

• 国外医学动态 •

临床疼痛治疗病人报告正性转归测量方法的开发

——来自 MM PACT 的推荐

1 介绍

临床上评价疼痛治疗效果经常会碰到两个难题: 正性转归的定义及如何最好地评测疗效。正性转归的定义可以反映出评价方的评议流程和评议标准。针对目前临床过程采用不同的预后评测指标从而妨碍了对治疗效果评估的现状, MMPACT (the Initiative on Methods Measurement and Pain Assessment in Clinical Trials) 推荐了评估慢性疼痛治疗的六项核心内容: (1) 疼痛状况; (2) 躯体机能; (3) 情绪机能; (4) 参与度及满意度; (5) 症状及副作用; (6) 安排部署。

根据不同的信息来源, 对正性转归效果评估的研究报告可分为以下几类: 病人自陈报告 (PROs)、实验室检验报告、仪器检测报告、行为观察报告、医师报告 (CROs) 以及第三方报告。每一类报告均能从不同角度提供数据, 例如, 病人自陈报告记录了他们对疾病和治疗影响健康状况和机能程度的自我体验, 包括病人对健康状况、疾病状况、坚持治疗情况和治疗满意度的评估, 以及疾病对躯体机能和生活满意度的影响。实验室检验报告、行为观察报告和仪器检测报告可通过等级评分获得客观、定量的行为或生理学数据。医师报告包括研究人员观察得到的数据, 也可由卫生保健工作人员根据病人信息获得的数据构成。被试者参与获得的医师报告有别于医生指导下完成的病人自陈报告, 因为前者有医生的判断或解释, 而后者则是病人自己的陈述。

事实上, 即使综合实验室检验报告、行为观察报告、医师报告和病人自陈报告的研究结果, 也很难完美地做到重点突出、系统全面、兼顾主客观地对正性转归效果进行评估。当医生无法客观地评价病人的感觉如疼痛及疲劳时, 病人自陈报告就会特别适用。

基于系列研究, MM PACT 推荐了一套可用于评估疼痛、躯体机能、情绪机能以及对治疗的参与度和满意度的病人自陈报告, 尽管给出了详细的测量方法, 但每种方法都有特定的使用条件, 尤其要注意的是: (1) 不应为了结果有意义而选择特定被试; (2) 指导语、实验内容以及评定动机必须清楚地记录。MMPACT 的建议中不仅鼓励收集现有方法可获取的数据, 同时也呼唤新的评估技术。第三次

MM PACT 会议的召开, 就是为了在评估和发展正性转归的测量方法上进一步达成一致。

2 专题会议流程

第三次 MM PACT 会议于 2003 年 11 月 20 日 ~ 22 日召开, 与会者包括 44 名来自不同领域 (学术界、政府、民间组织和制药行业) 的人士, 他们是实施和评估慢性疼痛治疗实验的基础研究、临床以及管理专家。会议的主要议题是, 开发新的病人自陈报告, 在充分恰当地对疼痛治疗临床实验的信度和效度的评估方面达成一致。

并非所有与会者都熟悉心理测量学原理的最新进展、项目反应理论 (IRT) 和电脑适性化测验 (CAT), 所以会前将一份关于项目反应理论背景介绍的文章发给了与会成员, 此外, 会上还由一位 IRT 和 CAT 的撰文作者做了综述报告。

3 疼痛试验中病人自陈报告开发的推荐步骤

开发病人自陈报告需要经过一系列连续的步骤, 确定潜变量, 搭建整体结构, 明确评价的目的、用途和被试者的特点。

测量方法的选择要基于病人的利益、治疗方法的特点以及预期的治疗结果。现有方法的局限性使新方法的开发势在必行。不过, 对一些现有的有效的量表来说可能需要做的只是增加部分项目。开发新的正性转归的评价方法时需要考虑测量方法的实际价值、是否适用于被试人群、保证结果可靠和有效的同时还要使病人的负担最小化。这就需要一系列完整的开发步骤。

开发新的测量方法会花费大量的时间、精力和财力, 因此在此之前必须对已有的方法进行检验, 以确保开发新方法的必要性。对已有的、可行的测量方法, 不仅要注意其对心理属性测量的效度, 同时还要关注其对相关人群实施时的可用性, 当然, 在合理、科学的范围内, 一定的灵活性也是允许的。然而, 一旦需要, 开发新的治疗结果评价方法就要经过严密的步骤。此外, 新方法的设计者还需说明该方法的新增优势, 例如降低病人负担、信度效度增强等。如果一种新的测量方法可以在更短时间内测出同样多的心理属性, 那么它就会受到欢迎。

(1) 测量要素

临床正性转归评价方法的开发初步完成后,要对其心理测试属性进行评价,特别是其准确性、可靠性和有效性。需要着重指出的是,测量的敏感度即对病人变化的反应在设计之初就应该密切关注,而且,测量结果必须要能够评价与临床相关的变量。上述这些属性都是必要的,缺少这些属性的新的测量方法将不会对评估正性转归提供任何帮助。

(2)改进的内容和项目

测量的内容应该覆盖需要评估的各种属性。例如,一种新的评价疼痛对情绪状态影响的测量方法,应当包括评价正性和负性情绪两方面的内容。许多研究表明,病人和医生对健康状况、疾病的影响和治疗结果的偏爱都有不同的认识。特殊疾病或症状的病人们对疾病和治疗对其躯体机能的影响有独特的看法,这点对于获得新的测量方法非常重要。研究病人对疾病的认识和个体经验,能够获得医生所不知的知识,这些在开发临床治疗的正性转归测量方法中都值得注意。

确定病人认为重要的内容模块可以采用焦点小组和个体面谈结合的方式。焦点小组的成员选取要包含疼痛程度不同的各种病人,因为症状程度的不同会对正性转归的评价产生不同的影响。同时,测评时,诸如年龄、性别、种族等的人口统计学变量也应该加以考虑。

还有一些重要的因素在开发正性转归的测量方法中没有得到足够的重视。首先,评价躯体机能的测量时,是测量疼痛影响到的躯体机能还是测量病人整体的躯体机能,或者二者均测量并不确定,因此,必须考虑病人的情况,以决定采用哪种测量方法。例如,在脊髓损伤研究时,患者不能上楼梯的行为更多涉及到机能损伤,而骨关节炎病人不能上下楼梯则更多涉及到的是疼痛,此时,评价疼痛对躯体机能的影响就比评价躯体功能受限程度的敏感度要高。同样,评价情绪状态既可以测量焦虑和抑郁的总体水平,也可以测量疼痛对情绪状态的影响。

其次,越来越多的国家在进行临床疼痛治疗试验,疗效的评估方法应能对其中的概念和内容作出轻松的解释并方便移植于不同的文化。但同时,即使许多内容可以逐句翻译,但在不同文化背景之下,必须注意翻译的概念是否一致,解释是否有意义。

再者,一旦选定了内容、明确了项目,就要开始注意指导语、选项措辞、测查时间、常模等内容。必须使被试清楚理解指导语和项目的内容,明确测查要求的时间(如上周、上个月还是现在),因为对不

同时间间隔的回答有不同的解释。此外,要注意对较长时间以前的回顾所形成的结果报告,其准确度会受到记忆力、当前症状和一些特定事件的影响。

量表初步制定好后,开始进行预实验,抽样选取部分被试,考查量表的指导语、项目措辞等是否适用于目标人群。预实验可揭示测量维度的缺项、平衡测量项目的级差,根据预实验结果可对量表修订和重测,这一过程反复多次,直到形成满意的量表。

对指导语、项目内容等的修改和测量过程需要详细记录。同时,样本量的大小和样本的典型性也是量表修订者需要提前考虑的心理测量属性。

(3)项目测验理论

经典测验理论(CTT)历史悠久,一直是检验测量方法是否合适的重要工具。经典测验理论假设回答者的分数与其对一个测量属性(如:疼痛)所涉及的一系列问题的反应线性相关。该理论的一个重要缺陷是其对被试的依赖性,对于一个量表,利用经典测验理论对某一类被试做出的可行性检验结论并不意味着该量表在其他被试那里也同样适用。经典测验理论的检测不可能海纳百川,涵盖所有的范围,因为我们在测量时,无法要求被试作答时额外增加许多明显与测量目的无关的负担。该理论的另一个缺陷是,对不同测量方法的评估数据无法比较,因为不同的测量方法可能评估的是不同的心理特性,而这些特性之间没有可比性。

RT则弥补了经典测验理论的不足。它是一种由各种非线性数学模型和相关的统计方法组成的统计学理论,它尝试将个体应答与连续分布中的位点对应,以反映出某些难以观察到的结构或潜在特性。RT模型将对一个项目尺度的反应概率作为被试者的量化特征和项目的特征参数。它可以清楚地解释个体在项目水平的反应和测试中项目特征参数之间的功能联系。项目反应和潜变量之间的非线性相关决定了建立在项目反应理论基础上的评估没有被试依赖性,也无需冗余项目,同时不同测量方法的项目维度也可以进行横向比较。

RT模型的目标是用尽可能少的项目来预测个体的健康状况和躯体机能,相反,经典测验理论关注的是测量的参数和维度的分级。相比于基于经典测验理论的测量方法,基于项目反应理论的测量对正性转归的评估有以下优点:①项目尺度中的每一个问题都表征了测量的要素和项目的功能,②对于不同的被试,项目属性和量表的得分均相对恒定,③得分可以进行比较。建立在RT基础上的测量方法,

其优点是整体样本不需要对每一个项目进行回答,因此 IRT 可以更有效地收集数据,同时也是项目评价的一种重要的分析工具。最近,美国国立卫生所资助了一个合作计划——病人报告正性转归测量方法信息系统 (PROMIS),评测 IRT 在疼痛、躯体机能和情绪状态相关领域的应用情况。目前,该项目还处于起步阶段,IRT 对疼痛相关领域的潜在研究价值仍需验证。

IRT 的特点促进了 CAT 在这一领域的应用。CAT 是一种能够将问题调整为适合被试个体回答的交互管理技术。根据被试的回答,测试内容可以进行调整,从利用合适的 IRT 模型进行校准而得到的项目库中选择出适合被试的题目。项目库一经确定和校准,计算机程序就会根据被试者对项目的回答评估被试者的特性水平,继而采用匹配搜索算法搜寻最适合被试者特性水平(如疼痛的严重程度)的项目,该步骤反复进行,直到达到预设的测量精度为止。这种方法减少了项目冗余的需求,降低了被试的负担。不过,对于某些项目假定为不能互换的多维结构,CAT 的应用还值得商榷。

尽管上面提到了许多优点,IRT 也有自身的局限。现今主流的 IRT 模型的核心假设是,项目是一元线性的,也即项目是可测的单一连续潜变量。事实上,当需要明确量表的测试目的时,选择一个项目库,对不同概念进行评价,然后通过 IRT 模型找出潜在的概念上的一元线性关系并确定问题,也是有可能的。于是,很重要的问题就是区分统计学上的一元线性关系和概念上的一元线性关系。疼痛临床实验的许多重要特性是多元线性的,即使 IRT 可以通过减少项目使其成为一元线性,但是删除的这些项目通常在临床疼痛研究中也是非常重要的。

MMPACT 会议在获得临床治疗正性转归测量方法中特别提供了一些重要的项目,如疼痛、躯体机能、情绪状态、病人整体改善状况和副作用,当然,即使把这些内容通过统计学方式组合起来,也还有可能丢失一些关于特定疗效的重要信息。为此,一种解决方案就是,采用因素分析查找反应的协变趋势,如果是多元线性关系,那么,我们就要把每个因素都作为独立的一个维度,从而保持与总体趋同。

IRT 的其它缺点还包括大样本量的需求、缺少方便友好的软件系统和标准化的识别指标。最后,也是最重要的一点,对于包括许多临床科研人员和健康工作者在内的大多数用户而言,IRT 的模型是复杂和难以操作的。

总体而言,IRT 是一种非常有前景的方法,但在推荐临床试验研究前需要进一步的改进。IRT 最有可能用于评估躯体机能,但可能不适用于评估一些诸如疼痛或者疲劳等的症状,因为 IRT 预测潜在的维度(如疼痛经验)时可能产生错误。同时,IRT 模型并不是解决经典测验理论所有遗留问题的灵丹妙药,仍需要在项目和方法上进行系统的改进工作。

4 讨论

在疼痛治疗的临床实验中,缺少一种能够为被试提供足够的可比较信息、适合于不同研究目标、使研究者可以将不同的人口统计学指标以及疾病指征进行合并比较的标准全面的测量方法。如前文所述,基于经典测验理论的测量方法存在许多局限,包括被试的负担、数据无法比较以及线性相关假说的不确定性。IRT 有很多优势,人们逐渐意识到以 IRT 补充甚至替代传统心理测量方法的可能性。随着 IRT 技术的发展和日渐为人们所熟悉,CAT 技术也将得到更多的应用。但是,这些新的技术仍有许多不足之处,前文已经列举。

MM PACT 的建议对于正在着手或者计划进行临床治疗、干预疼痛的研究者和管理者都有指导意义。比如 MM PACT 之前一份取得一致意见的文件介绍,一套小规模临床治疗正性转归测量方法已经可以应用于慢性疼痛的临床治疗。研究该方法的数据表,对于开发正性转归的评价方法是大有裨益的。收集和分析那些已发表的公认的正性转归测量方法中的实验数据,将有助于开发一种高效动态系统,从而建立可与被试匹配的项目库,方便进行治疗结果的比较,并且提高治疗结果的评估精度。一项类似的多学科交叉研究课题已经在美国国立卫生院支持下开展,这项合作不仅是对项目反应理论方法实用性验证的一次应用性的尝试,也是对包括疼痛、睡眠、情绪、躯体机能在内的常规测量项目从多学科、多技术的角度收集数据的一次方法学上的飞跃,这将有助于联合临床研究者、管理者、医药行业,在以病人为导向的临床疼痛实验研究中增强合力,开发更新更好的测评技术。

5 声明

本文所表达的观点仅为作者的观点,没有得到官方(包括美国经验事务部、美国食品药品监督管理局、美国国立卫生院以及提供大量经费赞助的药品公司)的认可。

(Turk DC, et al Pain, 2006 125: 208~ 215, 郑秀丽,郭建友,罗非译)