临床研究 DOI：10.13406/j.cnki.cyxb.003703

# 中文版肌萎缩侧索硬化多域功能障碍量表信效度分析

万石磊[1]，杨　丹[2]，刘念荃[2]，蒋千平[1]，徐丹丹[2]，赵　焰[2]，周　晶[2]
（1.湖北中医药大学针灸骨伤学院，武汉　430065；
2.湖北省中医院/湖北中医药大学附属医院推拿科/康复医学科，武汉　430061）

**【摘　要】目的：**汉化肌萎缩侧索硬化多域功能障碍量表（amyotrophic lateral sclerosis impairment multidomain scale，AIMS），并验证其在肌萎缩侧索硬化（amyotrophic lateral sclerosis，ALS）患者中的信度和效度。**方法：**纳入2023年8月至2024年7月在湖北省中医院运动神经元病中西医结合诊疗中心注册登记系统中的161例ALS患者作为研究对象，采用经翻译、修订后的中文版AIMS进行调查。使用Rasch模型对问卷进行单维性检验以及内部一致性、内容效度、结构效度、效标关联效度评价，依据经典测量理论进行项目分析、重测信度分析、效标关联效度分析。**结果：**单维度检验显示AIMS的3个维度均具备良好的单维性；除条目18以外，加权的均方拟合统计量（information-weighted fit statistic mean square，Infit MNSQ）和未加权的均方拟合统计量（outlier-sensitive fit statistic mean square，Outfit MNSQ）均在0.5～1.5，条目总分相关性（point-measure correlation，PT-Measure）在0.63～0.87；各条目相邻两选项的阈值>1.4 logit；怀特图显示条目难度在-2.24~2.08 logit；项目功能差异分析显示不同特征亚组患者差异值的绝对值均<1 logit；总量表和各维度的信度均>0.9，个体信度均>0.7，除呼吸维度分离指数为1.56外，总量表和其余维度分离指数均>2；项目分析显示各维度的条目与其所在维度的相关系数在0.622～0.865（$P$<0.01）；总量表与延髓、运动、呼吸3个维度的重测信度分别为0.994、0.970、0.990、0.972（$P$<0.01）；效标关联效度分析结果显示中文版AIMS与修订版肌萎缩侧索硬化功能评分量表（amyotrophic lateral sclerosis impairment multidomain scale，ALSFRS-R）总分和相应维度分数的相关系数分别为0.863、0.829、0.876、0.755（$P$<0.01）。**结论：**中文版AIMS具有良好的拟合度和较高的信效度，其条目难度适中，选项设置合理，整体质量良好，适用于我国ALS患者的功能评估。

**【关键词】**肌萎缩侧索硬化；Rasch模型；经典测量理论；信度；效度
**【中图分类号】**R744.8　　**【文献标志码】**A

# Reliability and validity of the Chinese version of Amyotrophic Lateral Sclerosis Impairment Multidomain Scale

*Wan Shilei*[1]，*Yang Dan*[2]，*Liu Nianquan*[2]，*Jiang Qianping*[1]，*Xu Dandan*[2]，*Zhao Yan*[2]，*Zhou Jing*[2]
（*1. College of Acupuncture and Orthopedics，Hubei University of Chinese Medicine；*
*2. Department of Tuina / Rehabilitation Medicine，Hubei Provincial Hospital of Traditional Chinese Medicine / Affiliated Hospital of Hubei University of Chinese Medicine*）

**【Abstract】*Objective*：**To translate Amyotrophic Lateral Sclerosis Impairment Multidomain Scale（AIMS）into Chinese，and to investigate its reliability and validity among Amyotrophic Lateral Sclerosis（ALS）patients. ***Methods*：**A total of 161 ALS patients who were registered in the registry system of Integrated Traditional Chinese and Western Medicine Treatment Center for Motor Neuron Disease in Hubei Provincial Hospital of Traditional Chinese Medicine from August 2023 to July 2024 were enrolled as subjects，and the Chinese version of AIMS was used for investigation after translation and revision. The Rasch model was used to assess the unidimensionality，internal consistency，content validity，construct validity，and criterion-related validity of the questionnaire，and the classical test theory was used for item analysis，test-retest reliability analysis，and criterion-related validity analysis. ***Results*：**Unidimensionality testing showed that all three dimensions of AIMS have good unidimensionality. For all items except item 18，information-weighted fit statistic mean square（Infit MNSQ）and outlier-sensitive fit statistic mean square（Outfit MNSQ）ranged from 0.5 to 1.5，and point-measure correlation（PT-Measure）ranged from 0.63 to 0.87. The threshold between adjacent options of each item was >1.4 logit. Wright maps showed that item difficulties ranged from -2.24 to 2.08 logit，and the differential item functioning analysis showed that the absolute values of differences between patients across different characteristic subgroups were <1 logit. The overall scale and each dimension had a reliability of >0.9，with an individual reliability of >0.7. The dimension of respiration had a separation index of 1.56，while the overall scale and the other dimensions had a separation in-

**作者介绍：**万石磊，Email：wanshilei2000@163.com，
研究方向：肌萎缩侧索硬化中西医结合治疗。
**通信作者：**周　晶，Email：zhoujing@hbhtcm.com。
**基金项目：**2023年湖北省自然科学基金资助计划（联合基金项目）（编号：2023AFD128）；2022年湖北省重点研发计划资助项目（编号：2022BCA055）。
**优先出版：**https://link.cnki.net/urlid/50.1046.R.20250113.0909.002（2025-01-13）

dex of >2. The item analysis showed that the correlation coefficient between the items in each dimension and their respective dimensions ranged from 0.622 to 0.865（$P$<0.01）. The overall scale and the dimensions of medulla oblongata, motor, and respiration had a test-retest reliability of 0.994, 0.970, 0.990, and 0.972, respectively（$P$<0.01）. The criterion-related validity analysis showed that the correlation coefficients of total score and the score of each dimension between the Chinese version of AIMS and the revised version of AIMS were 0.863, 0.829, 0.876, and 0.755, respectively（$P$<0.01）. ***Conclusion***: The Chinese version of AIMS has a good degree of fitting and relatively high levels of reliability and validity, with moderate item difficulty, reasonable option settings, and high overall quality, and it is suitable for the functional assessment of ALS patients in China.

**【Key words】**amyotrophic lateral sclerosis; Rasch model; classical test theory; reliability; validity

肌萎缩侧索硬化（amyotrophic lateral sclerosis，ALS）是1种选择性影响大脑皮质、皮质脊髓束、脑干和脊髓等上下运动神经元的神经退行性疾病[1]。ALS患者通常表现为骨骼肌进行性无力、萎缩、痉挛，进而导致运动、言语、吞咽、呼吸等多种功能障碍[2]。该病中位生存期仅为3～5年，多数ALS患者因呼吸功能衰竭或营养不良等原因而死亡，生活质量及预后极差[3]。因此，精确、全面和量化地评估ALS患者的功能障碍非常重要，这不仅帮助医生及时采取多学科管理进行干预，也在药物和其他临床试验中确定终点具有重要意义。

目前，在ALS临床治疗场景及临床试验中，基于经典测量理论（classical test theory，CTT）开发的修订版肌萎缩侧索硬化功能评分量表（amyotrophic lateral sclerosis impairment multidomain scale，ALSFRS-R）应用最为广泛[4]。然而，近年来的研究指出ALSFRS-R在应用中存在各维度在总分中的权重不均衡[5]，分数与病情严重程度未呈线性关系[6]，敏感性不足[7]，评分会随着某些干预措施而改变等问题。为了提高评估的准确性，荷兰乌得勒支大学的Leonard van den Berg教授团队通过研究ALS国际指南、成立专家小组及ALS患者访谈，采用CTT和Rasch模型编制了肌萎缩侧索硬化多域功能障碍量表（amyotrophic lateral sclerosis impairment multidomain scale，AIMS），用于评估ALS患者功能障碍[8]。该研究发现此量表与ALS患者生存时间密切相关，具有良好的信度和效度。与ALSFRS-R相比，AIMS的天花板效应和地板效应明显降低，提高了临床检测ALS疾病进展的准确性。本研究拟对AIMS进行汉化，并基于Rasch模型和CTT评估其在我国ALS患者中的信度和效度，旨在为ALS患者功能障碍的量化评估和临床试验中的终点设定提供更多工具和参考。

## 1 资料与方法

### 1.1 研究对象

于2023年8月至2024年7月在湖北省中医院运动神经元病中西医结合诊疗中心注册登记系统中的ALS患者作为研究对象。纳入标准：①符合黄金海岸标准[9-10]的ALS患者；②年龄18~80岁，性别不限；③认知功能正常；④自愿参与并签署知情同意书。排除标准：①处于疾病终末期（有创机械通气或因并发症引起意识改变的患者）；②严重精神疾患；③排除肌萎缩侧索硬化合并额颞叶痴呆患者。

### 1.2 研究工具

1.2.1 一般资料调查表 自行设计一般资料调查表收集患者基线资料，主要内容包括性别、年龄、文化程度、发病时间、发病部位、是否服用利鲁唑、肌电图显示累及节段等。

1.2.2 AIMS 该量表包括球部（7条）、运动（11条）、呼吸（5条）3个维度，涵盖言语不利、吞咽困难、饮水呛咳、肢体功能异常、呼吸困难等常见症状，共计23个条目，均为标准选择题形式，每个选择题提供3个备选项，依照自身病情选择对应选项，分别赋予0～2分，0分代表程度较重，1分代表程度较轻，2分代表没有异常，总分为46分，分数越高，表明患者症状越轻。

1.2.3 ALSFRS-R 该量表包括12个评估项目，ALSFRS-R可划分为3个维度[11]，其各维度可以单独评价，具体包括延髓功能（3条）：言语、流涎、吞咽；运动功能（6条）：书写（ALS起病前的优势手）、使用餐具（未行胃造瘘）/使用餐具（行胃造瘘）、穿衣和洗漱、床上翻身和调整被褥、行走、爬楼梯；呼吸功能（3条）：呼吸困难、端坐呼吸、呼吸机使用。每个条目提供5个选项，依次计为0~4分，延髓12分、运动24分、呼吸12分，总分为48分，得分越高，表明神经功能受损程度越小。ALSFRS-R目前已被翻译成多个语言版本，具有良好的评分者间信度、评分者内信度和电话随访信度[12-13]。本研究使用该量表作为校标测量中文版AIMS的效标关联效度。

### 1.3 研究方法

1.3.1 AIMS量表汉化 经电子邮件获取Leonard van den Berg团队汉化授权，按照Brislin翻译模型[14]对英文版量表进行翻译和修订。①正译：由2名康复科医生分别独立将原量表翻译为中文后，经2人协商讨论，比较整合各自翻译版本，

形成AIMS中文版1。②回译：由1名双语（中、英）神经病学专家和1名湖北中医药大学英语教研室教师在未接触原量表内容的前提下，分别回译AIMS中文版1为英语。经研究组成员将2份回译版与原量表对比分析，达成统一意见后形成AIMS中文版2。③文化调试：邀请6名具有10年及以上工作经验的神经内科、康复医学科高级职称专家对AIMS中文版2的翻译作出评价以及修改意见，综合各专家意见直至没有异议，最终形成AIMS中文版3。④预调查：采用方便抽样的方法选取符合纳入排除标准的30例ALS患者，发放AIMS中文版3，向患者解释调查目的和注意事项，填写完成后当场收回。询问患者是否能够充分理解每个条目的含义，每个选项是否清楚、有无歧义，以及存在的其他问题和建议。汇总、整理、修改后展开第二次专家咨询，形成最终AIMS中文版本。

1.3.2 资料收集 取得患者同意并签署知情同意书后，使用统一的指导语向患者介绍研究目的、主要内容，告知患者填写注意事项，同时向患者发放最终AIMS中文版问卷，对于无法握笔书写的患者，由研究者通过问答方式代其填写。Rasch模型分析对样本量并无特别要求，仅有相关研究对其做指导性规定，相关研究建议如果要使参数估计达到稳定，纳入样本量至少为100[15]，但如果要获得±0.5 logit区间的稳定性，Rasch分析需要至少150个样本[16]。由于ALS为罕见病，故本次研究计划至少纳入150例患者。

### 1.4 统计学方法

通过Excel录入患者数据，计数资料用频数、百分比表示，计量资料符合正态分布时用均数±标准差（$\bar{x} \pm s$）表示，偏态分布时使用中位数（四分位间距）[$M_d$（$P_{25}$，$P_{75}$）]表示。使用SPSS 29.0进行项目分析、重测信度分析、效标关联效度分析。使用Winsteps 3.72.3对AIMS相关数据进行Rasch模型分析，包括单维性检验、条目拟合情况、阈值有序性、条目与个体匹配度、项目功能差异分析、区分度检验及项目特征曲线等。单维性是指量表的所有条目仅测量单个潜在特质，是Rasch模型分析的前提条件[15]。本研究使用Rasch模型进行残差主成分分析以检验量表的单维性，Rasch模型首成分能够解释的比例越高越好，Linacre JM[17]建议Rasch分析中首成分解释的原始方差≥50%，才能满足单维性的要求。Racsh模型采用加权的均方拟合统计量（information-weighted fit statistic mean square，Infit MNSQ）和未加权的均方拟合统计量（outlier-sensitive fit statistic mean square，Outfit MNSQ）拟合统计量验证观测数据是否符合预测模型。Infit MNSQ主要反映项目难度与个人能力水平相当的数据，而Outfit MNSQ反映极端值或异常数据[18]。每个条目的Infit MNSQ和Outfit MNSQ的数值应该在0.5~1.5，数值越接近1，拟合效果越好，但Outfit MNSQ容易受到极端值的影响[19]。条目总分相关性（point-measure correlation，PT-Measure）用于表示个体在单个条目上与所在维度的相关性，一般认为0.4~0.8表示相关性良好，相关系数越高，该条目越能反映测量特质[20]。检验水准$\alpha$=0.05。

## 2 结 果

### 2.1 患者一般资料

本次研究共纳入161例受试者，其中男96例，女64例；年龄26~69岁，平均年龄为（50.83±9.25）岁；平均体质指数（body mass index，BMI）为（23.34±3.45）kg/m$^2$；教育程度：小学及以下19例（11.8%），初中34例（21.12%），高中/中专38例（23.6%），大专28例（17.39%），本科及以上42例（26.09%）；肢体起病136例（84.47%），延髓起病25例（15.53%）；服用利鲁唑136例（84.47%）；目前肌电图表现累及节段：1个节段25例（15.53%），2个节段39例（24.22%），3个节段47例（29.19%），4个节段50例（31.06%）；病程中位数为19（12，30）个月；发病年龄中位数为48（43，56）岁；ALSFRS-R评分中位数为36（28，42）。

### 2.2 AIMS量表汉化结果

为了贴合我国民众的文化习惯和语言表达，本研究在原有量表条目的编制及格式的基础上，对部分条目及其补充说明进行适当修改，但没有增减条目。具体修改如下：将条目1补充说明中的“语音计算机”修改为“语音计算机或屏幕朗读”；条目2补充说明中的“牛排”修改为“牛肉”；条目3中“管饲”修改为“胃管或鼻饲管进食”，以便于患者理解；条目9中“楼梯升降机”在我国不太常见，因此修改为“楼梯升降机或电梯”；将条目12“I have trouble opening or closing the zip and/or buttons on my trousers”翻译为“很难将裤子上的拉链拉上/拉开和（或）纽扣解开/系上”，以符合我国语言表述习惯；条目22条“很难将痰从肺或喉咙中咳出”因患者难以理解，将其修改为“咳痰没有力气”。

### 2.3 Rasch分析结果

2.3.1 单维性检验 将各维度分别进行单维性检验，结果显示各维度中解释的原始方差均≥50%（延髓维度59.0%，运动维度56.0%，呼吸维度58.4%），量表有良好的单维性。

2.3.2 条目拟合统计 本研究结果显示，3个维度的Infit MNSQ和Outfit MNSQ值除了Q18以外的条目均在0.5~1.5，表明拟合效果较为理想，PT-Measure在0.63~0.87显示条目与所在维度存在较为良好的相关性（表1）。

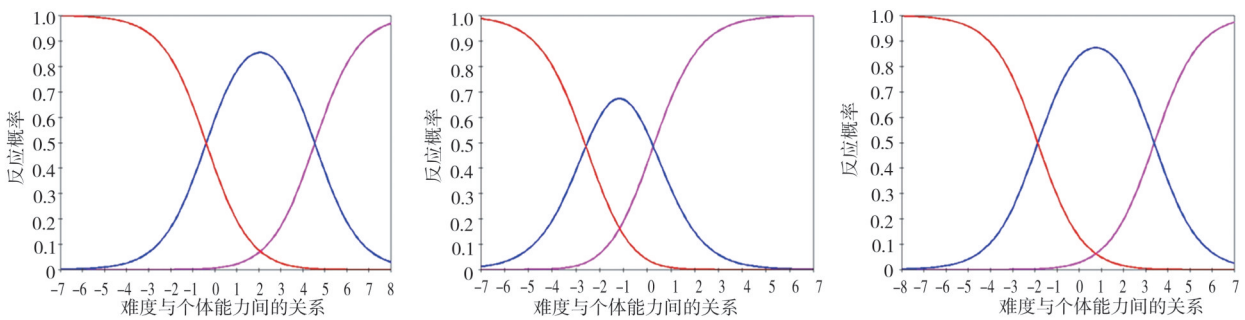
2.3.3 阈值有序性及项目特征曲线 阈值的有序性反映了各条目选项的阈值是否按照适当的顺序排列，条目的阈值参数由2个相邻分类概率曲线的交点决定[21]。理论上，等级的设置应随选项难度的提升而单调递增，每个条目的阈值应从小到大排序。原量表条目提供3个选项，对应0~2分，为3点量表，故存在2个阈值。一般而言，量表相邻选项的难度差值应在1.4~5.0 logit，表明各等级有差异，且合理有序[22]。在延髓、运动、呼吸3个维度中，相邻2个选项的阈值差异均大于1.4 logit（表2），表明各个维度的条目选项难度设置较为合理。延髓、运动、呼吸维度的项目特征曲线显示23个项目中没有一个显示无序阈值，见图1A~C。

表1　中文版AIMS的项目拟合分析

| 条目 | 难度值 | Infit | | Outfit | | PT-MEASURE |
|---|---|---|---|---|---|---|
| | | MNSQ | ZSTD | MNSQ | ZSTD | |
| 延髓功能 | | | | | | |
| Q1 比以前更少说话，因为现在变得更加困难了 | 2.08 | 1.25 | 1.60 | 1.16 | 0.80 | 0.85 |
| Q2 咀嚼和吞咽食物需要更长时间 | 0.87 | 0.82 | −1.50 | 0.79 | −1.40 | 0.87 |
| Q3 发现饮水困难 | −0.35 | 0.82 | −1.50 | 0.79 | −1.40 | 0.83 |
| Q4 喉咙里有很多痰并需要经常咳嗽 | −1.00 | 1.07 | 0.60 | 1.02 | 0.20 | 0.77 |
| Q5 需要擦嘴巴避免口水流下来 | −0.11 | 0.88 | −1.00 | 0.83 | −1.10 | 0.83 |
| Q6 醒来时，枕头上有口水 | −1.06 | 0.86 | −1.10 | 0.73 | −1.30 | 0.83 |
| Q7 食欲不如以前了 | −0.44 | 1.33 | 2.50 | 1.33 | 1.90 | 0.75 |
| 运动功能 | | | | | | |
| Q8 发现在家里短距离行走很困难 | −1.16 | 1.40 | 3.10 | 1.49 | 2.70 | 0.68 |
| Q9 爬楼梯困难 | 0.75 | 1.47 | 3.70 | 1.31 | 2.00 | 0.63 |
| Q10 在不平坦的地面上行走困难 | 0.05 | 1.27 | 2.40 | 1.21 | 1.60 | 0.68 |
| Q11 梳头或刷牙困难 | 0.35 | 0.72 | −2.90 | 0.67 | −2.90 | 0.82 |
| Q12 很难将裤子上的拉链拉上/拉开和（或）纽扣解开/系上 | 0.63 | 0.69 | −3.10 | 0.63 | −3.00 | 0.81 |
| Q13 写字困难，字迹不清晰 | 0.53 | 0.90 | −0.90 | 0.80 | −1.60 | 0.78 |
| Q14 用钥匙打开门锁困难 | 0.13 | 0.79 | −2.00 | 0.73 | −2.30 | 0.82 |
| Q15 自己进食困难 | 0.15 | 0.72 | −2.80 | 0.81 | −1.60 | 0.79 |
| Q16 收拾餐具困难 | 1.03 | 0.74 | −2.60 | 0.64 | −2.60 | 0.78 |
| Q17 在床上翻身困难 | −0.75 | 0.85 | −1.40 | 0.96 | −0.30 | 0.75 |
| Q18 抬头困难 | −1.70 | 1.54 | 3.80 | 2.14 | 4.40 | 0.65 |
| 呼吸功能 | | | | | | |
| Q19 当走路、穿衣、说话或吃东西时，会感到呼吸急促 | 0.76 | 1.19 | 1.60 | 1.23 | 1.60 | 0.77 |
| Q20 静坐时都呼吸急促 | −2.24 | 0.78 | −1.20 | 0.51 | −1.30 | 0.78 |
| Q21 当平躺时，呼吸变得更加困难或感到呼吸急促 | 0.71 | 1.03 | 0.30 | 1.00 | 0.10 | 0.84 |
| Q22 咳痰没有力气 | 1.41 | 0.92 | −0.70 | 0.92 | −0.50 | 0.87 |
| Q23 使用呼吸机 | −0.64 | 0.93 | −0.50 | 0.87 | −0.60 | 0.78 |

表2　各维度相邻选项间的阈值

| 维度 | 0 ~ 1 | 1 ~ 2 |
|---|---|---|
| 延髓功能 | −2.46 | 2.46 |
| 运动功能 | −1.42 | 1.42 |
| 呼吸功能 | −2.63 | 2.63 |

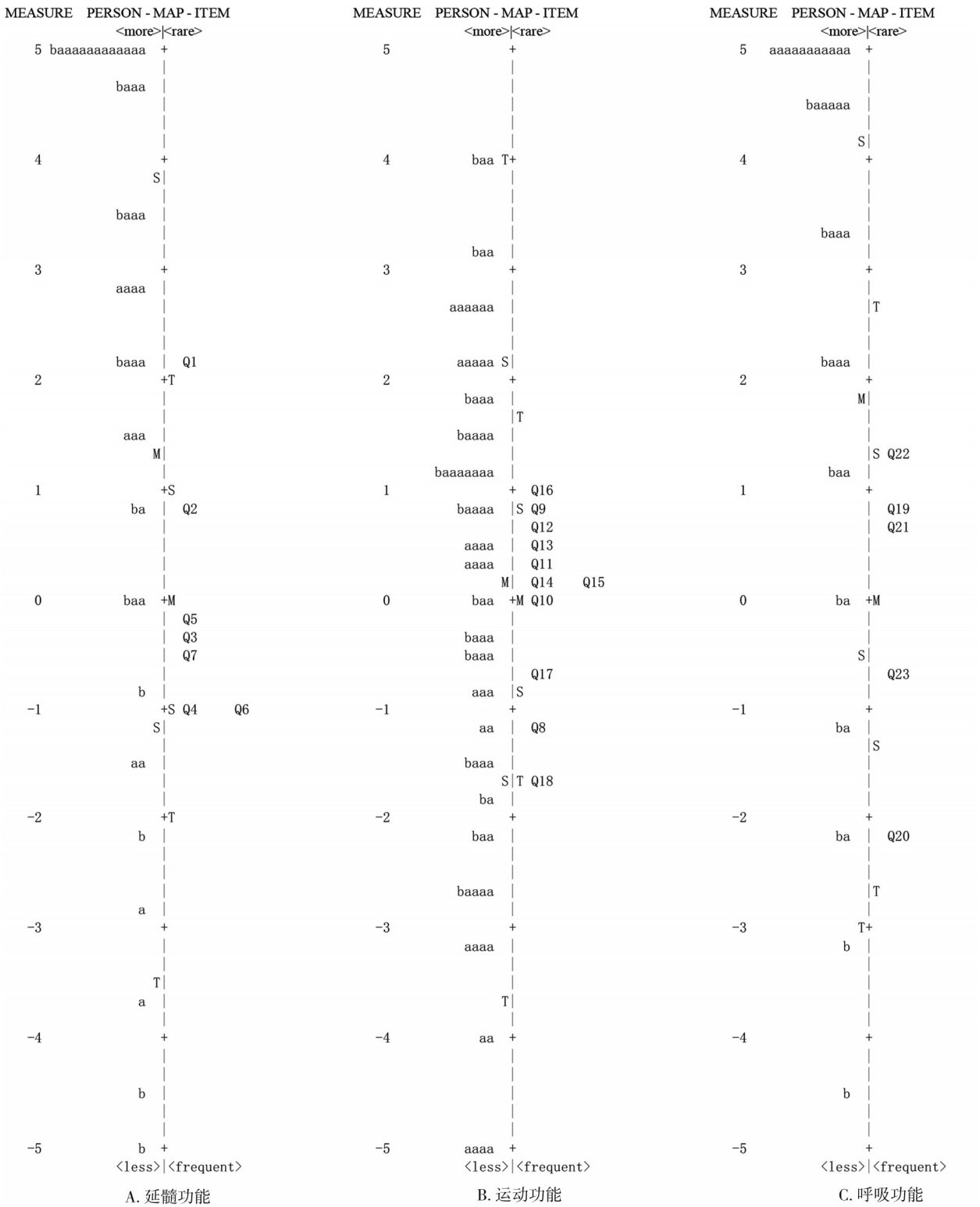



注：-类别概率，0；-类别概率，1；-类别概率，2；X轴为条目难度与个体能力间的关系，Y轴为反应概率；对于功能障碍程度最轻的患者，最可能选择2分的选项，随着整体功能障碍程度的增加，最可能选择1分的选项，随着整体功能障碍程度增加到最高水平，最可能的选项逐渐变为0分的选项；23个条目阈值均有序，延髓、运动、呼吸3个维度各举1个条目示例如Q1（A）、Q8（B）、Q19（C）

图1　条目特征曲线

2.3.4　怀特图　怀特图（Wright Map）是Rasch分析中体现量表整体质量的指标之一，它将个体能力与条目难度放在同个量尺中，根据各自在量尺中的位置，直接显示二者的分布情况，以及条目设置的难度是否能满足受试者的需要[23]。左侧为被试者的能力分布，右侧为条目的难度值分布，在理想状态下，两侧的均值应接近0且差值<1 logit[24]。分别对3个维

度进行分析，结果显示个体能力与条目难度的匹配不佳（图2A~C），个人能力分布跨度大于条目难度分布，仅有运动功能维度两侧均值接近0且差值未超过1 logit。各维度均显示部分个体的能力过低，在其能力值范围内，未出现难度相当的条目。AIMS延髓、运动、呼吸3个维度的条目难度范围分别在−1.06~2.08 logit、−1.7~1.03 logit、−2.24~1.41 logit（表1）。




注：虚线为量尺；左边代表被试能力分布状况，从−5到5代表能力值递高；“a”与“b”表示例数，在图2A延髓功能中“a”代表4位患者，“b”代表1~3位患者；在图2B呼吸功能中“a”代表2患者，“b”代表1位患者；在图2C运动功能中“a”代表5位患者，“b”代表1~4位患者；右边代表条目难度分布情况，从−5到5代表难度依次增大；被试的能力水平与条目难度越接近，越能反映被试者的能力；M表示平均值，左侧的M表示个体能力的平均值，右侧的M表示条目难度的平均值；S表示1个标准差，T表示2个标准差

**图2 中文版AIMS量表各维度怀特图**

此外还将入组患者的ALSFRS-R评分进行Rasch分析，结果显示ALSFRS-R的延髓、运动、呼吸的3个维度的难度范围分别为-0.96~1.56 logit，-0.82~0.84 logit，-0.82~0.79 logit。具体如下，延髓功能中言语（1.56 logit）、流涎（-0.96 logit）、吞咽（-0.60 logit）；运动功能中书写（-0.15 logit）、使用餐具（0.02 logit）、穿衣和洗漱（0.44 logit）、床上翻身和调整被褥（-0.57 logit）、行走（-0.58 logit）、爬楼梯（0.84 logit）。

2.3.5 项目功能差异分析 项目功能差异分析（differetial item functioning，DIF）通常是用来检测相同的条目在评估不同受试者时，是否会因具有某一特征而导致差异，如性别、年龄、人种、教育程度等。一般认为2组之间DIF的绝对值大于0.5 logit，T的绝对值>2，则表明该条目存在偏倚，但DIF的绝对值<1 logit时也是可接受的[25]。本研究按照性别、是否低于平均年龄进行亚组分析，“运动功能”维度的Q9在性别间、“延髓功能”维度的Q4以及“运动功能”维度的Q18在年龄间不满足上述条件外，其余条目均在可接受范围（表3）。

**表3 不同项目功能差异分析**

| 条目 | 性别 | | 年龄 | |
|---|---|---|---|---|
| | DIF差值 | T | DIF差值 | T |
| 延髓功能 | | | | |
| Q1 | 0.50 | 1.10 | 0.36 | 0.78 |
| Q2 | -0.26 | -0.59 | -0.24 | -0.55 |
| Q3 | 0.39 | 0.88 | -0.20 | -0.05 |
| Q4 | 0.31 | 0.67 | -1.08[a] | -2.28 |
| Q5 | 0.35 | 0.81 | 0.47 | 1.07 |
| Q6 | -0.71 | -1.48 | 0.76 | 1.63 |
| Q7 | -0.65 | -1.42 | -0.24 | -0.54 |
| 运动功能 | | | | |
| Q8 | -0.08 | -0.22 | 0.19 | 0.55 |
| Q9 | 0.66[a] | 2.05 | -0.11 | -0.33 |
| Q10 | 0.33 | 1.04 | 0.06 | 0.20 |
| Q11 | -0.08 | -0.26 | 0.18 | 0.58 |
| Q12 | -0.57 | -1.76 | 0.04 | 0.14 |
| Q13 | -0.08 | -0.25 | 0.34 | 1.08 |
| Q14 | 0.00 | 0.00 | 0.00 | 0.00 |
| Q15 | -0.03 | -0.09 | 0.17 | 0.54 |
| Q16 | -0.34 | -1.06 | -0.16 | -0.48 |
| Q17 | 0.13 | 0.38 | -0.16 | -0.49 |
| Q18 | 0.10 | 0.36 | -0.74[a] | -2.05 |
| 呼吸功能 | | | | |
| Q19 | 0.55 | 1.26 | -0.26 | -0.58 |
| Q20 | -0.54 | -0.86 | -0.26 | -0.44 |
| Q21 | 0.06 | 0.13 | 0.00 | 0.00 |
| Q22 | -0.02 | -0.60 | 0.33 | 0.76 |
| Q23 | -0.37 | -0.75 | 0.03 | 0.06 |

注：a，DIF的绝对值>1 logit，或（且）T的绝对值>2

2.3.6 区分度检验 在Rasch模型中，个体分离指数（person separation index，PSI）与项目分离指数（item separation index，ISI）分别评价量表对个体的区分度和量表的结构效度，通常认为分离指数≥2，信度≥0.7较为合适[26]。总量表与各维度的样本与条目分离系数、信度见表4，所有维度和总量表的ISI在4.33~6.55，PSI在1.56~3.20。

**表4 中文版AIMS量表分离指数及信度**

| 项目 | 个体 | | 条目 | |
|---|---|---|---|---|
| | 分离指数 | 信度 | 分离指数 | 信度 |
| 总量表 | 3.20 | 0.91 | 6.55 | 0.98 |
| 延髓功能 | 2.02 | 0.80 | 4.33 | 0.95 |
| 运动功能 | 2.65 | 0.88 | 4.54 | 0.95 |
| 呼吸功能 | 1.56 | 0.71 | 5.26 | 0.97 |

2.4 经典测量理论

2.4.1 项目分析 对于问卷中的条目与各维度总分，进行了Pearson相关性分析。当单一条目与总分关联较强时，表明条目间具有较高程度的一致性；在相关系数分析中，那些未达到显著性标准或相关系数低于0.4的条目，根据具体情况考虑是否删去[27]，分析结果显示各维度的条目与其所在维度的相关系数在0.622~0.865（$P$<0.01）（表5）。

**表5 中文版AIMS各维度项目分析**

| 条目 | $r$值 | $P$值 |
|---|---|---|
| 延髓功能 | | |
| Q1 | 0.830 | <0.01 |
| Q2 | 0.865 | <0.01 |
| Q3 | 0.848 | <0.01 |
| Q4 | 0.786 | <0.01 |
| Q5 | 0.842 | <0.01 |
| Q6 | 0.860 | <0.01 |
| Q7 | 0.745 | <0.01 |
| 运动功能 | | |
| Q8 | 0.658 | <0.01 |
| Q9 | 0.622 | <0.01 |
| Q10 | 0.676 | <0.01 |
| Q11 | 0.846 | <0.01 |
| Q12 | 0.845 | <0.01 |
| Q13 | 0.815 | <0.01 |
| Q14 | 0.851 | <0.01 |
| Q15 | 0.808 | <0.01 |
| Q16 | 0.805 | <0.01 |
| Q17 | 0.737 | <0.01 |
| Q18 | 0.646 | <0.01 |
| 呼吸功能 | | |
| Q19 | 0.761 | <0.01 |
| Q20 | 0.791 | <0.01 |
| Q21 | 0.839 | <0.01 |
| Q22 | 0.865 | <0.01 |
| Q23 | 0.774 | <0.01 |

2.4.2 重测信度 采用抽样便利法，选取30例研究对象在完成第1次评估后2周进行重测，2次得分的内在相关性通过Pearson相关系数来计算，以便进行问卷稳定性验证。对2次结果进行相关性分析，延髓、运动、呼吸的重测信度分别为0.970、0.990、0.972（$P$<0.01），总量表的重测信度为0.994（$P$<0.01），均大于0.8，表示量表有良好的稳定性和可靠性。

2.4.3 效标关联效度分析 效标关联效度指测量量表与外在效标量表间的相关程度，使用Pearsen相关分析方法，外在效标选择目前常用的ALSFRS-R量表。中文版AIMS总分与ALSFRS-R总分的相关系数为0.863，2个量表相应维度的总分的相关系数分别为0.829、0.876、0.755，均呈现正相关（表6）。

表6 AIMS与ALSFRS-R评分的相关性

| 项目 | 延髓功能 | 运动功能 | 呼吸功能 | 总分 |
|---|---|---|---|---|
| $r$值 | 0.829 | 0.876 | 0.755 | 0.863 |
| $P$值 | <0.01 | <0.01 | <0.01 | <0.01 |

## 3 讨 论

### 3.1 AIMS具有良好的信度及区分度

在重测信度分析中，各维度及总量表的重测信度系数介于0.970～0.994，均>0.8，表明量表具有良好的稳定性和可靠性。在Rasch模型分析中，各维度的条目信度为0.95~0.98，样本信度则为0.71~0.91，进一步证明了量表的信度和可靠性。分离指数用于反映区分不同能力和难度的效果，但PSI容易受到所选样本的影响。所有维度和总量表的ISI均>2，表明条目难度具有良好的区分度。总量表（PSI=3.2）能够将患者的功能区分为4个层次；延髓（PSI=2.02）和运动（PSI=2.65）子量表可以将患者分为3个层次。而呼吸子量表的PSI为1.56，显示个体能力之间的差异较小，仅能分出2个层次。这可能与所选样本中患者的病情相对相似有关，此次研究对象中无气切及使用气管插管的患者，大部分患者的呼吸功能损害较轻，ALSFRS-R的呼吸功能评分主要集中在11~12分。使用Rasch模型对患者数据进行了项目功能差异分析，其中按性别和是否低于平均年龄进行分组。结果显示，除条目4、条目9和条目18外，AIMS的大多数条目在不同特征的患者中均显示出可接受的差异性。一般而言，量表具有10%~20%有差异性的题目仍然是可以接受的[28]。因此，本课题组决定保留这3个条目，以维持量表的完整性和覆盖面。

### 3.2 AIMS具有良好的效度

AIMS的结构效度通过Rasch模型分析的残差主成分分析、条目拟合统计等来验证。使用残差主成分分析，各维度所解释的原始方差均≥50%，各维度均呈现单维性，表明各维度没有特质交叉。条目拟合统计结果显示，除条目18条目（抬头困难）拟合效果不佳外，其余22个条目的Infit MNSQ和Outfit MNSQ值在0.51~1.49，说明数据与Rasch模型拟合度较好，量表能够较准确地预测ALS患者病情严重程度。PT-Measure反映的是条目与分量表之间的相关性，相关系数在0.63~0.87；基于CTT的项目分析显示，各维度的条目与其所在维度的相关系数在0.622~0.865，2种测量方法均表明显条目与量表有良好的相关性。当条目在Rasch模型中拟合不佳时，表明与量表中的其他条目不协调，应考虑将其删除[29]，但并不表明该条目不重要，可能是在本次调查中未能有效测出。因此本课题组进行了项目分析，并通过Pearson相关性分析发现条目18的相关系数为0.646（>0.4）。考虑到对ALS患者来说，抬头困难是一个重要的预后因素[30]，且在本研究的样本中，抬头困难严重、轻度困难和没有困难的患者分别占14.20%、29.63%和56.17%，本课题组决定保留该条目。在量表难度分析中（图1），题目难度分布大约5个logit，被试者能力分布大约为10个logit，说明在各维度题目难度为中等难度，对于功能较强或较弱的患者缺乏与之对应的条目，这可能是与ALS呈现出区域受累以及进展程度的特点相关[31]。AIMS与ALSFRS-R相关性结果表明，AIMS具有较好的校标效度（>0.4）。本研究还对选项进行了阈值分析，结果显示选项设置合理，没有出现阈值混乱。

### 3.3 AIMS应用价值

在当前的ALS临床研究中，由于缺乏客观的功能障碍评估和预后判断指标，主要的临床结局评价依赖于各种量表，如ALSFRS-R。因此，量表作为测量工具的敏感性及其是否能准确反映患者病情的严重程度，对于确保临床研究得出正确的结论至关重要。目前被广泛使用的ALSFRS-R是基于CTT开发并验证的，但这种计分方式的刻度与病情之间是非线性关系[32]。例如，在运动功能的子量表中，

ALS患者爬楼梯能力从无需帮助到需要帮助时，评分下降3分，此时3分的下降仅反映轻度功能下降；同样，如果患者从行走正常变为仅有非行走功能，也会损失3分，这3分的下降则代表严重的功能丧失。这表明，在ALSFRS-R量表上，变化相同的分数并不能准确代表具体的功能变化幅度，每个分数的变化在不同的情境中可能代表不同的严重性。相比之下，Rasch模型可将非线性的数据转化成等距的“logit scale”，使客观测量成为可能[17]。基于Rasch模型开发的AIMS在各自维度上的选项阈值是有序排列的（表2），因此在任何一个维度上每1分的变化代表着一致且等价的病情变化。随着ALS病情的发展将累及不同区域（延髓，脊髓颈、胸、腰骶段），会不同程度地影响言语、吞咽、呼吸、各肢体活动等功能[31]。因此总评分往往难以准确反映出某一局部功能的变化。在临床试验中，使用总评分来评估治疗效果可能会导致结果被“稀释”，如苯丁酸钠和牛磺酸治疗ALS的临床试验中，药物的治疗效果主要体现在ALSFRS-R的运动维度的评分增加，而总评分改善较小[33]。使用多维度的AIMS比仅提供总分的量表更为适用此种情况。同时与ALSFRS-R（12条）相比，AIMS（23条）各维度条目均有所增加，在“延髓功能”维度中AIMS增添了痰液、口水、食欲问题，在“运动功能”维度中将行走功能分为“家里”“不平坦路面”两个情景，另外增加了使用钥匙、收拾餐具、抬头等问题，在“呼吸功能”维度中增加了咳痰问题，评估更加细致。另外对本课题组对ALSFRS-R各维度的进行了Rasch分析，而AIMS难度范围更大（表1），这表明AIMS评估的ALS残疾程度范围比ALSFRSR更广，可以更好地评估ALS疾病进展和治疗效果，有望在未来临床研究中成为评估ALS患者更敏感的结果指标。另外与ALSFRS-R各维度相比，AIMS各维度的信噪比更高，减少了在临床试验观察中结果的随机波动或不确定性，增加了评估时的准确度，减少了检测给定治疗效果所需的样本量，6个月和12个月临床试验的估计样本量分别减少了16.3%和25.9%[8]。

综上所述，本研究中的中文版AIMS涵盖3个维度，共23个条目。通过CTT和Rasch模型的分析表明，AIMS具有良好的拟合度，信度和效度较高，条目难度适中，选项设计合理，总体质量优良，适合用于评估我国ALS患者的功能状态。由于ALS是一种罕见疾病，其在我国的发病率仅为1.62/10万人·年[34]，这使得大样本的获取具有挑战性，因此，本次研究的临床对象均来自单中心，且未对终末期患者进行评估。未来，本课题组计划在多中心进行大样本的重复验证，并探讨AIMS在网络和电话随访中的可信度。此外，还将通过纵向研究，研究AIMS评分及其下降率与患者生存率之间的关系。

## 参 考 文 献


[1] Brown RH, Al-Chalabi A. Amyotrophic Lateral Sclerosis[J]. N Engl J Med, 2017, 377(2): 162-172.

[2] Masrori P, Van Damme P. Amyotrophic lateral sclerosis: a clinical review[J]. Eur J Neurol, 2020, 27(10): 1918-1929.

[3] Grassano M, Moglia C, Palumbo F, et al. Sex differences in amyotrophic lateral sclerosis survival and progression: a multidimensional analysis[J]. Ann Neurol, 2024, 96(1): 159-169.

[4] Bakker LA, Schröder CD, Tan HHG, et al. Development and assessment of the inter-rater and intra-rater reproducibility of a self-administration version of the ALSFRS-R[J]. J Neurol Neurosurg Psychiatry, 2020, 91(1): 75-81.

[5] Franchignoni F, Mandrioli J, Giordano A, et al. A further Rasch study confirms that ALSFRS-R does not conform to fundamental measurement requirements[J]. Amyotroph Lateral Scler Frontotemporal Degener, 2015, 16(5/6): 331-337.

[6] van Eijk RPA, de Jongh AD, Nikolakopoulos S, et al. An old friend who has overstayed their welcome: the ALSFRS-R total score as primary endpoint for ALS clinical trials[J]. Amyotroph Lateral Scler Frontotemporal Degener, 2021, 22(3/4): 300-307.

[7] Bedlack RS, Vaughan T, Wicks P, et al. How common are ALS plateaus and reversals?[J]. Neurology, 2016, 86(9): 808-812.

[8] de Jongh AD, van Eijk RPA, Bakker LA, et al. Development of a rasch-built amyotrophic lateral sclerosis impairment multidomain scale to measure disease progression in ALS[J]. Neurology, 2023, 101(6): 602-612.

[9] Shefner JM, Al-Chalabi A, Baker MR, et al. A proposal for new diagnostic criteria for ALS[J]. Clin Neurophysiol, 2020, 131(8): 1975-1978.

[10] 王虹杨，卢祖能. 肌萎缩侧索硬化症：2020诊断标准（黄金海岸标准）[J]. 神经损伤与功能重建, 2021, 16(12): 719-721, 763.
Wang HY, Lu ZN. Amyotrophic Lateral Sclerosis: 2020 Diagnostic Criteria (Gold Coast Criteria)[J]. Neural Injury and Functional Reconstruction, 2021, 16(12): 719-721, 763.

[11] Franchignoni F, Mora G, Giordano A, et al. Evidence of multidi-

mensionality in the ALSFRS-R scale: a critical appraisal on its measurement properties using rasch analysis[J]. J Neurol Neurosurg Psychiatry,2013,84(12):1340-1345.

[12] Bakker LA, Schröder CD, van Es MA, et al. Assessment of the factorial validity and reliability of the ALSFRS-R: a revision of its measurement model[J]. J Neurol,2017,264(7):1413-1420.

[13] Mehdipour A, Teshler L, Dal Bello-Haas V, et al. Assessing the measurement properties of the self-administered amyotrophic lateral sclerosis functional rating scale-revised (ALSFRS-R): a rasch analysis [J]. Phys Ther,2023,103(11):pzad109.

[14] Jones PS, Lee JW, Phillips LR, et al. An adaptation of Brislin's translation model for cross-cultural research[J]. Nurs Res,2001,50(5):300-304.

[15] Linacre JM, Heinemann AW, Wright BD, et al. The structure and stability of the functional independence measure[J]. Arch Phys Med Rehabil,1994,75(2):127-132.

[16] Lerdal A, Opheim R, Gay CL, et al. Psychometric limitations of the 13-item Sense of Coherence Scale assessed by Rasch analysis[J]. BMC Psychol,2017,5(1):18.

[17] Linacre JM. Rasch analysis of rank-ordered data[J]. J Appl Meas,2006,7(1):129-139.

[18] Wolfe EW, Jr Smith EV. Instrument development tools and activities for measure validation using rasch models: part Ⅱ—validation activities[J]. J Appl Meas,2007,8(2):204-234.

[19] Norhayati MN, Mofreh SAM, Azman YM. Vicarious traumatization questionnaire: psychometric properties using rasch model and structural equation modeling[J]. Int J Environ Res Public Health,2021,18(9).

[20] Wang L, Wu YX, Lin YQ, et al. Reliability and validity of the Pittsburgh Sleep Quality Index among frontline COVID-19 health care workers using classical test theory and item response theory[J]. Journal of Clinical Sleep Medicine,2022,18(2):541-551.

[21] Jayaram M, Wang B S C, Shauver MJ, et al. Using rasch analysis to validate the Michigan hand outcomes questionnaire from the wrist and radius injury surgical trial[J]. Plast Reconstr Surg,2021,148(4):558-567.

[22] Linacre JM. Optimizing rating scale category effectiveness[J]. J Appl Meas,2002,3(1):85-106.

[23] McAlinden C, Skiadaresi E, Moore J, et al. Subscale assessment of the NEI-RQL-42 questionnaire with Rasch analysis[J]. Invest Ophthalmol Vis Sci,2011,52(8):5685-5694.

[24] Buntragulpoontawee M, Khunachiva J, Euawongyarti P, et al. Investigating psychometric properties of the arm activity measure-Thai version (ArmA-TH) sub-scales using the Rasch model[J]. BMC Med Res Methodol,2021,21(1):46.

[25] Fan CW, Chen JS, Addo FM, et al. Examining the validity of the drivers of COVID-19 vaccination acceptance scale using Rasch analysis [J]. Expert Rev Vaccines,2022,21(2):253-260.

[26] Bobevski I, Kissane D, McKenzie D, et al. The Demoralization Interview: Reliability and validity of a new brief diagnostic measure among medically ill patients[J]. Gen Hosp Psychiatry,2022,79:50-59.

[27] 吴明隆. 问卷统计分析实务：SPSS操作与应用[M]. 重庆：重庆大学出版社,2010:543.

Wu ML. Practice of questionnaire statistical analysis: SPSS operation and application[M]. Chongqing: Chongqing University Press,2010:543.

[28] Walker, C. M .What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation [J]. Journal of Psychoeducational Assessment,2011,29(4):364-376.

[29] Frost AG, Orban JA. An examination of an appropriateness index and its effect on validity coefficients[J]. J Bus Psychol,1990,5(1):23-36.

[30] Pinto S, Gromicho M, Swash M, et al. Cervical muscle weakness is a marker of respiratory dysfunction in amyotrophic lateral sclerosis[J]. J Neurol Neurosurg Psychiatry,2020,91(3):323-324.

[31] Maranzano A, Verde F, Colombo E, et al. Regional spreading pattern is associated with clinical phenotype in amyotrophic lateral sclerosis [J]. Brain,2023,146(10):4105-4116.

[32] Fournier CN, Bedlack R, Quinn C, et al. Development and validation of the rasch-built overall amyotrophic lateral sclerosis disability scale(ROADS)[J]. JAMA Neurol,2020,77(4):480-488.

[33] Paganoni S, Macklin EA, Hendrix S, et al. Trial of sodium phenylbutyrate-taurursodiol for amyotrophic lateral sclerosis[J]. N Engl J Med,2020,383(10):919-930.

[34] Xu L, Chen L, Wang S, et al. Incidence and prevalence of amyotrophic lateral sclerosis in urban China: a national population-based study[J].Journal of Neurology Neurosurgery & Psychiatry,2020,91(5):520-525.

（收稿：2024-07-19；修回：2024-09-13；录用：2024-10-07）

（责任编辑：曾　玲）

本文引用格式：

万石磊，杨　丹，刘念荃，等. 中文版肌萎缩侧索硬化多域功能障碍量表信效度分析[J]. 重庆医科大学学报，2025，50(11)：1578-1586.