**: 337-344, 2013 [in Japanese].
17. Prior TW. Technical standards and guidelines for myotonic dystrophy type 1 testing. Genet Med **11**: 552-555, 2009.
18. Clinical Pulmonary Functions Committee of the Japanese Respira-

tory Society. Guideline of respiratory function tests - spirometry, flow-volume curve, diffusion capacity of the lung. Nihon Kokyuki Gakkai Zasshi (J Jpn Respir Soc) **Suppl**: 1-56, 2004 [in Japanese].

**19.** Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw **67**: 1-48, 2015.

**20.** Faraway JJ. Chapter 10 and 11. In: Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. 2nd ed. Chapman and Hall/CRC, Boca Raton, FL, 2016.

**21.** Thil C, Agrinier N, Chenuel B, Poussel M. Longitudinal course of lung function in myotonic dystrophy type 1. Muscle Nerve **56**: 816-818, 2017.

**22.** Kaminsky P, Brembilla-Perrot B, Pruna L, Poussel M, Chenuel B. Age, conduction defects and restrictive lung disease independently predict cardiac events and death in myotonic dystrophy. Int J Cardiol **162**: 172-178, 2013.

**23.** Yetimakman AF, Bayrakçı B, Esquinas AM. Myotonic Dystrophy type 1, individualised respiratory care rather than standart prognostication. J Neurol Sci **401**: 125-126, 2019.