

Psychometric evaluation of The Index of Myalgic Encephalomyelitis Symptoms (TIMES). Part II: Criterion-related and discriminant validity, test-retest reliability and minimal detectable difference.

Sarah F Tyson¹, Mike C Horton², Russell Fleming³

1. School of Health Science, University of Manchester
2. Psychometric Laboratory for Health Sciences, University of Leeds
3. ME Association, UK

Corresponding Author: Prof Sarah Tyson (sarah.tyson@manchester.ac.uk)

Keywords: Myalgic encephalomyelitis, chronic fatigue syndrome, symptoms, reliability, minimal dateable difference, psychometrics, co-production, scale development, validity

Disclosure statement: None of the authors have any conflicts of interest but note that Mr Fleming is employed by the ME Association as Head of Project Development.

Funding statement: The work was funded by the ME Association, UK

Acknowledgments The authors thank the advisory groups' members for their unstinting support and wise counsel throughout the project.

Abstract

Objective: To evaluate the criterion-related and discriminant validity, test-retest reliability and minimal detectable difference of The Index of ME Symptoms (TIMES) in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS)

Methods: People with ME/CFS in the UK completed the TIMES online (n=1055). Rasch-transformed interval data and parametric statistics were used: Pearson correlations (with the ME severity scale); analysis of variance; intra-class correlations (ICC) and standard error of measurement of ICC measured criterion-related and discriminant validity, test-retest reliability and minimal detectable difference respectively.

Results: Highly significant ($P < 0.001$) moderate ($r = 0.400 - 0.528$) correlations were seen between the TIMES scales and severity of ME/CFS except the gastro-intestinal and immune systems scales ($r = 0.315$ and 0.302 $P < 0.001$ respectively). Discriminant validity was demonstrated with significant differences in TIMES scores between all five levels of ME severity, except between levels 4 and 5 in some cases, which were underpowered due to the small group numbers. Test-retest reliability was excellent ($ICC > 0.7$, $p < 0.001$) except the cranial nerves and immune system scales which were good ($ICC = 0.681$ and 0.669 , $p < 0.001$) and minimal detectable difference was excellent (3.95-17.45%).

Conclusions: The Index of ME Symptoms (TIMES) scales are valid, reliable, sensitive assessments of symptoms in ME/CFS. They are freely available for use.

Introduction

Myalgic encephalomyelitis (ME), also known as chronic fatigue syndrome (CFS) is a common disabling condition. Although aetiology is uncertain, recent genetic studies suggest that neurological and immune systems are involved¹. The core symptoms are debilitating fatigue/exhaustion (including post-exertional malaise), cognitive dysfunction, sleep disturbance and pain. However immune, autonomic, metabolic, and endocrine abnormalities are also widely reported^{2,3}. As there are no biomarkers for ME/CFS, diagnosis relies on fulfilment of symptom-based criteria. These are based on clinical academics' experience rather than empirical data^{4,5}.

There are few validated measures of symptoms in ME/CFS, most reports of symptoms are based on unvalidated lists used to investigate the diagnostic criteria^{4,5}. The exception is the DePaul Symptom Questionnaire (DSQ)⁶. This lengthy self-report measure of ME/CFS symptoms has 99 questions covering personal characteristics, history and symptoms. Fifty-four symptoms are included arranged to reflect the Clinical Canadian Criteria (CCC)⁷ for diagnosis. The original questionnaire has been modified over time such that original, expanded, brief, and paediatric versions are now available⁸. The psychometric properties of the original DSQ are reported to be strong (but the scores are not presented) in terms of test-retest reliability (based on kappa co-efficient and Spearman correlations)⁹ and criterion-related validity¹⁰, based on correlations with the Medical Outcomes Short Form 36¹¹. These were reported to be significant ($p < 0.05$) but the correlation co-efficients are not presented. Minimal detectable difference has not been reported, nor the construct validity in terms of item response theory.

Given the incomplete psychometrics and feasibility issues regarding the DSQ raised by our ME and clinicians' advisory groups (mainly the excessive length), we developed a new assessment of ME/CFS symptoms (The Index of ME Symptoms, TIMES). This was part of a project to co-produce a clinical assessment toolkit for ME/CFS with people with ME/CFS and clinicians working in specialist NHS ME/CFS services, which focussed on the information needed for clinical assessment rather than diagnosis.

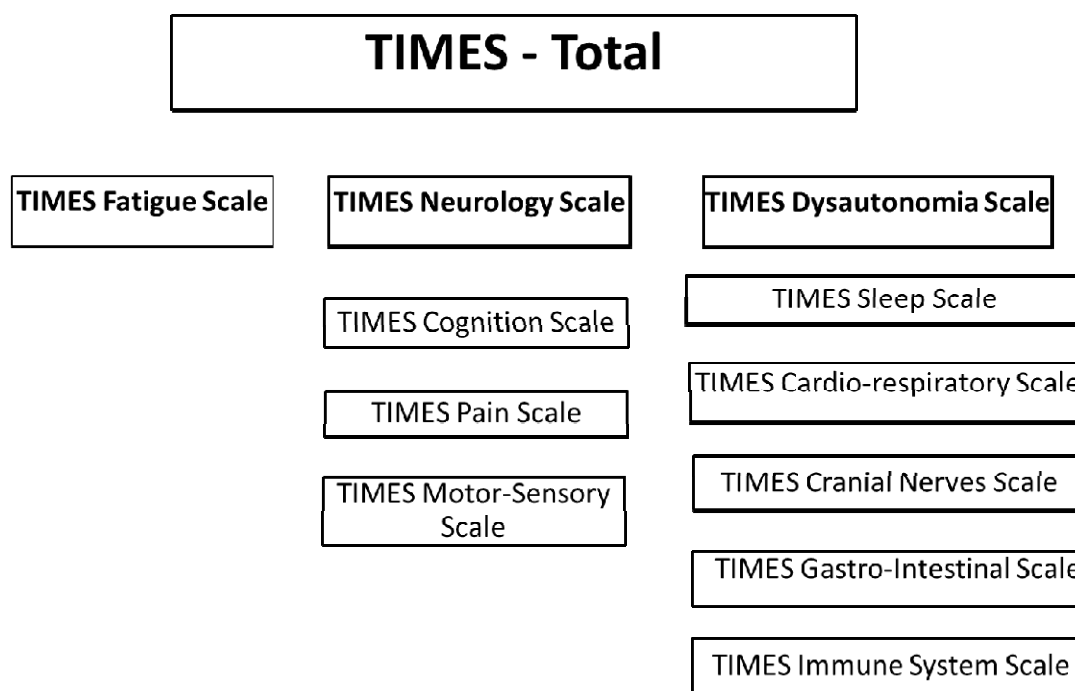
The work to develop and evaluate the TIMES using Rasch analysis has been detailed in the companion paper to this publication¹². Here we report the 'classic theory testing' in terms of

the criterion-related and discriminant validity, test-retest reliability and minimal detectable difference.

The TIMES assesses ME/CFS symptoms by asking questions about 58 symptoms on a four-point Likert scale (0 = I do not have this symptom to 3 = very severe). The summed total score (TIMES-Total) indicates the overall symptom burden. The symptoms are arranged in nine domains (fatigue, cognition, pain, sleep, motor-sensory symptoms, cardio-respiratory symptoms, gastro-intestinal symptoms, cranial nerves, and immune system symptoms). These provide robust, valid ‘stand-alone’ assessments of specific symptoms. The domains excluding fatigue also form valid sub-scales which assess neurological symptoms and dysautonomia. A schematic diagram summarising the portfolio of TIMES scales is found in Figure 1.

Figure 1 about here

Figure 1 Schematic summarising the TIMES portfolio of assessments



As well as providing summated scores summarising the level of symptom burden, which is useful for research audit, service development etc, the domains provide a detailed a profile of the nature and severity of users’ symptoms. This enables a more detailed examination of the patients’ illness which is needed in clinical practice.

The TIMES are presented in Supplementary File 1. It is freely available for others to use. Clinic ready versions are available to download from the ME Association website ([The ME Association Clinical Assessment Toolkit \(MEA-CAT\) - The ME Association](#)) and an online version (which includes automatic calculation of scores and production of the symptom profile is also available ([The ME Association App](#))).

Method

The TIMES was co-produced with people with ME/CFS and clinicians working in specialist NHS ME/CFS services using the methods detailed in the companion paper.¹² They are summarised here. An ME/CFS advisory group with a wide range of age, duration and severity of ME/CFS was convened from volunteers following publicity in the ME Association's newsletter. Members of the clinical advisory group were drawn from volunteers from the membership of the British Association of Clinicians in ME (BACME). Both groups contributed to all stages of the project.

Data were collected via an online survey of the TIMES using the Qualtrics Survey tool. People who had been diagnosed with ME/CFS in the UK were recruited via publicity in the UK's ME Association social media and publicity channels, ME support groups, and the 25% ME Group (a charity supporting people with severe ME).

Three cohorts were recruited during the development of the TIMES. Initially, a cohort were recruited to test the first iteration of the TIMES (TIMES1). Two weeks after submission, participants were asked to repeat the survey to assess test-retest reliability and minimal detectable difference. In addition to completing the TIMES, participants were asked whether, over the previous two weeks, their symptoms were '*overall the same, better or worse*'. The data from those who responded that their ME/CFS was 'the same' were used in the analysis of test-retest reliability and minimal detectable difference. Data from these two cohorts were transformed to reflect the final format of the TIMES¹². A further cohort were recruited to evaluate whether the revised version (TIMES2) fitted the Rasch model¹². Data from TIMES 1 and TIMES2 were combined and used for the analysis of validity as reported in this paper.

In line with best practice^{13,14}, raw (ordinal) scores were converted to a 0-100 interval level data through a logit-based transformation (also known as Rasch transformation). This accounts for the non-linear relationship between ordinal response patterns and the underlying

latent trait. The advantage of this procedure is that it increases measurement precision and enables parametric statistical analyses to be used. The table detailing that conversion is shown in Supplementary File 2 and is freely available for use.

A classic theory approach was used to assess criterion-related (also known as convergent validity) and discriminant validity, test-retest reliability and minimal detectable difference following the COSMIN Design Checklist for Patient-Reported Outcome Measurement Instruments to structure the analysis and reporting¹⁵.

Criterion-related and discriminant validity was examined by comparison with the ME/CFS Severity Scale¹⁶. The ME/CFS Severity scale is a 5-level ordinal scale. A higher score indicates greater severity. Pearson correlations assessed criterion-related validity, testing the hypothesis that symptom burden would increase with ME/CFS severity. Analysis of variance assessed divergent validity by testing whether there were significant differences in symptom burden between different levels of ME/CFS severity.

Intra-class correlations (two-way random effects model) assessed test-retest reliability, while minimal detectable difference (MDD) was calculated from the standard error of measurement of the intra-class correlation (ICC) where:

$$\text{Minimal detectable difference} = \text{standard deviation} \times (\sqrt{1-ICC})$$

Results

One thousand and fifty-five participants were recruited. The three cohorts of participants had very similar characteristics (Table 1). Reflecting the demographics of ME/CFS, they were predominantly middle-aged women with moderately severe ME/CFS and extensive lived experience.

Table 1 about here

Table 1. Characteristics of the recruited cohorts

	TIMES 1	TIMES 2	Retest cohort whose ME/CFS was ‘the same’
N	701	354	100
Gender (m/f/other)	116 (16%) 572 (82%) 13 (2%)	58 (16.4%) 295 (83%) 1 (0.3%)	19 (19%) 78 (78%) 3(3%)
Age	53.7 (13.5)	56.3 (13.6)	53.0 (14.1)
Duration	14.7 (12.2)	16.2 (12.2)	15.4 (12.8)
Severity	Mean (sd) 2.85 (0.89) L1: 41 (5.8%) L2: 194 (27.7%) L3: 310 (44.2%) L4: 138 (19.7%) L5: 18 (2.6%)	2.9 (0.95) L1: 23 (6.5%) L2: 100 (28.2%) L3: 132(37%) L4: 89(25%) L5: 10 (2.8%)	2.81 (0.93) L1: 6(6%) L2: 32 (32%) L3: 41 (41%) L4: 17(17%) L5: 4 (4%)
Location			
England	606 (86%)	309 (87%)	85 (85%)
Wales	25 (3.6%)	1 (3.4%)	1 (1%)
Scotland	62 (8.8)	30 (8.5%)	14 (14%)
Northern Ireland	8 (1.1%)	0	0

L1= Level 1, very mild ME/CFS; L2 = Level 2, mild ME/CFS; L3= level 3, moderately severe ME/CFS; L4= level 4 severe ME/CFS; L5= level 5, very severe ME/CFS.

Criterion-related and discriminant validity

Highly significant ($P < 0.001$) moderate ($r = 0.400 - 0.528$) correlations were seen between the TIMES scales and severity of ME/CFS (Table 2), except the gastro-intestinal and immune systems scales, which had correlation co-efficients of 0.315 and 0.302 respectively. All the TIMES scales demonstrated excellent discriminant validity, with significant differences in scores between all five levels of ME severity, except between levels 4 and 5 in some cases (Table 2). The comparisons between levels 4 and 5 were probably underpowered due to the small number of people in these groups.

Table 2 about here

Table 2: Criterion and discriminate validity for the TIMES scales.

	Criterion-related validity (Pearson's correlation, r)	Discriminant validity (ANOVA, p value)
Total TIMES score	0.504	All comparisons significant at P<0.001
Fatigue scale	0.528	All comparisons significant at P<0.001 except level 4 and 5 (p=1)
Neurology sub-scale	0.471**	All comparisons significant at P<0.001
Cognition scale	0.414**	All comparisons significant at P<0.001
Pain scale	0.405**	All comparisons significant at P<0.05
Motor-sensory scale	0.431**	All comparisons significant at P<0.001
Dysautonomia sub-scale	0.450**	All comparisons significant at P<0.01
Sleep scale	0.400	All comparisons significant at P<0.01
Cranial nerves scale	0.402**	All comparisons significant at P<0.001 except level 4 and 5 (P=0.062)
Gastro-intestinal scale	0.315**	All comparisons significant at P<0.001 except level 4 and 5 (p=1)
Cardio-respiratory scale	.450**	All comparisons significant at P<0.001
Immune system scale	.302**	All comparisons significant at P<0.05

Test-retest reliability and sensitivity to change

The TIMES scales demonstrated excellent test-retest reliability (ICC>0.7, p<0.001) except the cranial nerves and immune system scales which were good (ICC = 0.681 and 0.669 respectively, p<0.001) (Table 3). The minimal detectable difference was excellent (3.95-17.45%, Table 3).

Table 3 about here

Table 3 Test-retest reliability and minimal detectable difference of the TIMES scales

Scale	Test-retest reliability (ICC)	Minimal detectable difference (0-100 scale)
Total Times score	0.795	3.95
Fatigue scale	0.732	15.31
Neurology scale	0.773	7.13
Cognition Scale	0.700	17.45
Pain Scale	0.733	14.85
Motor-sensory Scale	0.775	12.42
Dysautonomia scale	0.781	6.27
Sleep Scale	0.768	13.02
Cardio-respiratory Scale	0.751	13.18
Cranial nerves Scale	0.681	12.08
Gastro-intestinal Scale	0.715	12.53
Immune Scale	0.669	15.71

Discussion

The results of this study demonstrate that the TIMES scales are valid, reliable, sensitive measures of ME/CFS symptomology. They are designed primarily as a tool to characterise ME/CFS symptoms for use in clinical practice, but they can also be used as an outcome measure, or to monitor changes over time.

Although highly significant, criterion-related validity correlation values were only ‘moderate’. A possible explanation is the high degree of inter-individual variability in the nature and severity of individuals’ symptoms, which is widely reported anecdotally in the ME community. A further issue is the fluctuating nature of individuals’ symptoms, which is also widely reported, although empirical data is lacking. Such fluctuations would cause participants to report that their overall condition is ‘the same’, although their symptoms could have changed considerably within the two-week assessment period. Such fluctuations would result in poorer reliability and minimal detectable difference sensitivity scores due to changes in the participants’ condition, rather than a short-coming of the measurement tool. A two-week period was used between the test and retest data collection to avoid recall bias, as is the convention when evaluating measurement tools for chronic conditions. However, there has

been very little research into how ME/CFS symptoms changes over the short or long term, so it was unclear whether the participants' symptoms could be assumed to be stable over this period. A shorter recall period (e.g. a week) could reduce the impact of fluctuations and give a clearer indication of the scale's properties but increases the likelihood of recall bias. Further research is needed to explore how ME/CFS symptoms change over both the short- and longer-term, and to investigate how to interpret the results of test-retest reliability evaluations in fluctuating conditions.

The use of Rasch-transformed interval level data is novel and successful. We did not formally compare the precision of the raw (ordinal) and transformed (interval) data, but substantial improvements in measurement error (45%) have been reported¹⁷. This has several benefits. By increasing the statistical power, it enables more accurate detection of changes over time and possible treatment effects, while reducing the sample sizes needed to detect a change. It also allows information about people with ME's symptoms to be included in parametric analyses with other continuous data (e.g. physiological or biomedical markers) which increases the breadth and depth of such investigations.

The only other assessment of ME symptoms which has undergone psychometric evaluation is the DePaul Symptom Questionnaire (DSQ)^{6,8-10}. Direct comparison with the results presented here are limited as most work to evaluate the DSQ focusses on its primary purpose; to operationalise diagnostic criteria. This work used factor analysis^{10,18,19} in relation to diagnostic criteria, and sensitivity and specificity to discriminate between people with ME/CFS and controls. However, Brown and Jason (2014)¹⁰ reported “*good convergent and discriminant validity*” based on significant ($p < 0.05$) comparisons with the Medical Outcomes Short Form 36¹¹. The correlation co-efficients and specific p-values are not presented, so direct comparison are not possible, but the correlations presented in this paper are all highly significant, suggesting the criterion-related validity at least matches the DSQ.

Test-retest reliability of the DSQ is described as having “*good to excellent test-retest reliability, with Pearson's or kappa correlation coefficients that were 0.70 or higher for the majority of items*”⁹. However, the scores for individual items, or the summated scores are not presented. For TIMES, we used intra-class correlations (as recommended in the COSMIN guidelines¹⁵) which showed good-excellent test-retest reliability, so it appears the reliability

of the two assessments is similar. The minimal detectable difference, or other measures of sensitivity to change have not been reported for the DePaul Symptom Questionnaire.

The main strengths of this study are use of co-production to ensure the content comprehensively covers the issues that are important relevant to people with ME/CFS in a feasible and acceptable format, its large sample size, use of Rasch transformed interval data and rigorous application of the COSMIN guidelines¹⁵. Possible limitations relate to the representativeness of the sample. There have been few community-based epidemiology studies of ME/CFS, so the demographics of the disease are unclear. However, the characteristics of our participants broadly reflect other large adult ME/CFS studies with convenience samples¹⁹⁻²¹. Thus, we are confident the sample is representative of the ME/CFS community. Nevertheless, our recruitment methods would be biased towards English speakers who are active on social media and involved with ME organisations. In keeping with other large-scale surveys of ME/CFS, we recruited people with a “formal diagnosis” of ME/CFS, but we did not confirm the diagnosis via health records. This is because the recording of ME/CFS in both primary and secondary care is notoriously inaccurate, incomplete and inconsistent^{22,23}. Consequently we have may recruited a few people who were misdiagnosed. Further research is needed to better understand the demographics of ME/CFS, and how to accurately diagnose and record it.

Conclusions

The TIMES scales are valid, reliable and sensitive assessments of symptomology in ME/CFS.

References

1. Genetics Delivery Team, Boutin T, Bretherick AD, Dibble JJ, Ewaoluwabemiga E, Northwood E, Samms GL, Vitart V, Project and Cohort Delivery Team, Almelid Ø, Baker T. Initial findings from the DecodeME genome-wide association study of myalgic encephalomyelitis/chronic fatigue syndrome. medRxiv. 2025 Aug 8:2025-08.
2. Missailidis D, Annesley SJ, Fisher PR. Pathological Mechanisms Underlying Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Diagnostics*. 2019;9:80.
3. Shepherd C (2022). ME/CFS/PVFS The ME Association Clinical and Research Guide (13th Ed). The ME Association, UK.

4. Lim EJ, Son CG. Review of case definitions for myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Journal of translational medicine*. 2020 Jul 29;18(1):289.
5. Brurberg KG, Fønhus MS, Larun L, Flottorp S, Malterud K. Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. *BMJ open*. 2014 Feb 1;4(2):e003973.
6. Jason LA, Evans M, Porter N, Brown M, Brown AA, Hunnell J, et al. The development of a revised Canadian myalgic encephalomyelitis chronic fatigue syndrome case definition. *A J Biochem Biotechnol*. (2010) 6:120–35.
7. Carruthers BM, Jain AK, De Meirleir KL, et al. Myalgic encephalomyelitis/chronic fatigue syndrome: clinical working case definition, diagnostic and treatment protocols. *J Chronic Fatigue Syndr*. 2003;11(1):7–115.
8. Jason LA, Sunnquist M. The development of the DePaul Symptom Questionnaire: Original, expanded, brief, and pediatric versions. *Frontiers in pediatrics*. 2018 Nov 6;6:330.
9. Jason LA, So S, Brown AA, Sunnquist M, Evans M. Test-retest reliability of the DePaul Symptom Questionnaire. *Fatigue* (2015) 3:16–32. doi: 10.1080/21641846.2014.978110
10. Brown AA, Jason LA. Validating a measure of myalgic encephalomyelitis/chronic fatigue syndrome symptomatology. *Fatigue: Biomedicine, Health & Behavior*. 2014 Jul 3;2(3):132-52.
11. Ware JE Jr & Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, **30** (6), 473–483.
12. MC Horton, SF Tyson , R Fleming, Gladwell P (2026). Development and psychometric evaluation of The Index of Myalgic Encephalomyelitis Symptoms (TIMES) Part I: Rasch Analysis and Content Validity. MedRxiv Preprint doi: <https://doi.org/10.64898/2026.02.16.26346394> February 17, 2026
13. Tennant A and Küçükdeveci AA (2023) Application of the Rasch measurement model in rehabilitation research and practice: early developments, current practice, and future challenges. *Front. Rehabil. Sci.* 4:1208670. doi: 10.3389/fresc.2023.1208670
14. Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurol* 2007; 6: 1094–105
15. Mokkink LB, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, de Vet HCW et al. The COSMIN study design checklist for patient-reported outcome measurement instruments [COSMIN checklist with 4-point scale](#) Last access Jan 2026.

16. National Institute of Health and Clinical Excellence. Chronic fatigue syndrome/myalgic encephalomyelitis (or encephalopathy) Guidelines (CG53, 2007). [NICE-Guideline-MECFS-CG53-2007-Downloaded-01.08.20.pdf \(meassociation.org.uk\)](https://www.nice.org.uk/guidance/CG53/2007) Last accessed January 2026.
17. Bartholomew EJ, Medvedev ON, Petrie KJ, Chalder T. Rasch analysis of the hospital anxiety and depression scale in patients with chronic fatigue syndrome. *Journal of Psychosomatic Research*. 2025 Sep 3:112370.
18. Jason LA, Sunnquist M, Brown A, Furst J, Cid M, Farietta J, et al. Factor analysis of the DePaul Symptom Questionnaire: Identifying core domains. *J Neurol Neurobiol*. (2015) 1. doi doi: 10.16966/2379 7150.114
19. Huber K, Sunnquist M, Jason LA. Latent class analysis of a heterogeneous international sample of patients with myalgic encephalomyelitis/chronic fatigue syndrome. *Fatigue* (2018) 6:163–78. doi: 10.1080/21641846.2018.1494530
20. Tyson SF, Stanley K, Gronlund TA, Leary S, Dean SE, Dransfield C et al Research priorities for myalgic encephalomyelitis/ chronic fatigue syndrome (ME/CFS): the results of a James Lind alliance priority setting exercise. *Fatigue: Biomedicine, Health & Behavior*, 2022: 10:4, 200-211
21. Bretherick AD, McGrath SJ, Devereux-Cooke A, Leary S, Northwood E, Redshaw A, Stacey P, Tripp C, Wilson J, Chowdhury S, Lewis I. Typing myalgic encephalomyelitis by infection at onset: A DecodeME study. *NIHR open research*. 2023 Aug 21;3:20.
22. Solve ME Initiative (2022) Summary of ME/CFS ICD-10-CM code changes: https://bit.ly/MECFS_ICDCCode. Last accessed January 2026
23. Valdez AR, Hancock EE, Adebayo S, Kiernicki DJ, Proskauer D, Attewell JR et al. Estimating Prevalence, Demographics, and Costs of ME/CFS Using Large Scale Medical Claims Data and Machine Learning. *Front Pediatr*. 2019 Jan 8;6:412.