

基于Hammersmith功能运动量表扩展版的脊髓性肌萎缩症功能评估项目反应理论模型研究

范亚蓓¹, 陈瑾瑾², 葛盼丽¹, 程 曦¹, 张旻宸², 林 枫^{1,3*}

1 南京医科大学第一附属医院, 江苏 南京 210029;

2 南京医科大学康复医学院, 江苏 南京 210003;

3 南京医科大学附属逸夫医院, 江苏 南京 211100

* 通信作者: 林枫, E-mail: peterduus@njmu.edu.cn

收稿日期: 2024-02-24; 接受日期: 2024-09-20

基金项目: 国家重点研发计划项目(2022YFC2009700); 江苏省重点研发计划重点项目(BE2023023-4)

DOI: 10.3724/SP.J.1329.2024.06011

开放科学(资源服务)标识码(OSID):



摘要 目的 基于Hammersmith功能运动量表扩展版(HFMSE)探索建立用于评估脊髓性肌萎缩症(SMA)的项目反应理论(IRT)模型。**方法** 选择2021年2月—2023年6月在南京医科大学第一附属医院老年神经内科及儿科接受修正治疗的SMA患者23例,采用堆栈设计对重复测量数据进行建模;基于似然比计算组成最优模型的HFMSE项目区分度、猜测度、困难度和患者能力水平(θ)参数;采用测试功能差异(DTF)分析量表是否受到康复治疗或性别影响而存在偏倚;采用蒙特卡洛模拟证明模型项目难度分布未受采样过程影响,根据难度预期值中位数绘制人项图,并探讨HFMSE量表中各项目难度、患者能力水平在首次评估和后续评估时的差异。**结果** 本研究共纳入32个项目构建4个备选模型,其中3参数逻辑斯蒂模型(3PLM)具有最佳拟合优度($P > 0.05$, M_2 统计量=479.75)。3PLM中所有项目谨慎度均为1,可计算出项目的区分度、猜测度和难度,且该模型在性别和康复治疗这2个因素上无明显评估偏倚。3PLM有较高信度(Cronbach's $\alpha = 0.955$, LCRC=0.981)和效度[θ 值与原始总分存在正相关关系($r = 0.99$, $P < 0.001$)],可通过二项式换算患者能力水平 θ 值与量表总分。通过可展示项目参数和个人参数在患者能力水平(θ)标尺上位置分布的人项图,观察 θ 值与项目难度的关系。**结论** 基于HFMSE建立的IRT模型可为不同场景SMA患者功能评估和制订康复治疗决策提供建议,但应注意其项目猜测度。

关键词 脊髓性肌萎缩症; Hammersmith功能运动量表扩展版; 项目反应理论; 功能分析

脊髓性肌萎缩症(spinal muscular atrophy, SMA)是脊髓前角及延髓运动神经元变性导致的一种常染色体隐性遗传的退行性神经肌肉疾病^[1],主要临床特征为肢体近端和躯干进行性、对称性肌无力和肌萎缩,具有较高的致死率和致残率^[2-3]。中国新生儿SMA发病率为1/9 788,人群携带率高达1/50^[3-4]。随着疾病修正治疗药物临床应用与推广,SMA患者自然病程呈良性趋势^[5]。疾病修正治疗结合康复可改善SMA患者运动功能、心理状态和生活质量,这

有赖于准确且高效的评估^[6-7]。

Hammersmith功能运动量表扩展版(Hammersmith function motor scale expanded, HFMSE)在Hammersmith功能运动量表(Hammersmith function motor scale, HFMS)基础上添加了13个项目附加模块,适用于2岁及以上能够坐和步行的SMA患者粗大运动功能评估,耗时10~30 min。目前,HFMSE被广泛应用于SMA患者自然病程观察和临床药物试验评估中,具有良好的信度和效度^[3,8]。但是,HFMSE

引用格式: 范亚蓓, 陈瑾瑾, 葛盼丽, 等. 基于Hammersmith功能运动量表扩展版的脊髓性肌萎缩症功能评估项目反应理论模型研究[J]. 康复学报, 2024, 34(6): 604-610, 632.

FAN Y B, CHEN J J, GE P L, et al. Item response theory model for functional assessment of spinal muscular atrophy based on the Hammersmith functional motor scale expanded [J]. Rehabil Med, 2024, 34(6): 604-610, 632.

DOI: 10.3724/SP.J.1329.2024.06011

©《康复学报》编辑部, 开放获取 CC BY-NC-ND 4.0 协议

© Rehabilitation Medicine, OA under the CC BY-NC-ND 4.0

在评估效率和评治结合方面仍有待改进。首先,使用HFMSE对SMA患者进行评估时,是否因为接受康复治疗等因素而存在偏倚?其次,SMA患者需进行HFMSE的33个项目评估,而SMA患者存在体能差、移动难的问题,是否具有更为便捷的评估方法?再次,HFMSE现有评估项目难度与SMA患者个人能力间的关系如何?

项目反应理论(item response theory,IRT)为解决HFMSE评估上述疑问提供了理论基础和实践工具^[9-10]。IRT模型(item response models,IRMs)可以用相同单位(logit)对项目难度和个人能力进行比较,把个人能力和项目难度放到同一张表格中,赋予量表新的参数进行分析;IRT还可以分析测试功能差异(differential test functioning,DTF),排除治疗前后评估的偏倚,更有效地评价治疗前后的功能变化^[11-12]。因此,本研究基于IRT对SMA患者的HFMSE评估结果进行分析,为优化SMA患者的评估及指导康复治疗提供新思路。

1 临床资料

1.1 病例选择标准

1.1.1 诊断标准 参照北京医学会医学遗传学分会和北京罕见病诊疗与保障学会发布的《脊髓性肌萎缩症遗传学诊断专家共识》^[2]有关SMA的诊断标准。

1.1.2 纳入标准 ①符合《脊髓性肌萎缩症遗传学诊断专家共识》2或3型患者;②接受诺西那生钠注射液修正治疗(简称修正治疗);③可配合进行HFMSE评估;④患者本人或监护人知情同意并自愿签署知情同意。

1.1.3 排除标准 ①伴有其他严重疾病;②近期接受手术治疗。

1.2 一般资料

根据小样本研究的最大变异原则^[13],纳入便利样本。选择2021年2月—2023年6月在南京医科大学第一附属医院老年神经内科及儿科接受修正治疗的SMA患者23例。月龄13~417个月;男14例,女9例;SMA 2型患者13例,3型患者10例;修正治疗期间接受康复治疗8例,未接受康复治疗15例;平均评估次数(5.74±1.66)次,其中评估次数≥6次的SMA患者14例,评估3~5次的SMA患者9例;HFMSE评分总分2~62分,平均HFMSE评分(27.47±15.17)分。评估时间为首次修正治疗前1d,治疗后第14、28、63、183、303、423天,真实随访时间与

预定时间前后不超过1d。23例SMA患者共进行132次评估。本研究入组患者月龄、性别、评估次数和能力水平覆盖取值范围,且具有较大异质性,符合本研究的采样原则。本研究方案经南京医科大学附属逸夫医院伦理委员会审批通过(审批号:2018-SR-017)。

2 方法

2.1 数据采集

由1名获得HFMSE评估资质的专业人员对SMA患者进行一对一访谈及评估。采集SMA患者一般资料(包括年龄、性别、病程、疾病分型、是否接受了康复治疗等),并进行HFMSE评估。对每次HFMSE量表评估结果的数据进行整理,形成患者个人能力水平和项目难度的样本数据。HFMSE量表每个项目评分为3级制(0~2分),每个测试项目无法完成为0分;需要调整/适应/代偿的情况下完成动作为1分;无需调整/适应/代偿的情况下完成动作为2分^[14]。

2.2 IRT建模

参考本课题组前期研究方法进行IRT建模^[10-11,15-17]。根据重复测量设计对主要参数和分析过程进行调整。

2.2.1 数据预处理 剔除缺失值≥30%和取值始终不变的项目。对于缺失值<30%的项目,采用基于潜在类别分析的多重数据填补过程进行填补。对其余项目进行二值化处理,即评估得分由0-1-2转写为0-1-1。经此转换,得分越高表示患者功能状况或健康水平越高。

2.2.2 模型建构 采用堆栈设计(stack design)对重复测量数据进行建模^[1],即对于*i*例被试和*j*个测次,共有*i*×*j*条观测。在这些观测基础上建立IRT模型,共有4个备选logistic模型,根据参数数量分别记为单参数(1PLM)、2参数(2PLM)、3PLM和4PLM。

2.2.3 模型选择 对每个模型进行基于 M_2 统计量的拟合优度检验,对拟合优良的模型($P>0.05$)进行基于似然比的模型间比较,以贝叶斯信息准则(Bayesian information criterion,BIC)取值更小的模型作为最终IRM。以标记卡方检验(signed chi-squared test, $S-\chi^2$)分析每个项目的拟合优度($P>0.05$ 表示拟合优良)。

2.2.4 模型假设检验 模型需要符合3个假设,分别是单维性[模型只测算一种潜在特质(SMA患者能力水平)]、局部独立性(项目只受潜在特质影

响)、单调性(能力越高的个体得分越高)。本研究通过计算同质系数(homogeneity coefficients, H)判断单维性^[18],采用STRAAT等^[19]提出的条件关联性检测算法验证局部独立性,通过绘制项目特征曲线(item characteristic curves, ICCs)判断单调性。

2.2.5 模型参数 为所选模型计算参数,包括区分度(测评项目区分个人能力高低的程度)、困难度(测评项目得分的难度)、猜测度(测评项目因猜测而得分的可能性)和谨慎度(测评项目因不慎未得分的可能性)。

2.2.6 模型信度和效度 根据克隆巴赫 α (Cronbach's α)和潜在类别信度系数(latent class reliability coefficient, LCRC)检验模型信度,以信度指标 >0.9 为高信度。本研究通过模型估算的 θ 值与原始得分的相关性分析检验模型效度。

2.2.7 偏倚分析 采用测试功能差异(differential test functioning, DTF)分析量表是否受到康复治疗或性别影响而存在偏倚^[20]。

2.2.8 蒙特卡洛模拟和人项图 针对小样本问题,本研究使用蒙特卡洛模拟证明模型项目难度分布未受采样过程影响。构建500个模拟数据集,各含1 000条复本,以此计算项目难度的经验值与估计值。根据难度预期值中位数绘制人项图,探讨HFMSE量表中各项目的难度、患者能力水平在首次评估和后续评估时的差异。

2.3 统计学方法

数据分析采用R软件(4.2.2版)。数据填补采用基于POLCA包的MILCA法^[21]。IRT建模、DTF检

验和蒙特卡洛模拟采用mokken和mirt工具包^[22-23],根据摩肯量表的4个条件(单维性、局部独立性、单调递增性和定序性)对HFMSE数据进行摩肯量表分析(MSA);模型的单维性采用自动化条目筛选过程进行检验;局部独立性检验采用基于似然比的残差相关性分析;单调性采用单侧单样本Z检验^[23]。相关关系和组间比较可视化采用ggstatsplot包^[24],效度分析采用Pearson卡方检验,实测值与蒙特卡洛模拟的预期中位数采用Kendall秩相关检验。统计检验水平以 $P<0.05$ 为差异具有统计学意义。

3 结果

3.1 IRT模型

在33个测试项目中,项目1(坐在治疗床/椅子上)取值始终为2,该项目无法区分个体能力高低,予以剔除。本研究共纳入32个项目构建4个备选模型。其中,3PLM拟合良好($P>0.05$, M_2 统计量=479.75, $df=432$),整体同质系数 $H=0.846$,标准误(SE)=0.020。这32个项目构成的量表具有强同质性^[23],可以认为其符合单维性。结合 H_θ 的取值分布(>0)和条件关联性检测算法,32个项目均未违反局部独立性。所有项目的ICCs均为上升的S形曲线,符合单调性形态。3PLM的项目参数和项目拟合指标,谨慎度均为1。尽管项目5拟合不良($P<0.05$),经综合考虑DTF结果和量表完整性,决定保留该项目^[25]。32个测试项目按照区分度递减排序。见表1。

表1 3PLM模型项目参数和拟合指标

Table 1 Item parameters and goodness-of-fit indexes of 3PLM

项目编号	项目指令	区分度	猜测度	困难度	S- χ^2 值	df	RMSEA
30	使用扶手上台阶	66.408 9	0	1.098 1	1.668 1	26	0
9	仰卧向左翻到俯卧	63.177 6	0	-0.890 9	1.496 5	26	0
19	独站	62.994 2	0	0.688 0	2.541 9	26	0
3	单手摸头	46.615 8	0.601 1	-0.530 3	10.893 6	26	0
31	使用扶手下台阶	42.148 7	0	1.184 7	1.206 5	26	0
5	仰卧翻至侧卧	37.785 0	0	-1.394 5	233.623 8	26	0.246 9
6	俯卧向右翻至仰卧	36.314 0	0.077 2	-0.989 3	9.618 0	26	0
4	双手摸头	16.982 6	0.168 0	-0.375 4	14.895 9	26	0
20	行走	15.874 7	0	0.830 1	12.112 2	26	0
28	下蹲	13.285 1	0	1.046 3	6.082 9	26	0
29	向前跳跃30 cm	12.040 1	0	1.963 2	0.774 7	26	0
10	坐位到仰卧	10.777 6	0.092 9	-0.369 6	14.007 6	26	0
8	仰卧向右翻到俯卧	8.854 2	0	-0.888 7	9.815 2	26	0
16	爬行	8.125 3	0	0.202 6	6.775 1	26	0

续表1

项目编号	项目指令	区分度	猜测度	困难度	S- χ^2 值	df	RMSEA
27	站立转换为坐在地上	7.353 8	0	1.132 8	13.116 6	26	0
13	直臂支撑	6.573 2	0.080 5	-0.261 5	15.859 1	26	0
33	不使用扶手下台阶	6.050 5	0	1.539 7	31.735 2	26	0.041 0
32	不使用扶手上台阶	5.744 8	0	1.607 2	38.435 8	26	0.060 4
11	前臂支撑	5.612 8	0.031 1	-1.052 3	5.492 9	26	0
15	四点支撑	5.558 8	0.056 0	-0.442 8	8.849 4	26	0
18	扶持站立	5.431 6	0	0.481 8	17.891 4	26	0
7	俯卧向左翻至仰卧	5.297 6	0	-1.047 9	23.216 2	26	0
22	仰卧位左髋屈曲	3.953 9	0	-1.242 4	33.471 7	26	0.046 8
24	直跪转换至左膝半跪	3.861 3	0	0.861 0	11.723 1	26	0
21	仰卧位右髋屈曲	3.843 5	0	-1.126 2	14.174 8	26	0
23	直跪转换至右膝半跪	3.827 1	0	0.806 7	11.195 9	26	0
12	俯卧抬头	3.757 3	0.150 4	-0.441 6	9.353 2	26	0
14	仰卧坐起	3.326 5	0	0.054 0	36.915 4	26	0.056 6
26	右腿从跪立转换为站立	2.403 3	0	1.976 7	9.483 3	26	0
25	左腿从跪立转换为站立	2.403 1	0	1.976 9	9.485 3	26	0
17	仰卧位抬头	2.211 1	0	0.382 9	17.143 8	26	0
2	直腿坐	1.927 5	0.281 4	-0.848 7	26.002 2	26	0.000 8

根据性别和康复治疗2种变量进行测度不变性检验。分别以女性和未接受康复治疗为对照组,以男性和接受康复治疗为焦点组。结果显示,

HFMSE评估在性别、康复治疗2个因素上DTF差异无统计学意义($P>0.05$),与性别和康复治疗实施的预期间卷得分分布曲线基本重叠。见图1。

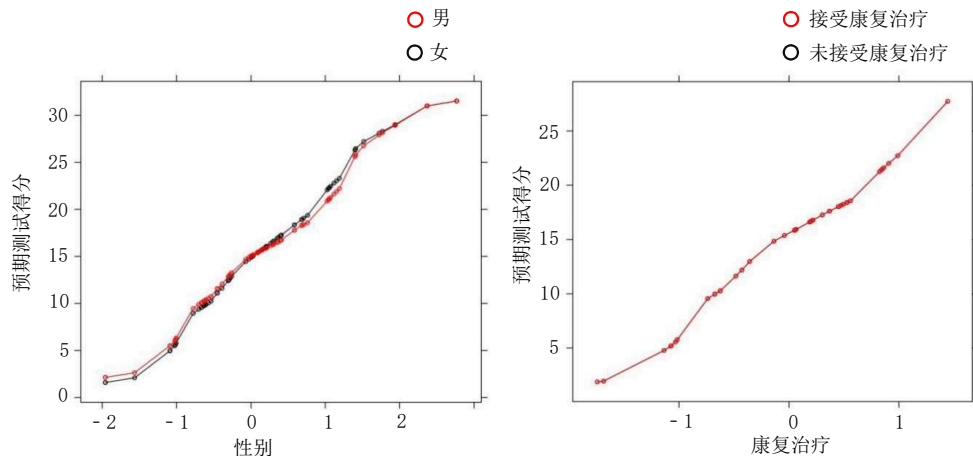


图1 基于2个因素的DTF检验

Figure 1 DTF test based on two factors

3.2 模型信度和效度

所选3PLM有较高信度(Cronbach's $\alpha=0.955$, LCRC=0.981)和效度[θ 值与原始总分存在正相关关系($r=0.99, P<0.001$)].因此患者能力水平 θ 值与量表总分(TTS)可以用如下二项式进行换算:

$$\theta = -1.63 + 0.065 1 TTS - 1.39 \times 10^{-3} TTS^2 \quad (\text{公式} 1)$$

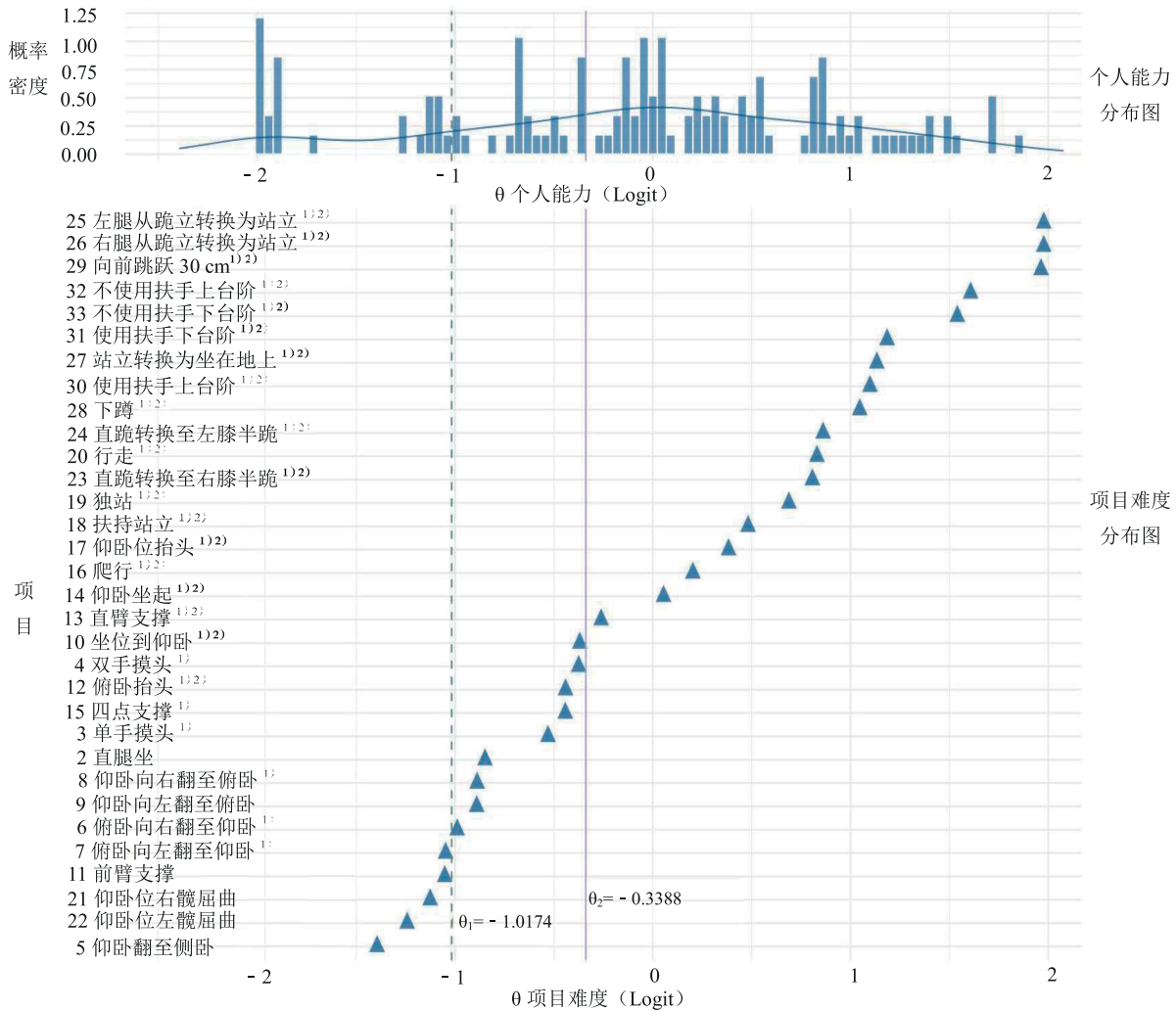
3.3 个人能力与项目难度比较

蒙特卡洛模拟得到预期值的中位数与实测值进行Kendall秩相关检验结果表明,两者呈明显相关性($\tau=0.996, P<0.001$).因此本研究根据难度预期值的中位数,绘制个人能力水平与项目难度关系的人项图。见图2。上图为带有概率密度曲线的直

方图,显示患者能力 θ 值分布的直方图和概率密度曲线,曲线下面积表示患者分布在相应 θ 值取值范围内的概率。此图呈现为与正态分布近似的钟型分布。下图为项目难度分布图,从左下到右上难度递增。由于上下2个图统一采用Logit作为度量单位,因而上图横轴的个人能力值与下图横轴的项目难度值可以直接比较。该图使用时可根据某个患者的能力值绘制一条垂直线,将下图的项目分为左侧胜任区(垂直线左侧项目难度低于患者能力)和右侧挑战区(右侧项目难度高于患者能力)。

以1位3岁2型SMA男童为例,图2中虚竖线表

示该患者第1次评估时的个人能力水平($\theta_1=-1.0174$ Logit);实竖线表示该患者第5次评估时的个人能力水平($\theta_2=-0.3388$ Logit),随着能力的提升,他能胜任HFMSE评估的项目增多。与此例相似,所有患者的个人能力水平均可根据评估结果,通过“3.2”项下“公式1”计算出对应的数值,并在图2横坐标中找到对应位置。绘制图2的虚、实竖线,从而直接观察该患者个人能力水平的变化情况,及时发现SMA患者存在功能障碍的评估项目,为下一阶段康复训练提供参考方向。



注:1) 表示第1次评估存在障碍的项目;2) 表示第5次评估存在障碍的项目。

Note: 1) indicates item with an obstacle in the first evaluation; 2) indicates item with an obstacle in the fifth evaluation.

图2 人项图

Figure 2 Person-item plot

4 讨论

本研究尝试基于HFMSE评估探索建立用于评估SMA的项目反应理论模型,结果显示3PLM具有较好的拟合度和信效度,可以通过建立IRT模型的方式对HFMSE的测量学特征进行探讨。在此基础上,将HFMSE评估用于SMA患者时,不会因患者性别或是否接受康复治疗而产生评估结果的偏差,这与既往相关研究一致^[3,8]。本研究结果还推算出HFMSE量表各评估项目的区分度、猜测度和困难度,促使其更好地使用,为SMA患者临床康复治疗决策提供指导建议。

4.1 基于HFMSE的IRT模型可为不同场景SMA患者功能评估提供建议

IRT提供了一个框架,将被试项目难度和个体能力水平联系起来^[26-27]。本研究结果显示,HFMSE有12个测试项目(编号30、9、19、3、31、5、6、4、20、28、29和10)具有较大的区分度。若这12个项目通过二值化(“有”或“无”)方式评估SMA患者功能障碍情况,可以很大程度上降低多层次打分标准带来的评估困难,可用于SMA患者居家自测,及时发现SMA患者功能状况变化。而其他区分度较低的测试项目,可供专业医护人员在医疗机构对SMA患者进行评估,建议区分度较低的项目在现有基础上增加评分级别,以更好地辨别SMA患者在某一项活动能力水平上的变化。

4.2 基于HFMSE的IRT模型评估应关注部分项目猜测度

本研究结果显示,HFMSE中第2、3、4、12等项目猜测度较高,这些项目活动难度较低,而打分要求较为严格。如项目2直腿坐,要求被测试人员需做到髌骨向上,脚踝之间距离 $<10\text{ cm}$,髋关节处于中立位(无内外旋)等。若评估人员对HFMSE具体评分要求掌握不够充分,仅根据项目内容默认被测试人员达到直腿坐就赋分,容易导致评估偏差。因此,在开展HFMSE评估培训及日常评估时,应重点关注猜测度较高的项目,以避免在评估中出现误差。

4.3 基于HFMSE的IRT模型评估结果可为SMA患者康复治疗决策提供参考

本研究结果显示,HFMSE项目5在图2的左下方,难度最低;项目25在图2的右上方,难度最高。项目5要求SMA患者可采用任何方式(除了抓垫子边缘外)实现翻身至侧卧,动作要求较低;与其他评估项目比较,翻身是SMA患者生活中较常用的活动,

患者也更熟悉其相关活动技巧,该项目在评估时更容易得分。这与BIENIASZEWSKA等^[28]研究结果一致。项目25、26、29对SMA患者肌力、动作技巧等有较高要求,项目难度较大,SMA患者在评估过程中较难得分。这提示基于HFMSE的IRT模型评估结果可体现SMA患者不同能力水平,根据人项图可为SMA患者选择难度适宜的项目进行康复训练,提升康复效率。

有研究指出应根据SMA患者功能状态制订康复目标,但既往研究较少将SMA患者能力水平与评估项目之间建立关系^[2,29]。本研究基于HFMSE构建的IRT模型,综合考虑了项目难度和患者能力水平,医务人员和患者均可根据评估结果,分析患者个人能力水平及其变化,及时发现患者能力水平线下仍未得分的项目或稍高于其能力水平的项目。这些内容可优先考虑作为该患者下一步的功能目标,从而为康复治疗提供可量化的决策依据^[16],助力临床更合理、有效地开展康复训练。

此外,人项图显示,HFMSE测评项目的难度水平分布并不均衡,部分难度系数对应的评测项目较多,而少部分难度系数对应的评测项目很少甚至空白。为了更严谨、全面地评估SMA患者功能变化情况,建议进一步优化、补充HFMSE的测评项目。

5 小结

基于HFMSE的IRT模型可用于评估不同场景SMA患者功能变化,并为SMA患者康复治疗提供参考,但应注意其项目猜测度。本研究仍存在一些不足之处:①纳入样本量偏小,仅采用了小样本的最大变异原则和堆栈法建模设计;②本研究采取数据二值化原则进行评估,部分完成动作即可得分,未能较好地区分患者某个项目完成度的不同;③观察项目以粗大运动功能为主,未能全面分析SMA患者的功能情况;④未针对不同年龄、功能状况SMA患者进行分类分析;⑤HFMSE项目未能覆盖所有难度区间。因此,还需开展多中心、大样本研究,细分不同年龄、能力水平,优化、补充HFMSE测评项目,以推动基于HFMSE的IRT模型在SMA患者功能评估和康复治疗中的应用。

参考文献

- [1] ANDRICH D, MARAIS I. A course in rasch measurement theory: measuring in the educational, social and health sciences [M]. Singapore: Springer Nature Singapore, 2019: 299-305.
- [2] 北京医学会医学遗传学分会, 北京罕见病诊疗与保障学会. 脊髓性肌萎缩症遗传学诊断专家共识[J]. 中华医学杂志, 2020,

- 100(40):3130-3140.
 Medical Genetics Branch of Beijing Medical Association, Beijing Society for Diagnosis of Treatment and Guarantee of Rare Diseases. Expert consensus on genetic diagnosis of spinal muscular atrophy [J]. *Natl Med J China*, 2020, 100(40):3130-3140.
- [3] 中国医师协会儿科医师分会, 中国医师协会儿科医师分会儿童呼吸学组. 脊髓性肌萎缩症呼吸管理专家共识(2022版)[J]. *中华实用儿科临床杂志*, 2022, 37(6):401-411.
 Pediatrician Branch of Chinese Medical Doctor Association, Sub-specialty Group of Respiratory Diseases of Pediatrician Branch of Chinese Medical Doctor Association. Consensus statement for standard of pulmonary management in spinal muscular atrophy (2022 version) [J]. *Chin J Appl Clin Pediatr*, 2022, 37(6):401-411.
- [4] LI C, GENG Y F, ZHU X D, et al. The prevalence of spinal muscular atrophy carrier in China: evidences from epidemiological surveys [J]. *Medicine (Baltimore)*, 2020, 99(5):e18975.
- [5] PITARCH CASTELLANO I, CABRERA-SERRANO M, CALVO MEDINA R, et al. Delphi consensus on recommendations for the treatment of spinal muscular atrophy in Spain (RET-AME consensus) [J]. *Neurol Engl Ed*, 2022, 37(3):216-228.
- [6] PRANANTA M S, RAHMAWATI M M, GOESASI R Z. Habilitation of child with spinal muscular atrophy type 2: a case report [J]. *Surabaya Phys Med Rehabil J*, 2022, 4(2):81-89.
- [7] MIREA A, LEANCA M C, ONOSE G, et al. Physical therapy and nusinersen impact on spinal muscular atrophy rehabilitative outcome [J]. *Front Biosci (Landmark Ed)*, 2022, 27(6):179.
- [8] SALAZAR R, MONTES J, DUNAWAY YOUNG S, et al. Quantitative evaluation of lower extremity joint contractures in spinal muscular atrophy: implications for motor function [J]. *Pediatr Phys Ther*, 2018, 30(3):209-215.
- [9] THISSEN D, STEINBERG L. The history of item response theory in 20th century [J]. *Chin Eng J Educ Meas Eval*, 2020, 1(1):43-59.
- [10] JIANG Y E, ZHANG D M, JIANG Z L, et al. ICF-based simple scale for children with cerebral palsy: application of Mokken scale analysis and Rasch modeling [J]. *Dev Neurorehabil*, 2023, 26(2):71-88.
- [11] 徐浪, 张芹, 江钟立, 等. 基于ICF和项目反应理论建模评估老年终末期肾病血液透析患者的活动和参与的研究[J]. *中国康复医学杂志*, 2023, 38(10):1385-1392.
 XU L, ZHANG Q, JIANG Z L, et al. ICF-based assessment of activities and participation for aged persons with end-stage renal disease and maintenance hemodialysis via item response theory modeling [J]. *Chin J Rehabil Med*, 2023, 38(10):1385-1392.
- [12] NUGENT W R. Understanding DIF and DTF: description, methods, and implications for social work research [J]. *J Soc Soc Work Res*, 2017, 8(2):305-334.
- [13] PATTON M Q. *Qualitative research & evaluation methods: integrating theory and practice* [M]. 4th Ed. California, USA: Sage Publications, 2015:283-284.
- [14] STĘPIEŃ A, GAJEWSKA E, REKOWSKI W. Motor function of children with SMA1 and SMA2 depends on the neck and trunk muscle strength, deformation of the spine, and the range of motion in the limb joints [J]. *Int J Environ Res Public Health*, 2021, 18(17):9134.
- [15] CHEN J J, ZHU Z Y, BIAN J J, et al. Nutrition-associated health levels in persons with cancer: item response modelling based on the international classification of functioning, disability and health [J]. *Eur J Phys Rehabil Med*, 2023, 59(5):593-604.
- [16] FENG C, GENG B F, LIU S G, et al. Activity and participation in haemophiliacs: item response modelling based on international classification of functioning, disability and health [J]. *Haemophilia*, 2023, 29(1):308-316.
- [17] FENG C, JIANG Z L, SUN M X, et al. Simplified post-stroke functioning assessment based on ICF via dichotomous mokken scale analysis and rasch modeling [J]. *Front Neurol*, 2022, 13:827247.
- [18] STRAAT J H, VAN DER ARK L A, SIJTSMA K. Comparing optimization algorithms for item selection in mokken scale analysis [J]. *J Classif*, 2013, 30(1):75-99.
- [19] STRAAT J H, VAN DER ARK L A, SIJTSMA K. Using conditional association to identify locally independent item sets [J]. *Methodology*, 2016, 12(4):117-123.
- [20] SCHNEIDER L, STROBL C, ZEILEIS A, et al. An R toolbox for score-based measurement invariance tests in IRT models [J]. *Behav Res Meth*, 2022, 54(5):2101-2113.
- [21] SULIS I, PORCU M. Handling missing data in item response theory. Assessing the accuracy of a multiple imputation procedure based on latent class analysis [J]. *J Classif*, 2017, 34(2):327-359.
- [22] CHALMERS R P. Mirt: a multidimensional item response theory package for the R environment [J]. *J Stat Soft*, 2012, 48(6):1-29.
- [23] SIJTSMA K, VAN DER ARK L A. A tutorial on how to do a Mokken scale analysis on your test and questionnaire data [J]. *Br J Math Stat Psychol*, 2017, 70(1):137-158.
- [24] PATIL I. Visualizations with statistical details: the 'ggstatsplot' approach [J]. *J Open Source Softw*, 2021, 6(61):3167.
- [25] ZHAO Y, HAMBLETON R K. Practical consequences of item response theory model misfit in the context of test equating with mixed-format test data [J]. *Front Psychol*, 2017, 8:484.
- [26] VAN DE WINCKEL A, OTTIGER B, BOHLHALTER S, et al. Comprehensive ADL outcome measurement after stroke: rasch validation of the lucerne ICF-based multidisciplinary observation scale (LIMOS) [J]. *Arch Phys Med Rehabil*, 2019, 100(12):2314-2323.
- [27] BOONE W J, STAVER J R. *Advances in Rasch analyses in the human sciences* [M]. Cham, Switzerland: Springer International Publishing, 2020:216-217.
- [28] BIENIASZEWSKA A, SOBIESKA M, STEINBORN B, et al. Examination of upper limb function and the relationship with gross motor functional and structural parameters in patients with spinal muscular atrophy [J]. *Biomedicine*, 2023, 11(4):1005.
- [29] MERCURI E, FINKEL R S, MUNTONI F, et al. Diagnosis and management of spinal muscular atrophy: part 1: recommendations for diagnosis, rehabilitation, orthopedic and nutritional care [J]. *Neuromuscul Disord*, 2018, 28(2):103-115.

(下转第632页)