



Universiteit
Leiden
The Netherlands

Deriving reference values for nerve conduction studies from existing data using mixture model clustering

Reijntjes, R.H.; Potters, W.V.; Kerkhof, F.I.; Zwet, E. van; Rossum, I.A. van; Verhamme, C.; Tannemaat, M.R.

Citation

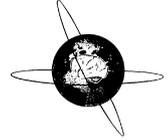
Reijntjes, R. H., Potters, W. V., Kerkhof, F. I., Zwet, E. van, Rossum, I. A. van, Verhamme, C., & Tannemaat, M. R. (2021). Deriving reference values for nerve conduction studies from existing data using mixture model clustering. *Clinical Neurophysiology*, 132(8), 1820-1829. doi:10.1016/j.clinph.2021.04.013

Version: Publisher's Version

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/3249515>

Note: To cite this publication please use the final published version (if applicable).



Deriving reference values for nerve conduction studies from existing data using mixture model clustering



R.H. Reijntjes^a, W.V. Potters^b, F.I. Kerkhof^a, E. van Zwet^c, I.A. van Rossum^a, C. Verhamme^b, M.R. Tannemaat^{a,*}

^a Department of Neurology, Leiden University Medical Center, Leiden, the Netherlands

^b Department of Neurology, Amsterdam UMC, University of Amsterdam, Amsterdam Neuroscience, Amsterdam, the Netherlands

^c Department of Biostatistics, Leiden University Medical Center, Leiden, the Netherlands

ARTICLE INFO

Article history:

Accepted 10 April 2021

Available online 21 May 2021

Keywords:

Nerve conduction studies

Mixture model clustering

Reference values

Clinical neurophysiology

HIGHLIGHTS

- We made reference values for nerve conduction studies with mixture model clustering.
- Precise, individual reference values were based on age, sex, height and temperature.
- Z-scores can be calculated to quantify the (ab)normality of a test result.

ABSTRACT

Objective: to obtain locally valid reference values (RVs) from existing nerve conduction study (NCS) data.

Methods: we used age, sex, height and limb temperature-based mixture model clustering (MMC) to identify normal and abnormal measurements on NCS data from two university hospitals. We compared MMC-derived RVs to published data; examined the effect of using different variables; validated MMC-derived RVs using independent data from 26 healthy control subjects and investigated their clinical applicability for the diagnosis of polyneuropathy.

Results: MMC-derived RVs were similar to published RVs. Clustering can be achieved using only sex and age as variables. MMC is likely to yield reliable results with fewer abnormal than normal measurements and when the total number of measurements is at least 300. Measurements from healthy controls fell within the 95% MMC-derived prediction interval in 97.4% of cases.

Conclusions: MMC can be used to obtain RVs from existing data, providing a locally valid, accurate reflection of the (ab)normality of an NCS result.

Significance: MMC can be used to generate locally valid RVs for any test for which sufficient data are available.¹

© 2021 International Federation of Clinical Neurophysiology. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Abbreviations: AUMC, Amsterdam University Medical Center; CMAP, Compound Muscle Action Potential; CDF, Cumulative Distribution Function; DML, Distal Motor Latency; LUMC, Leiden University Medical Center; MMC, Mixture Model Clustering; MN2SD, Mean \pm 2 Standard Deviations; NCS, Nerve Conduction Study; NCV, Nerve Conduction Velocity; RV, Reference Value; SD, Standard Deviation.

* Corresponding author at: Albinusdreef 2, 2333 ZA, Leiden, the Netherlands.

E-mail addresses: r.reijntjes@lumc.nl (R.H. Reijntjes), w.v.potters@amsterdamumc.nl (W.V. Potters), f.i.kerkhof@lumc.nl (F.I. Kerkhof), e.w.van_zwet@lumc.nl (E. van Zwet), i.a.van_rossum@lumc.nl (I.A. van Rossum), c.verhamme@amsterdamumc.nl (C. Verhamme), m.r.tannemaat@lumc.nl (M.R. Tannemaat).

¹ All code and a test dataset have been placed online (gitlab.com/lumc/clinicalneurophysiology/ReferenceValues).

<https://doi.org/10.1016/j.clinph.2021.04.013>

1388–2457/© 2021 International Federation of Clinical Neurophysiology. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Reference values (RVs) are of major importance for proper interpretation of clinical neurophysiological measurements, especially nerve conduction studies (NCS). Considerable debate remains about the use of RVs available in the literature and RVs collected in individual clinical neurophysiology departments. Publicly available RVs for NCS are rare: a recent systematic review only found one set of RVs of sufficient quality for a small number of measurements (Chen et al., 2016). In addition, universal application of publicly available RVs is hampered by the fact that they are likely to differ between clinical neurophysiology departments (Litchy et al., 2014; Brown et al., 2017), as they are influenced by

factors such as electrode placement and size, filter settings and temperature. Therefore, several guidelines recommend that clinics develop their own RVs (Dillingham et al., 2016; Stålberg et al., 2019). However, as RVs are influenced by patient-related factors including age, body height and sex (Salerno et al., 1998; Dinesh Kumbhare et al., 2015), such an approach requires an exhaustive battery of NCSs on a large number of healthy subjects. To our knowledge, very few clinical neurophysiology departments actually create locally valid RVs.

Of note, a large number of measurements taken during routine clinical practice are likely to be normal and could be used to develop RVs, if they can be identified as such. Here, we assumed that any large enough set of NCS measurements consists of a mixture of two populations: a population of “normal” measurements belonging to healthy nerves and a population of “abnormal” measurements. These populations can be separated by a statistical technique called mixture model clustering (MMC). Here we used supervised MMC with statistical models, which combines clustering and linear regression models. MMC separates the data and creates a model for both populations that is based on factors that are known to affect the measured value, such as age, sex, height and temperature. The use of linear regression models in MMC has three advantages: 1) it improves clustering precision, 2) the model representing “normal” measurements can be used to calculate individual RVs (or “MMC predicted values”) that are corrected for all relevant factors affecting the measurement and 3) the actual measured value can be compared to the MMC-predicted value based on these specific characteristics, which allows for the calculation of a z-score giving a precise quantification of the (ab)normality of a particular test result.

2. Methods

2.1. Patient data acquisition

All NCS were performed at Leiden University Medical Center (LUMC) and the Amsterdam University Medical Center (AUMC), location AMC. Both are tertiary referral centers for neuromuscular diseases. Details on stimulus and filter settings and electrode placement are provided as supplementary data (Appendix A). At the LUMC, all NCS and patient data from 1/1/2011–2/31/2017 were retrieved from reports. At the AUMC, all NCS data from 7/23/2008–11/18/2019 were retrieved directly from the EMG software using a custom-made script and anonymized. Recorded patient data included: patient age, sex, height and limb temperature. For motor conduction studies, we extracted distal motor latency (DML), negative peak compound muscle action potential (CMAP) amplitude, negative peak CMAP area, negative peak CMAP duration, nerve conduction velocity (NCV) of the most distal nerve segment in the arm and leg and minimal F-wave latency. In addition, we recorded the reason for referral and the conclusion from the report from LUMC data. At the AUMC, patient height was routinely recorded, but at the LUMC, this was usually only done when F-waves of leg nerves were recorded, as these were corrected for height (Buschbacher, 1999c, 1999b).

According to Dutch law, formal approval by a local ethics committee and individual informed consent is not required for retrospective research on anonymized data gathered exclusively for patient care. We focused on a set of commonly used motor nerve conduction measurements, for which widely accepted RVs have been published: DML, CMAP amplitude, CMAP area, CMAP duration, NCV and minimal F-wave latency of median, ulnar, tibial and peroneal nerves (Dinesh Kumbhare et al., 2015). Results for the peroneal nerve will be shown throughout the manuscript as an example.

2.2. Log transformation of NCS data

Prior to clustering analysis, CMAP amplitude and area data were transformed to a logarithmic scale, as histograms were not consistent with two normal distributions (Fig. 1A). Logarithmic transformation is not possible on measurements with a value of 0. To determine the effects of leaving out these measurements, we performed the clustering analysis twice on log-transformed LUMC data: once after removing all measurements with value 0, and once after replacing 0 with 0.01.

2.3. Mixture model clustering (MMC)

A mixture model is a probabilistic model to determine the presence of a predefined number of clusters. MMC has previously been used successfully to classify Alzheimer Disease patients based on cerebrospinal fluid biomarkers (De Meyer et al., 2010; Toledo et al., 2015). Model-based mixture models allow for the simultaneous probabilistic classification into clusters and estimation of regression models in each cluster. Here, we used the flexmix package (Leisch, 2004; Grün and Leisch, 2008) for the programming language R (R Core Team, 2019), using a linear iterative expectation–maximization model (Dempster et al., 1977). In the expectation step, cluster memberships are estimated using regression models resulting in each measurement having a cluster-specific probability. In a subsequent maximization step, unknown model parameters (intercept, slope, and variance) for each of the clusters are estimated by maximizing the cluster-specific log-likelihood and using the probabilities calculated in the expectation step as weights. This process is iterated, until the values converge (the video in Appendix B provides an animation visualizing a simulation of expectation and maximization steps performed during mixture model clustering). The result of the iterative process is a matrix where each column represents the number of clusters and the rows represent each of the measurements. This matrix specifies the probability that the measurement of a subject belongs to each of the clusters. As the predefined number of clusters was two, a probability of more than fifty percent assigned the measurement to that cluster.

Based on the literature and availability of data, we used the following variables as independent variables (Buschbacher, 1999b, 1999c): for measurements from the AUMC and measurements on leg nerves from the LUMC, we used sex, age, body height, temperature, interaction between sex and age and interaction between sex and body height. For measurements of arm nerves from the LUMC, we used sex, age, temperature and the interaction between sex and age, because height data were missing relatively frequently from these measurements.

After clustering, we designated clusters with higher values for CMAP amplitude, CMAP area and NCV and lower values for DML, CMAP duration and minimal F-wave latency as the normal cluster.

2.4. Comparison of MMC results with published RVs

For comparison with published RVs, we stratified normal clusters according to the sex, height, and age categories used by Buschbacher (Dinesh Kumbhare et al., 2015). We then calculated mean, third percentile, standard deviation (SD) and mean \pm 2SD (MN2SD) cut-off values of the normal cluster. Groups with fewer than 10 measurements were excluded from analysis.

2.5. Exploring robustness of MMC

We explored the effects of using multiple combinations of sex, age, height, and temperature and all possible interactions between these parameters in the clustering step of MMC for peroneal nerve

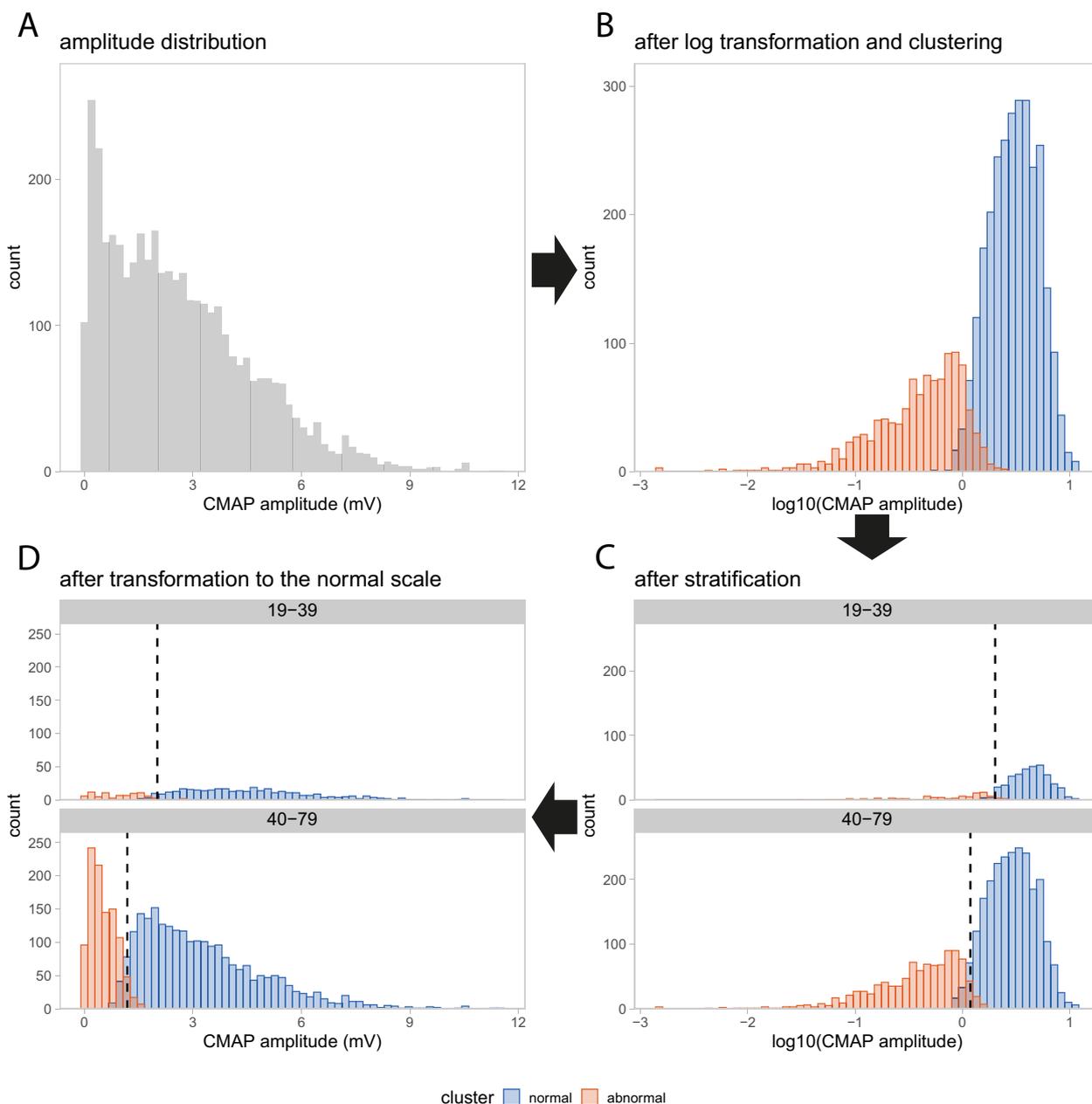


Fig. 1. Visual representation of the mixture model and stratification steps to create reference values. (A) raw data plotted in a frequency histogram, (B) data after log transformation and clustering, (C) stratification on the log scale based on age groups previously used by Buschbacher and (D) transformation of the data back to the normal scale. Importantly, mean values and mean-2SD cut-off values are calculated on log transformed data. Vertical dashed lines indicate the mean-2SD reference values. CMAP: Compound Muscle Action Potential.

CMAP amplitude data using AUMC data. Several linear models were tested. The most complicated model had main effect for sex, age, body height and temperature and interaction affect for sex and age and sex and body height. For each combination, we calculated the mean, SD, adjusted R^2 of the normal cluster and the percentage of abnormal measurements.

Next, we determined the robustness of MMC by adding increasing number of simulated abnormal data to existing peroneal nerve CMAP amplitude data from the LUMC. Simulated data were generated by creating a linear mixed model of the abnormal values and generating and adding values by running a simulation of this model using the simstudy package (Goldfeld, 2019), increasing the abnormal population two- and fourfold and calculating the effect on size, average, and standard deviation of the normal cluster.

Finally, we aimed to explore the effect of sample size on reliability of the obtained RV by performing a second simulation of 1500 normal and abnormal CMAP amplitude measurements of the peroneal nerve in a 4:1 ratio. We then calculated the MN2SD cut-off value at the lower end of the normal cluster. By performing this simulation 30 times, we calculated the mean and the cut-off value for a sample size of 1500 measurements. We then iteratively reduced the number of measurements by 100 and recalculated mean and SD of the MN2SD each time.

2.6. External validation

The AUMC possesses NCS RVs, derived from 26 healthy individuals as described previously (Verhamme et al., 2009). These mea-

measurements are not part of the larger database used for MMC. We used this independent control dataset to validate MMC-derived RVs. For this aim, MMC was used to generate statistical models for each measurement. We modelled the interaction between sex and age and between sex and body height. MMC-based predictions with a 90% prediction interval were made of each measurement and compared with the actual measured values. Because measurements are expected to only have a lower or upper bound (i.e. all data are expected to be abnormal in only one direction: lower than normal for CMAP amplitudes, CMAP areas and NCVs, and higher than normal for CMAP duration, DMLs, minimal F-wave latency), 95% of measured values from healthy controls are expected to fall in the MMC-predicted range.

2.7. Data availability

Data are available in a public, open access repository. We have placed all R code and a test dataset online as open source data (gitlab.com/lumc/clinicalneurophysiology/ReferenceValues).

3. Results

3.1. Demographics

Baseline data can be found in Table 1. In general, age and sex data were complete, temperature data were missing in 30% (LUMC) and 45% (AUMC). Height data was missing in 49% of NCSs at the LUMC.

3.2. Clustering and creating RVs

For all measurements, MMC resulted in two clear clusters (Fig. 1). In all cases, the normal clusters contained the highest number of values, showed a normal distribution and had a higher adjusted R², suggesting that these were indeed representative of “normal” measurements. In the normal cluster, significant effects of age, height, sex and temperature were present as expected from literature: e.g., CMAP decreased with age (data not shown).

There was a small mean difference of 0.6 mV between distributions of the normal clusters of peroneal nerve CMAP amplitudes of the LUMC and Amsterdam UMC ($p < 0.001$; Kolmogorov-Smirnov test) (Fig. 2).

After stratification according to the age and height classes used by Buschbacher (Dinesh Kumbhare et al., 2015), MMC-derived data from both hospitals were similar (Tables 2 and 3). Of note, MMC derived mean and cut-off value (MN2SD) for CMAP amplitudes were generally lower than those published by Buschbacher. MMC-derived 3rd percentile and MN2SD values for CMAP amplitudes were closer to each other than those published by Buschbacher. Replacing measurements with value 0 with 0.01 prior to log transformation did not affect the number, mean and SD of the normal cluster as all low measurements were grouped in the abnormal cluster (data not shown).

Appendix C provides estimates of the independent variables and other statistical results of the linear models for the normal clusters after MMC.

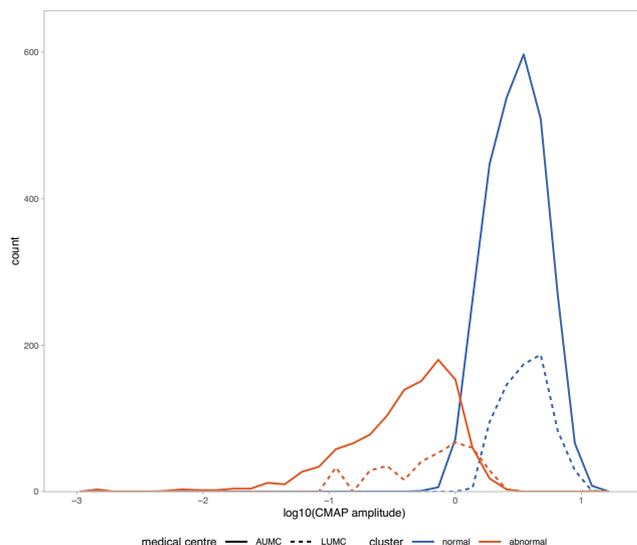


Fig. 2. Comparison of AUMC and LUMC clusters of peroneal nerve CMAP amplitudes. Although the distribution of the normal cluster differs significantly between both centers ($p < 0.001$; Kolmogorov-Smirnov test), the mean value and standard deviation of the normal cluster is similar. CMAP: Compound Muscle Action Potential. AUMC: Amsterdam University Medical Center. LUMC: Leiden University Medical Center.

3.3. Effect of different variables and changing cluster sizes

MMC of the peroneal nerve CMAP using only sex and age as variables resulted in two clear clusters in the AUMC data (Fig. 3). The addition of temperature and height as independent variables had limited effects on size, mean and fit of the normal cluster (e.g. adjusted R² was 0.17 using main effect for sex and age and interaction effect for sex and age, and 0.18 when data were clustered using main effect for sex, age, body height and temperature and interaction effect for sex and age and sex and body height).

We explored the effect of the number of abnormal measurements on the normal cluster by performing MMC on the peroneal nerve CMAP amplitude after adding simulated abnormal data to LUMC data (Fig. 4). Doubling the number of abnormal measurements had limited effects on size, mean and MN2SD of the normal cluster: MN2SD was 1.2 mV using the actual data set, 1.2 mV after doubling the abnormal cluster. However, increasing the abnormal cluster fourfold (resulting in clusters of approximately equal size) led to a change in the MN2SD of the normal cluster to 0.9 mV.

Finally, we explored the effect of sample size on reliability of the obtained value by iteratively reducing the number of measurements in a simulation from 1500 measurement to 200 measurements (Fig. 5). In this analysis, the mean cut-off was 1.3 mV and the SD was less than 0.1 mV. As expected, the SD of the obtained cut-off value gradually increases as the number of measurements used for MMC decreases, with a further increase in SD when the total number of measurements was lower than 300.

Table 1

Baseline characteristics. Values indicate mean ± standard deviation, unless stated otherwise. #Student’s t-test assuming unequal variances; §chi-square test of independence.

	AUMC	Missing (n)	LUMC	Missing (n)	p-value
Number of studies (n)	7546		5550		
Body height (cm)	173 ± 10	0	174 ± 10	2699	0.001#
Female sex, n (%)	3992 (53%)	0	3051 (55%)	0	0.020§
Age (years)	55.7 ± 14.6	0	56.3 ± 15.0	0	0.013#
Temperature (°C)	33.2 ± 1.8	3405	31.6 ± 1.4	1641	< 0.001#

Table 2

Buschbacher's RVs and MMC-derived RVs from the Amsterdam UMC and LUMC for arm nerves. Bb: Buschbacher, SD: standard deviation, DML: distal motor latency, NCV: nerve conduction velocity, AUMC: Amsterdam University Medical Center. LUMC: Leiden University Medical Center, N: number of measurements. Amplitude data expressed in mV, Area in mV.ms, DML, duration and F-wave latency in ms. Details with regards to settings and recording sites can be found in Appendix A. Mean ± 2SD and 3rd percentile columns: mean-2SD and 3rd percentile for CMAP amplitudes, CMAP areas and NCVs, mean + 2SD and the 97th percentile for CMAP duration, DML and minimal F-wave latency. Empty fields: insufficient data for MMC.

Nerve	Units of measure	Age (years)	sex	height (cm)	Mean			Mean ± 2SD			3rd percentile			N		
					Bb	AUMC	LUMC	Bb	AUMC	LUMC	Bb	AUMC	LUMC	AUMC	LUMC	
Median	Amplitude (mV)	19–39			11.9	6.1	5.3	4.7	3.1	2.4	5.9	3.2	2.5	382	48	
		40–59			9.8	5.2	4.9	4.2	2.4	2.2	4.2	2.5	2.1	1483	116	
		60–79			7.0	4.3	3.9	1.8	1.9	1.6	3.9	2.0	1.6	1457	123	
	Area (mV.ms)	19–49			37.4	18.4	17.9	11.6	8.8	8.1	14.6	8.7	7.9	914	84	
		50–59			30.9	15.6	15.7	13.7	6.9	6.5	15.3	7.0	6.7	965	63	
		60–79			23.7	13.5	12.9	5.1	5.7	5.0	11.9	5.8	5.5	1465	109	
	DML (ms)	19–49	Female			3.5	3.3	2.8	4.3	4.4	3.7	4.4	4.6	3.9	489	38
			Male			3.8	3.5	3.4	4.6	4.6	4.4	4.6	4.7	4.4	403	55
		50–79	Female			3.8	3.6	3.3	4.6	4.8	4.4	4.4	5.0	4.5	1066	66
			Male			4.0	3.9	3.7	4.8	5.1	4.7	4.7	5.2	4.7	1461	131
	Duration (ms)	19–79			5.9	5.4	5.8	7.7	7.0	7.2	8.0	7.1	7.1	3549	252	
	NCV (m/s)	19–39	Female			60.0	58.2	58.0	54.0	49.9	50.4	53.0	50.9	51.3	114	21
			Male			58.0	55.5	54.2	50.0	46.3	43.8	49.0	46.8	44.9	84	26
		40–79	Female			57.0	54.9	54.1	47.0	45.4	44.5	51.0	46.0	45.2	746	78
	Minimal F-wave latency (ms)	19–49		< 160			23.7	24.2		25.7	27.6		27.0		27	
				160–169			25.3	25.7		28.5	29.6		30.3		99	
				170–179			27.3	27.1		30.9	31.2		31.8		136	
				≥ 180			28.9	29.1		33.5	33.6		33.8		138	
		50–79	< 160			25.2	25.8		28.6	29.7		29.8		53		
			160–169			27.5	27.2		30.3	31.4		31.2		225		
170–179					28.7	29.3		31.5	34.1		34.3		337			
		≥ 180			30.4	31.2		34.2	36.0		35.7		426			
Ulnar	Amplitude (mV)	19–79			11.6	6.3	7.0	7.4	3.7	3.4	7.9	3.6	3.4	3844	237	
					35.9	19.1	22.4	21.7	10.9	13.0	23.9	10.7	13.4	3801	196	
	DML (ms)			3.0	2.9	2.6	3.6	3.8	3.3	3.7	3.8	3.3	4064	242		
				6.0	5.4	5.3	7.8	8.0	6.7	7.7	7.5	6.4	4284	201		
	NCV (m/s)			61.0	56.6	56.4	51.0	44.3	43.0	52.0	45.2	43.8	2188	241		
		Minimal F-wave latency (ms)	19–49	< 160			23.5	24.6		26.1	27.6		27.4		25	
	160–179					26.2	27.1	27.6	30.2	31.6	31.0	31.8	30.3	253	20	
	≥ 180					29.2	30.1	32.3	32.8	35.6	40.1	35.5	38.2	170	11	
	< 160					25.0	26.2		28.8	30.6		31.9		71		
	50–79	160–179			28.1	28.9	30.9	30.9	34.0	37.5	34.7	37.3	701	30		
≥ 180				30.4	32.0	32.8	33.8	37.3	37.3	37.2	37.0	503	18			

3.4. External validation

We calculated MMC-derived predicted values and their 95% prediction intervals for 21 measurements of an external, previously published set of 26 healthy subjects from the Amsterdam UMC (Verhamme et al., 2009). Of all measurements, 97.4% fell within the MMC derived 95% prediction interval.

4. Discussion

Here, we show that normal measurements can be extracted from existing NCSs using MMC. The derived cluster of normal measurements can be used to calculate a department-specific predicted value and z-score based on age, height, sex and temperature. This provides a more precise, individual estimate of the (ab)normality of a test result than the use of published RVs, because: 1) the estimation is based on very large data sets, 2) MMC-derived values incorporate all relevant patient characteristics and 3) values have been obtained using local equipment, settings and patient populations. In addition, this approach allows the user to attach a z-score to each measurement relative to the expected value of that patient, rather than merely stating that a measurement is within normal limits or not, as is currently common practice in most clinical neurophysiology departments.

4.1. External validation of MMC

MMC can be used in the absence of a formal gold standard, but we used several methods to show that the obtained values are indeed representative of normal measurements. First, MMC resulted in two clear clusters for each measurement, with the normal cluster following all expectations with regard to size, normal distribution and fit of the model. Second, MMC performed on two large independent datasets from different hospitals yielded values similar to each other and previously published RVs. Third, 97.4% of all measurements from an independent dataset of healthy controls fell within MMC-derived expected values, close to the expected value of 95%.

4.2. Comparison with published RVs

MMC-derived RVs were highly similar to previously published values (Dinesh Kumbhare et al., 2015). However, MMC-derived mean values for CMAP amplitudes and areas, while highly similar between LUMC and AUMC, were generally lower than those obtained by Buschbacher. This is probably because we calculated these values on the logarithmic scale, e.g., the mean tibial nerve CMAP amplitude would have been 7.2 mV instead of 6.3 mV when calculated on the natural scale. However, mean and SD should not be derived from skewed data (Fig. 1). This was previously noted by Buschbacher (Dinesh Kumbhare et al., 2015), although it appears that his published RVs were not based on log-transformed data.

Table 3

Buschbacher's RVs and MMC- derived RVs from the Amsterdam UMC and LUMC for leg nerves. Bb: Buschbacher, SD; standard deviation, DML: distal motor latency, NCV: nerve conduction velocity, AUMC: Amsterdam University Medical Center. LUMC: Leiden University Medical Center, N: number of measurements. Amplitude data expressed in mV, Area in mV.ms, DML, duration and F-wave latency in ms. Details with regards to settings and recording sites can be found in Appendix A. Mean ± 2SD and 3rd percentile columns: mean-2SD and 3rd percentile for CMAP amplitudes, CMAP areas and NCVs, mean + 2SD and the 97th percentile for CMAP duration, DML and minimal F-wave latency. Empty field: insufficient data for MMC.

Nerve	Units of measure	Age (years)	Height (cm)	Mean			Mean ± 2SD			3rd percentile			N		
				Bb	AUMC	LUMC	Bb	AUMC	LUMC	Bb	AUMC	LUMC	AUMC	LUMC	
Peroneal	Amplitude (mV)	19–39		6.8	4.2	4.9	1.8	1.9	2.3	2.6	2.0	2.4	360	92	
		40–79		5.1	2.9	3.5	0.1	1.1	1.5	1.1	1.2	1.7	2273	553	
	Area (mV.ms)	19–49		20.2	11.1	14.7	4.2	4.7	6.5	6.8	5.0	6.6	769	172	
		50–79		14.9	7.8	9.9	-0.3	2.9	4.3	3.6	3.2	4.8	1800	439	
	DML (ms)	19–79		4.8	4.9	3.8	6.4	6.6	5.1	6.4	6.6	5.0	3306	881	
		Duration (ms)		5.7	5.0	5.6	7.7	6.9	7.8	7.7	6.9	7.7	3504	841	
	NCV (m/s)	19–39	< 170	49.0	50.1	50.1	41.0		45.5	43.0	46.8		22		
			≥ 170	46.0		47.4	38.0		40.3	37.0	41.6		65		
		40–79	< 170	47.0		45.2	37.0		38.9	39.0	39.9		233		
			≥ 170	44.0		43.0	36.0		35.8	36.0	36.7		398		
		Minimal F-wave latency (ms)	19–39	< 160	45.7	41.1		51.5	47.2		45.5		12		
				160–169	48.8	45.1	45.8	56.6	54.6	50.8	53.7	49.5	29	12	
	Tibial	Amplitude (mV)	19–29		15.3	11.6	10.5	6.3	6.0	5.3	5.8	6.3	6.4	101	50
			30–59		12.9	8.3	7.4	3.9	3.3	2.9	5.3	3.3	3.0	1206	540
60–79				9.8	5.7	5.5	1.4	2.0	2.1	1.1	2.2	2.2	984	422	
Area (mV.ms)		19–49		38.9	21.3	18.2	10.3	9.0	7.6	14.2	8.6	8.1	672	266	
		50–79		29.2	13.8	13.0	2.6	4.6	5.0	5.0	4.9	5.1	1746	692	
DML (ms)		19–79		4.5	4.6	4.0	6.1	6.4	5.4	6.1	6.4	5.4	2883	1186	
	duration (ms)		6.3	5.4	5.5	8.7	9.8	8.1	8.7	8.2	7.8	3157	1211		
NCV (m/s)	19–49	< 160	51.0	45.7	45.4	43.0	37.4	39.0	44.0	40.1	40.9	14	10		
		160–169	49.0	46.6	46.1	37.0	38.7	38.2	42.0	41.0	39.2	71	74		
		≥ 170	47.0	43.3	44.8	37.0	34.5	36.2	37.0	34.8	37.4	280	204		
	50–79	< 160	49.0	43.3	44.2	39.0	35.9	35.6	40.0	36.6	37.9	61	68		
		160–169	45.0	42.8	43.5	35.0	35.1	35.8	37.0	35.6	37.1	259	233		
		≥ 170	44.0	40.3	40.9	34.0	31.5	32.5	34.0	31.9	33.3	794	600		
Minimal F-wave latency (ms)	19–39	< 160	43.2	41.1		47.6	46.1		46.1		15				
		160–169	47.2	46.7	47.7	53.2	54.3	53.3	55.2	52.1	27	24			
		170–179	52.0	50.7	52.9	60.0	59.3	60.5	58.7	60.2	52	41			
		≥ 180	53.1	55.4	56.7	61.9	65.0	65.0	65.3	62.8	56	23			
	40–79	< 160	46.8	46.5	50.2	56.0	53.8	58.6	53.5	58.3	54	51			
		160–169	50.5	50.4	53.1	57.5	59.0	60.7	60.2	60.0	248	177			
	170–179	53.9	54.6	56.5	61.1	63.8	65.2	63.6	64.9	310	230				
	≥ 180	57.9	60.5	60.7	68.5	71.4	70.0	70.7	68.6	345	167				

This is supported by the observation that the MN2SD and 3rd percentile values were highly similar in MMC-derived data, but not in Buschbacher's. In addition, differences may have been caused by differences in population, methods of data acquisition, such as filter settings, electrode size and electrode placement. Although equipment, electrode placement and filter settings are similar in our centers we observed small but significant differences between the datasets from both hospitals. We did not formally test whether all measurements differed significantly, but this is likely the case, underlining the importance of generating locally valid RVs (Tjon-A-Tsien et al., 1996; Van Dijk et al., 1999) We stratified data to create RV tables in accordance with groups used by Buschbacher to enable a comparison. However, we recommend using individual predicted values as this will yield more precise results than stratified data.

4.3. Comparison with other methods

Several other methods to derive RVs from existing data have been published. The Extrapolated Norms (E-Norms) procedure is based on the assumption that a cumulative distribution function (CDF) of values derived from biological measurements, will show an S-shaped curve, with a middle, linear segment representing the values of normal measurements (Jabre et al., 2015). RVs are derived by drawing a straight line through this middle segment and visually determining where it diverges from the S-curve. This

method has recently been refined using the extrapolated reference method (E-Ref), which calculates the inflection point by measuring the angle between adjacent points on the CDF plot (Nandedkar et al., 2018).

A drawback of both methods is that they require a relatively homogeneous data set as they do not allow correction for patient-related factors within the model (Stålberg et al., 2019). Multiple models are therefore needed based on different subsets of factors (e.g. age, height, sex), meaning that very large numbers of patients are required to obtain accurate RVs. Whereas MMC will result in a precise, tailored RV for each individual measurement, E-norms derived RVs are based on stratified group data, where large strata will lead to less precise results and smaller strata will require increasing large data sets. Furthermore, additional potentially relevant variables (e.g., height or body weight) can easily be added to MMC to increase precision. Finally, MMC enables the calculation of a z-score (in other words, a continuous value on the degree of abnormality), whereas E-norms will only yield a dichotomous outcome (i.e. normal or abnormal) that cannot be used for further statistical analysis.

4.4. Robustness of MMC

We show that clustering can be achieved with MMC for many measurements using only age and sex as independent variables. This is an advantage for common use, as these data will always

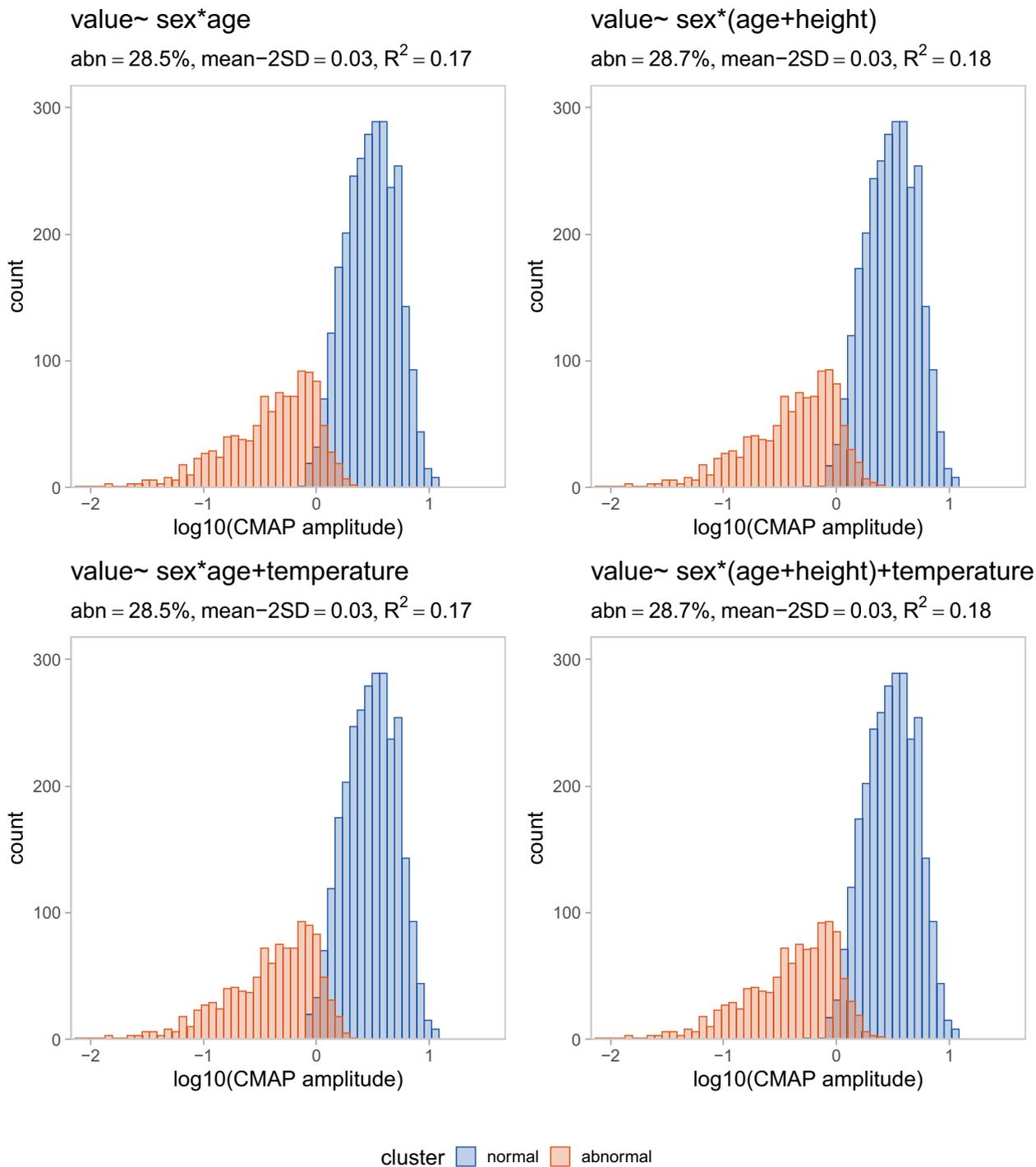


Fig. 3. Effect of different models on clustering of peroneal nerve CMAP amplitudes. Clustering is already fairly robust using sex and age. The “*” sign in the formulas means interaction: the most complicated model had main effects for sex, age, height and temperature and interaction effects for sex and age and sex and height. Calculations are based on those measurements for which all these variables were known (n = 3343); Amsterdam UMC data. CMAP: Compound Muscle Action Potential. abn: values in the abnormal cluster, percentage of total. R²: adjusted R². SD: standard deviation.

be available. Nonetheless, we recommend including temperature and height data when performing MMC if possible. These variables have significant effects in the obtained regression model, and models including these variables will therefore yield more precise RVs. Doubling the size of the abnormal cluster with simulated values based on existing peroneal nerve CMAP amplitude data had a limited effect on the final cut-off value of the normal value, but a further simulated increase of the abnormal cluster appeared to affect the SD and 3rd percentile cut-off of the nor-

mal cluster. Although this simulation is useful to show the limits of MMC, it should be noted that the abnormal cluster was far smaller for every actual measurement on which we performed MMC. An additional simulation reducing the amount of measurements showed that the reliability of MMC decreases noticeably when fewer than 300 measurements are used. These data are in line with the observation that a minimum of 300 measurements appears to be necessary to obtain relatively consistent results.

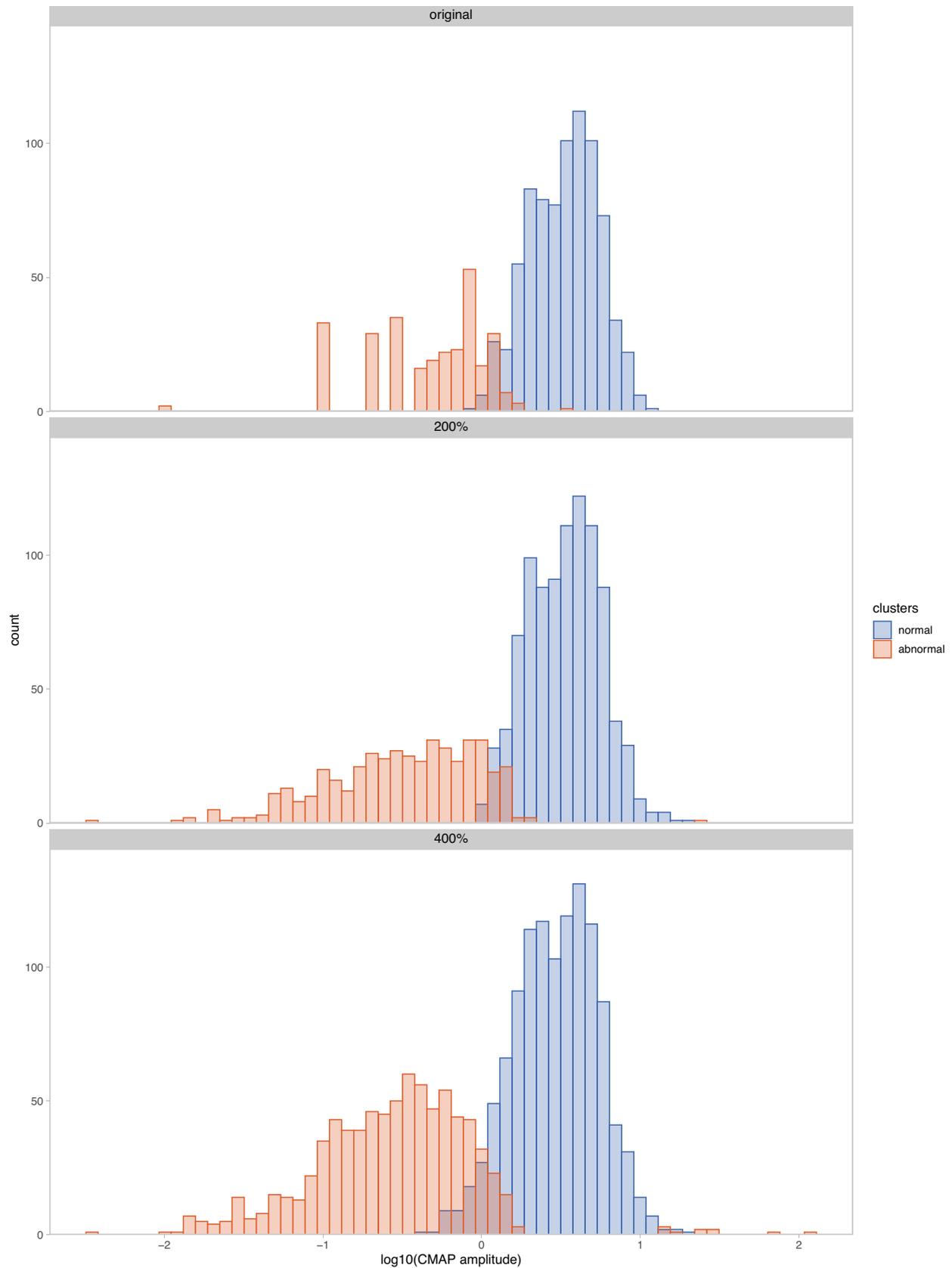


Fig. 4. Effect of abnormal cluster size. Doubling the number of abnormal measurements had limited effects on mean and MN2SD of the normal cluster: MN2SD was 1.2 mV using the actual data set, 1.2 mV after doubling the abnormal cluster. However, increasing the abnormal cluster fourfold (resulting in clusters of approximately equal size) led to a change in the MN2SD of the normal cluster to 0.9 mV. Calculations are based on LUMC data.

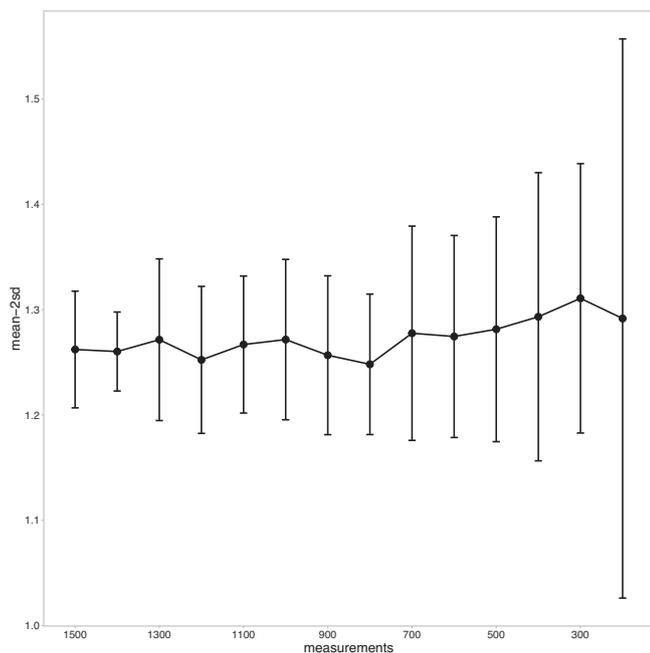


Fig. 5. Effect of sample size on precision of MMC predicted RVs. Simulation of 1500 normal and abnormal peroneal nerve Compound Muscle Action Potential (CMAP) amplitude measurements in a 4:1 ratio. The Y-axis shows the mean-2SD cut-off and standard deviation (SD) based on 30 simulations. Reducing the number of measurements in steps of 100 shows that the SD increases with fewer measurements, and becomes noticeably larger when the total number measurements falls below 300.

4.5. Limitations

Height and temperature data were frequently missing. In addition, precision of MMC can probably be improved by adding additional independent variables. Ethnicity (Shivji et al., 2019) and Body mass index (Salerno et al., 1998; Buschbacher, 1999a, 1999d) have been reported to affect NCS values. We found indications that the side of the measurement affected modelling for NCS measurements of the arms (data not shown). Indeed, NCS values may differ between the dominant and non-dominant arm (Werner and Franzblau, 1996; Kommalage and Gunawardena, 2013). As we did not record which limb was the dominant side, we decided not to use side as a factor in the current analysis.

MMC assumes that both clusters have a normal distribution, which was usually not the case for the abnormal cluster. We do not expect that this affected the properties of the normal clusters, as these always showed a normal distribution.

5. Conclusions

We provide a method that enables all clinical neurophysiology departments to generate locally valid RVs. To facilitate adoption of this method we have placed all R code and a test dataset online (gitlab.com/lumc/clinicalneurophysiology/ReferenceValues).

Future studies comparing data sets from more neurophysiology departments could shed light on the effects of local variations in NCS practice and allow the creation of RVs for NCS measurements for which reference data currently do not exist. MMC may also be useful to obtain optimal cut-off data for entrapment neuropathies, jitter values from single fiber EMG or NCS data from children (Pitt and Jabre, 2018). In addition, MMC is likely to be useful for any kind of neurophysiological data, including measurements such as nerve cross sectional areas obtained with ultrasound, quantitative EEG parameters and evoked potentials.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.clinph.2021.04.013>.

References

- Brown KE, Lohse KR, Mayer IMS, Strigaro G, Desikan M, Casula EP, et al. The reliability of commonly used electrophysiology measures. *Brain Stimul* 2017;10:1102–11. <https://doi.org/10.1016/j.brs.2017.07.011>.
- Buschbacher RM. Median nerve motor conduction to the abductor pollicis brevis. *Am J Phys Med Rehabil* 1999a;78:S1–8. <https://doi.org/10.1097/00002060-199911001-00002>.
- Buschbacher RM. Peroneal nerve F-wave latencies recorded from the extensor digitorum brevis. *Am J Phys Med Rehabil* 1999b;78:48–52. <https://doi.org/10.1097/00002060-199911001-00010>.
- Buschbacher RM. Tibial nerve F-waves recorded from the abductor hallucis. *Am J Phys Med Rehabil* 1999c;78:S43–7. <https://doi.org/10.1097/00002060-199911001-00009>.
- Buschbacher RM. Ulnar nerve motor conduction to the abductor digiti minimi. *Am J Phys Med Rehabil* 1999d;78:9–14. <https://doi.org/10.1097/00002060-199911001-00003>.
- Chen S, Andary M, Buschbacher R, Del Toro D, Smith B, So Y, et al. Electrodiagnostic reference values for upper and lower limb nerve conduction studies in adult populations. *Muscle Nerve* 2016;54:371–7. <https://doi.org/10.1002/mus.25203>.
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B* 1977;39:1–22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>.
- Van Dijk JG, Van Bente I, Kramer CGS, Stegeman DF. CMAP amplitude cartography of muscles innervated by the median, ulnar, peroneal, and tibial nerves. *Muscle Nerve* 1999;22:378–89. [https://doi.org/10.1002/\(SICI\)1097-4598\(199903\)22:3<378::AID-MUS11>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1097-4598(199903)22:3<378::AID-MUS11>3.0.CO;2-2).
- Dillingham T, Chen S, Andary M, Buschbacher R, Del Toro D, Smith B, et al. Establishing high-quality reference values for nerve conduction studies: A report from the normative data task force of the American Association of Neuromuscular & Electrodiagnostic Medicine. *Muscle Nerve* 2016;54:366–70. <https://doi.org/10.1002/mus.25204>.
- Dinesh Kumbhare MDMS, Robinson L, Buschbacher R. Buschbacher's Manual of Nerve Conduction Studies, Third Edition. 3rd ed. Springer Publishing Company; 2015.
- Goldfeld K. simstudy: Simulation of Study Data. 2019. <https://cran.r-project.org/package=simstudy>
- Grün B, Leisch F. FlexMix version 2: Finite mixtures with concomitant variables and varying and constant parameters. *J Stat Softw* 2008;28:1–35. <https://doi.org/10.18637/jss.v028.i04>.
- Jabre JF, Pitt MC, Deeb J, Chui KKH. E-Norms: A method to extrapolate reference values from a laboratory population. *J Clin Neurophysiol* 2015;32:265–70. <https://doi.org/10.1097/WNP.0b013e31827ed6aa>.
- Kommalage M, Gunawardena S. Influence of age, gender, and sidedness on ulnar nerve conduction. *J Clin Neurophysiol* 2013;30:98–101. <https://doi.org/10.1097/WNP.0b013e31827ed6aa>.
- Leisch F. FlexMix: A general framework for finite mixture models and latent class regression in R. *J Stat Softw* 2004;11:1–18. <https://doi.org/10.18637/jss.v011.i08>.
- Litchy WJ, Albers JW, Wolfe J, Bolton CF, Walsh N, Klein CJ, et al. Proficiency of nerve conduction using standard methods and reference values (cl. NPhys Trial 4). *Muscle Nerve* 2014;50:900–8. <https://doi.org/10.1002/mus.24243>.
- De Meyer G, Shapiro F, Vanderstichele H, Vanmechelen E, Engelborghs S, De Deyn PP, et al. Diagnosis-independent Alzheimer disease biomarker signature in cognitively normal elderly people. *Arch Neurol* 2010;67:949–56. <https://doi.org/10.1001/archneurol.2010.179>.
- Nandedkar SD, Sanders DB, Hobson-Webb LD, Billakota S, Barkhaus PE, Stålberg EV. The extrapolated reference values procedure: Theory, algorithm, and results in patients and control subjects. *Muscle Nerve* 2018;57:90–5. <https://doi.org/10.1002/mus.25606>.
- Pitt MC, Jabre J. The problem of lack of normative data in paediatric EMG and possible solutions. *Clin Neurophysiol* 2018;129:672–5. <https://doi.org/10.1016/j.clinph.2017.11.024>.

- R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2020. <https://www.R-project.org/>
- Salerno DF, Franzblau A, Werner R, Bromberg MB, Armstrong TJ, Albers JW. Median and ulnar nerve conduction studies among workers: Normative values. *Muscle Nerve* 1998;21:999–1005. [https://doi.org/10.1002/\(SICI\)1097-4598\(199808\)21:8<999::AID-MUS3>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1097-4598(199808)21:8<999::AID-MUS3>3.0.CO;2-0).
- Shivji Z, Jabeen A, Awan S, Khan S. Developing normative reference values for nerve conduction studies of commonly tested nerves among a sample Pakistani population. *J Neurosci Rural Pract* 2019;10:178–84. https://doi.org/10.4103/jnpr.jnpr_370_18.
- Stålberg E, van Dijk H, Falck B, Kimura J, Neuwirth C, Pitt M, et al. Standards for quantification of EMG and neurography. *Clin Neurophysiol* 2019;130:1688–729. <https://doi.org/10.1016/j.clinph.2019.05.008>.
- Tjon-A-Tsien AML, Lemkes HHPJ, van der Kamp-Huyts AJC, van Dijk JG. Large electrodes improve nerve conduction repeatability in controls as well as in patients with diabetic neuropathy. *Muscle Nerve* 1996;19:689–95. [https://doi.org/10.1002/\(SICI\)1097-4598\(199606\)19:6<689::AID-MUS1>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1097-4598(199606)19:6<689::AID-MUS1>3.0.CO;2-6).
- Toledo JB, Bjerke M, Da X, Landau SM, Foster NL, Jagust W, et al. Nonlinear association between cerebrospinal fluid and florbetapir F-18 β -amyloid measures across the spectrum of Alzheimer disease. *JAMA Neurol* 2015;72:571–81. <https://doi.org/10.1001/jamaneurol.2014.4829>.
- Verhamme C, Van Schaik IN, Koelman JHTM, De Haan RJ, De Visser M. The natural history of charcot-marie-tooth type 1A in adults: A 5-year follow-up study. *Brain* 2009;132:3252–62. <https://doi.org/10.1093/brain/awp251>.
- Werner RA, Franzblau A. Hand dominance effect on median and ulnar sensory evoked amplitude and latency in asymptomatic workers. *Arch Phys Med Rehabil* 1996;77:473–6. [https://doi.org/10.1016/S0003-9993\(96\)90036-2](https://doi.org/10.1016/S0003-9993(96)90036-2).