



**AO Foundation**  
Publishing

# AO骨科量表评鉴

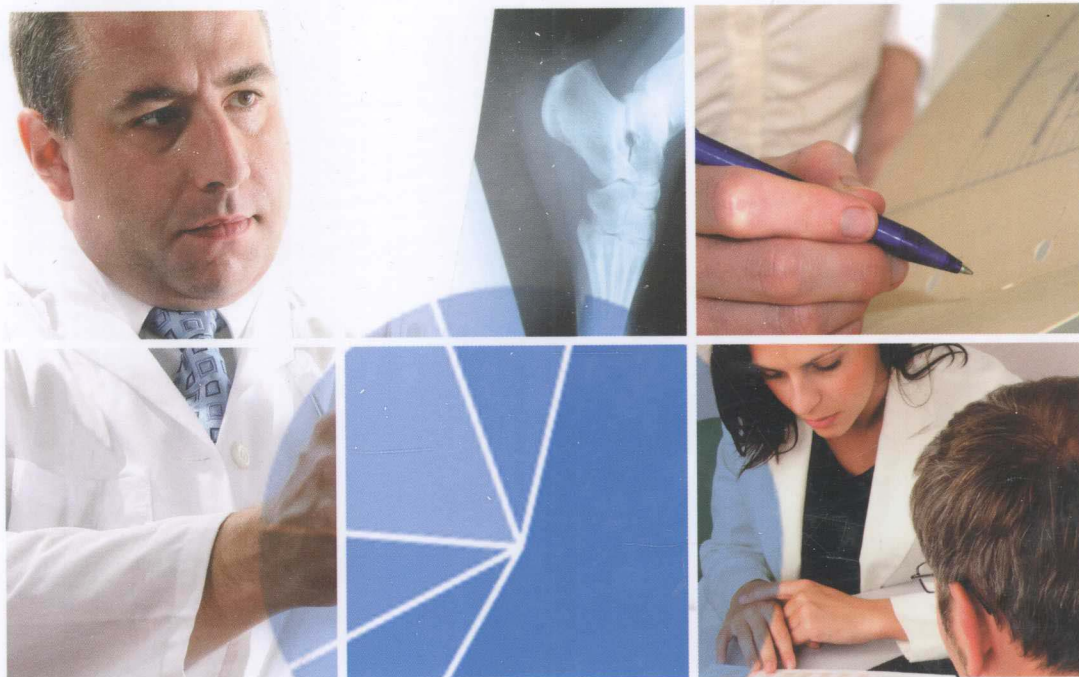
## (第2版)

Musculoskeletal Outcomes Measures and Instruments  
(2nd expanded edition)

编者 [美]Michael Suk [瑞士]Beate P Hanson

[美]Daniel C Norvell [美]David L Helfet

译者 周谋望 刘楠 杨延砚



山东科学技术出版社  
[www.lkj.com.cn](http://www.lkj.com.cn)



**AO Foundation**  
Publishing

---

# AO骨科量表评鉴

## (第2版)

---

Musculoskeletal Outcomes Measures and Instruments  
(2nd expanded edition)

编者 [美]Michael Suk [瑞士]Beate P Hanson

[美]Daniel C Norvell [美]David L Helfet

译者 周谋望 刘楠 杨延砚

## 图书在版编目 (CIP) 数据

AO 骨科量表评鉴:第2版/[美]苏克等编;周谋望,刘楠,杨延砚译. —济南:山东科学技术出版社,2013  
ISBN 978 - 7 - 5331 - 7068 - 4

I. ①A… II. ①苏… ②周… ③刘… ④杨…  
III. ①骨科学 IV. ①R68

中国版本图书馆 CIP 数据核字(2013)第 242070 号

Copyright © of the original English language edition 2009 by AO Publishing, Davos, Switzerland, Original title: “AO Handbook Musculoskeletal Outcomes – Measures and Instruments”, in 2 volumes, by Michael Suk/Beate P Hanson/Daniel C Norvell/David L Helfet

The simplified Chinese language edition © 2013 Shandong Science and Technology Press Co., Ltd.

图字 15 - 2011 - 150

## AO 骨科量表评鉴

(第2版)

编者 [美]Michael Suk  
[瑞士]Beate P Hanson  
[美]Daniel C Norvell  
[美]David L Helfet  
译者 周谋望 刘楠 杨延砚

---

出版者:山东科学技术出版社

地址:济南市玉函路16号  
邮编:250002 电话:(0531)82098088  
网址:www.lkj.com.cn  
电子邮件:sdkj@sdpress.com.cn

发行者:山东科学技术出版社

地址:济南市玉函路16号  
邮编:250002 电话:(0531)82098071

印刷者:山东新华印务有限责任公司

地址:济南市世纪大道2366号  
邮编:250104 电话:(0531)82079112

---

开本:889mm × 1194mm 1/16

印张:28.5

版次:2013年11月第1版第1次印刷

---

ISBN 978 - 7 - 5331 - 7068 - 4

定价:200.00 元

## 编者

Michael Suk, MD, JD, MPH  
Assistant Professor  
University of Florida  
Director, Orthopaedic Trauma Service  
Shands Jacksonville  
655 West 8th Street  
ACC Building, 2nd Floor/Ortho  
Jacksonville, FL 32209  
USA

Beate P Hanson, MD, MPH  
Assistant Professor  
Director of AOCID  
Clavadelerstrasse  
7270 Davos Platz  
Switzerland

Daniel C Norvell, PhD, MPH  
Spectrum Research, Inc.  
705 S 9th Street, Suite 203  
Tacoma, WA 98405  
USA

David L Helfet, MD, MBChB  
Professor of Orthopaedic Surgery  
Cornell University Medical College  
535 East 70th Street  
New York, NY 10021  
USA

## 译者

周谋望 刘楠 杨延砚

# 前言

许多临床医师和研究人员都希望制订一些通用的评定量表来表达和分享关于骨科疾病治疗的结果，本书就是这个理念的产物。本书不仅是骨科评定量表的汇总，更期望提供一个推进骨科学发展的坚实基础。

我们首先找出在文献报道中最常用的骨科评定量表（如：有长期使用历史的量表），然后按照解剖部位、类型（如：“患者自评”或“医务人员评定”）编排这些量表。单一评定报告或基于生理结果（如：关节活动度或放射影像学）的量表未纳入本书。

此外，本书收集了骨科文献中最常报道的健康相关生活质量评定量表，这些量表有：简明36项健康状况调查量表（SF-36）、Nottingham 健康分析量表（NHP）和肌肉骨骼功能评估量表（MFA）。

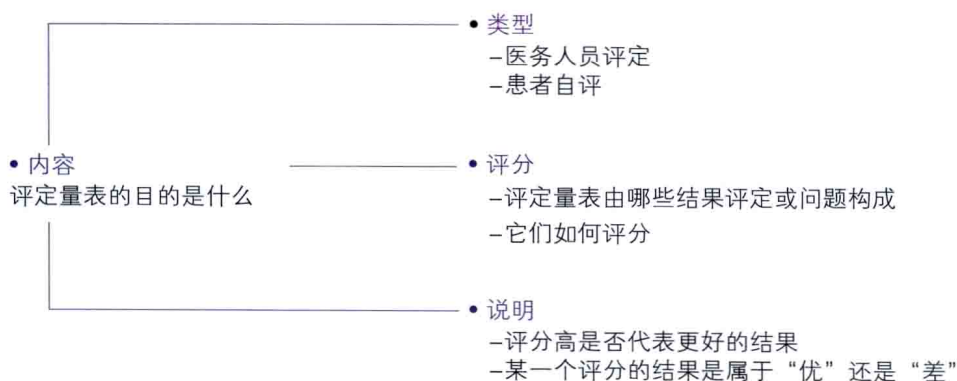
本书通过搜索MEDLINE<sup>®</sup>，尽可能多地找出包含骨科评定量表的参考文献，由于一些制订于数十年前的评定量表仍在使用（如：Merle D'Aubigne-Postel髋关节评分，1954），我们没有限制发表日期，以使我们的评定量表详尽和全面。

另外还通过查阅教科书、文献全文和参考书目，以及网络搜索引擎（如：Google<sup>™</sup>）进行查找，建立每个身体部位评定量表的条目。按照条目首先找到原文，然后找出文献中评估这一评定量表信度（可重复性和内部一致性）、效度（内容效度、标准效度和结构效度）以及敏感度的所有研究。尽管这是一个挑战，但是我们自信最终充分地找出了文献中最常用的骨科评定量表，包括患者自评（patient reported outcomes, PRO）和医务人员评定的量表（clinician based outcomes, CBO）。

查阅每个评定量表，分为四个主要部分进行总结：

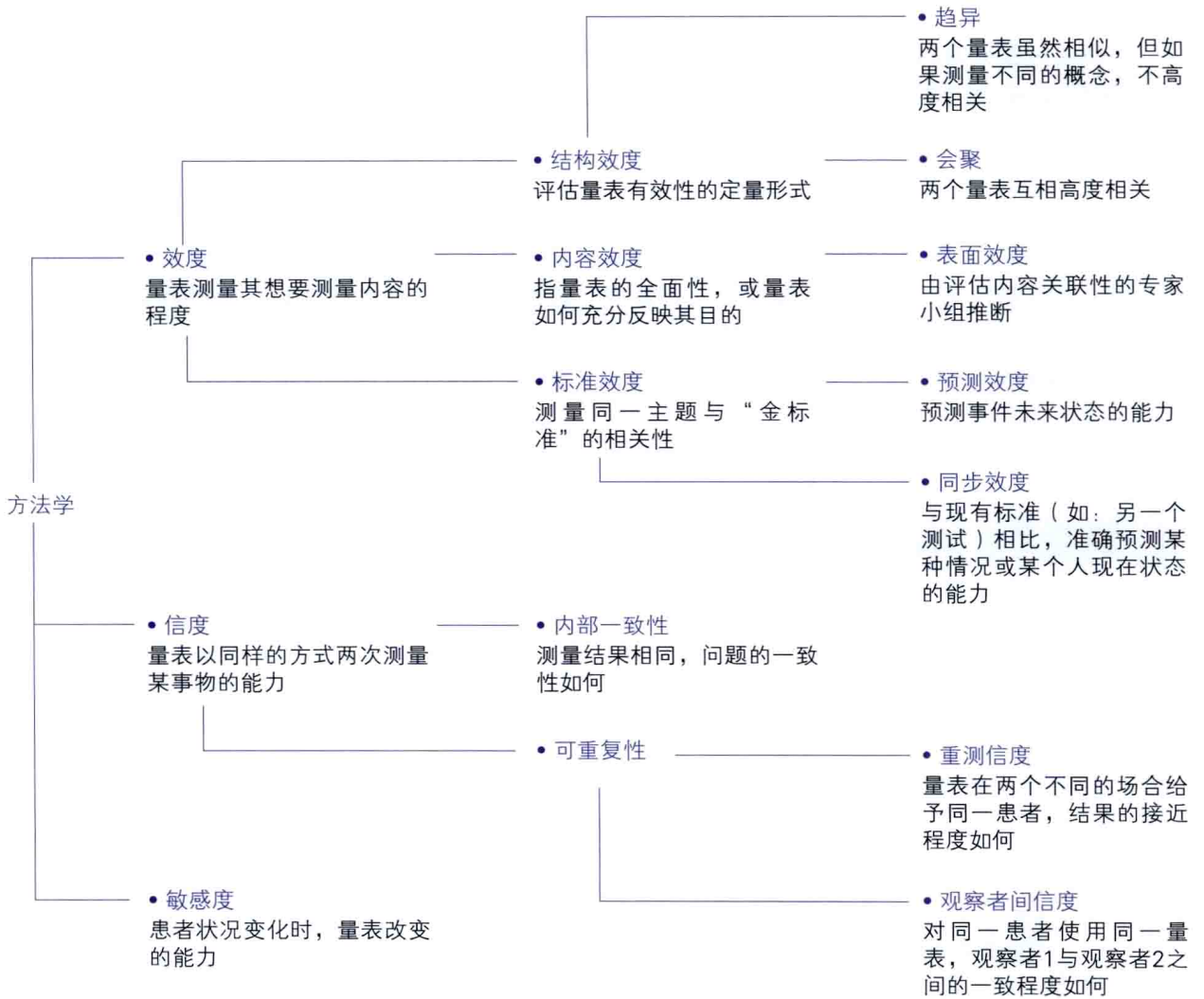
## 1 内容

- 类型
- 量表
- 说明



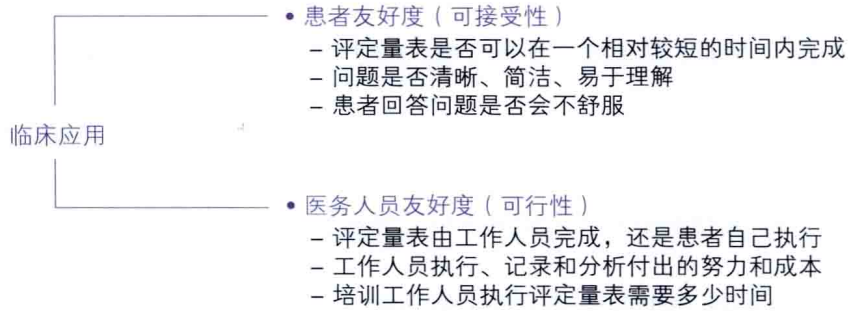
## 2 内容

- 效度
  - 结构效度
  - 内容效度
  - 标准效度
- 信度
  - 内部一致性
  - 可重复性
- 敏感度



### 3 临床应用

- 患者友好度
- 医务人员友好度



### 4 评分

完成搜索和总结之后, 我们对每个评定量表以总分10分进行评估, 并以一目了然的包含文字说明的图表表示; 同时列出了参考文献, 易于读者查找原文。

# 致谢

《AO骨科量表评鉴》得以出版，有赖于很多同事的无私奉献和辛勤工作，在这里表示衷心的感谢：

感谢Joseph R Dettori, PhD，是他制订了每个量表的评价和评分标准并对各个章节进行了统一编辑，这是本书很重要的工作之一。

感谢Eric Strauss, MD，作为医学生，其对评定量表进行第一轮搜集而付出的辛勤工作，使本书的出版成为可能。

感谢Nate Dettori, BS，是他对各种评定量表进行了细致耐心的搜集和整理，这一工作是本书得以出版的基础。

感谢共同编者Beate Hanson, MD和Daniel Norvell, PhD，他们将自己的信心和关注全部倾注于本书的整个出版过程。

感谢Urs Rüetschi，一位卓越的出版者，其持续的信心和对细节的关注，使本手册的出版成为现实。

感谢David Helfet, MD，导师、良师益友，他十分重视提高临床骨科医生的学术修养，这也是本书出版的灵感之一。

感谢AO基金会对这一重要项目的慷慨支持和资助。

Michael Suk



# 目 录

|                             |    |
|-----------------------------|----|
| <b>1 绪论</b> .....           | 1  |
| 1.1 现实的挑战 .....             | 1  |
| 1.2 手册 .....                | 1  |
| 1.3 读者对象 .....              | 1  |
| <b>2 选择适当评定的原因</b> .....    | 2  |
| 2.1 引言 .....                | 2  |
| 2.2 结果研究 .....              | 2  |
| 2.3 结果, 结果和更多的结果 .....      | 2  |
| 2.4 选择重要的结果 .....           | 3  |
| 2.5 结论 .....                | 3  |
| 2.6 参考文献 .....              | 3  |
| <b>3 高效的评定量表</b> .....      | 4  |
| 3.1 内容 .....                | 4  |
| 3.2 方法学 .....               | 4  |
| 3.3 临床应用 .....              | 9  |
| 3.4 评定量表的总体质量 .....         | 11 |
| 3.5 参考文献 .....              | 11 |
| <b>4 医务人员评定</b> .....       | 13 |
| 4.1 引言 .....                | 13 |
| 4.2 什么是医务人员评定 .....         | 13 |
| 4.3 医务人员评定是否比患者自评更加客观 ..... | 13 |
| 4.4 医务人员评定是否准确衡量患者的功能 ..... | 14 |
| 4.5 参考文献 .....              | 15 |
| <b>5 患者自评的目的</b> .....      | 16 |
| 5.1 什么是患者自评 .....           | 16 |
| 5.2 患者自评的出现 .....           | 16 |
| 5.3 量表类型 .....              | 16 |
| 5.4 参考文献 .....              | 17 |

|      |                  |     |
|------|------------------|-----|
| 6    | 建议               | 18  |
| 6.1  | 评估评定量表：选择正确的评定量表 | 18  |
| 6.2  | 制订评定量表：需要考虑的要点   | 19  |
| 6.3  | 参考文献             | 20  |
| 7    | 患者期望的结果          | 21  |
| 7.1  | 评定量表范例的发展        | 21  |
| 7.2  | 患者期望的作用          | 21  |
| 7.3  | 下一步的挑战           | 22  |
| 7.4  | 参考文献             | 22  |
| 8    | 鉴别和评价肌肉骨骼系统评定量表  | 24  |
| 8.1  | 搜索方法的解释          | 24  |
| 8.2  | 评定量表内容评价总结       | 24  |
| 8.3  | 评定量表方法学评价总结      | 25  |
| 8.4  | 评定量表临床应用评价总结     | 25  |
| 8.5  | 评定量表总分总结         | 25  |
| 8.6  | 我们对评分系统的调整       | 26  |
| 9    | 一般性评定量表和上肢评定量表   | 29  |
| 9.1  | 一般性评定量表          | 31  |
| 9.2  | 肩                | 41  |
| 9.3  | 肘                | 116 |
| 9.4  | 腕/手              | 146 |
| 10   | 下肢评定量表           | 208 |
| 10.1 | 骨盆               | 210 |
| 10.2 | 髌                | 214 |
| 10.3 | 膝                | 256 |
| 10.4 | 踝                | 330 |
| 10.5 | 足                | 379 |
| 10.6 | 跟骨               | 420 |
| A1   | 缩写词列表            | 431 |
| A2   | 词汇表及其定义          | 434 |
| A3   | 已评估评定量表列表        | 436 |

## 2 选择适当评定的原因

### 2.1 引言

在骨科疾病治疗的过程中，评定治疗结果时选择适当的评定量表是关键的一步。这可能是一项具有挑战性的任务，因为对于某个特定的期望结果（如：关节活动度），一种治疗方案或干预措施可能被认为好于另一种；但对于另一个结果（如：疼痛缓解）则不如另一种治疗方案或干预措施。

---

一项设计良好的研究虽明确描述某种治疗超过其他治疗，但提供的证据可能不足，或对“重要的”结果没有评定，则可能有误导。

因为作者没有考虑到对结果的选择，一些好的研究会留给我们很多问题。例如，一种治疗方式与其他方式相比，可能引起更少的近期并发症，但可能引起功能和生活质量下降。这些结果是否被评估了呢？对任何临床治疗或科研，评定治疗效果的关键是识别和评估重要的结果，那些被认为是重要的结果，可能存在于旁观者那里，应该更加重视这些方面。要理解选择适当结果的重要性，我们必须先了解结果研究的基本含义。

### 2.2 结果研究

本章的目的不是讨论结果研究的深度和广度，这超出了本手册的范围，但是理解结果研究的意义和重要性是选择“重要的”评定量表的条件。了解医疗卫生干预措施实际价值的需要，导致结果研究的再度兴起。在总结和评估使用评定量表之前，首先考虑结果研究的基本含义非常重要。下面的表格是对结果研究的定义的举例说明：

医疗卫生服务最终结果研究应该考虑患者的感受、喜好和评价<sup>[1]</sup>。

“结果研究是广泛的一系列研究活动，包括治疗的评估、治疗的衡量，以及何时采集数据、治疗管理<sup>[2]</sup>。”

疾病治疗结果的学术研究用于某种特定的疾病<sup>[3,4]</sup>。

结果研究涉及患者认为的其所接受的医疗护理的结果是什么。传统的研究涉及关节活动度、肌力、影像学等的标准评估。结果研究从不意味着代替评估骨科疾病治疗研究的通常方式，它是指增加另一个评估的维度<sup>[2]</sup>。

与决策者（从患者到整个社会）有关的描述、说明各种变化特别是干预措施对最后终点（从生存到医护满意度）影响的学科<sup>[5]</sup>。

一个与研究方式有关，并且使用各种互补的结果测量，致力于测量医疗卫生实际工作（有效性）的学科，范围从传统的临床测量（症状控制、骨性对线、关节活动度、疾病进展或生存）到健康相关生活质量和费用效益测量<sup>[6,7]</sup>。

### 2.3 结果，结果和更多的结果

一名外科医师在手术前面临着使用哪些装置或手术策略能更好地满足患者需要这一重大抉择。例如，对于肱骨近端骨折，有大约10种不同的内植物可供选择，其效果大致相同。如果不知道每种内植物的适用范围，就像看到许多关闭的门而不知道门后是什么；但如果每扇门有不同的外观，知道每种治疗方式的优势和缺陷，就使得外科医师更容易决定打开哪扇门。

选择评定量表更具挑战性，在骨科文献中有超过100种结果评定量表可供选择。仅就肩关节而言，可以从文献中发现大约30种评定量表，并且数量还在增加。如果不通过搜索文献就想知道每扇门之后有什么将是一个难以克服的障碍，那么通过本书的帮助，你可以拥有做出适当选择所需的知识。

## 2.4 选择重要的结果

本书不是告诉你在特定情况下需要哪种评定量表，最终你必须基于多种因素做出决定。选择一个重要的结果基于以下几个重要的步骤：

- 确定你所选择的结果用于临床或科研目的。
- 确定临床或科研目的。
- 确定哪些可采用的评定量表可以解决这一目的。
- 确定这些评定量表的质量。

遵循这些步骤，将会让你在选择可能“重要的”评定量表时，将其减少至一个可控的数量。

一旦找出一组评定量表，那么就可以比较它们的具体内容、适用群体和整体质量。后面的章节将陈述如何评估评定量表的整体质量。

## 2.5 结论

在为骨科临床和科研团队确定和推荐特定装置和干预措施的压力下，骨科结果研究具有独一无二的机会，来证明其验证这些方式在适当的患者群体、以适当的评定量表作为终点的真正临床获益的能力。尽管骨科文献已经记录其重要性，但是结果研究尚未证实其在临床和科研背景下的作用；亦尚未提供应用评定量表效度、信度和敏感度的完整评述。我们希望本手册对这一努力方向做出重大贡献。

## 2.6 参考文献

- [1] Clancy CM, Eisenberg JM (1998) Outcomes research: measuring the end results of health care. *Science*; 282: 245–246.
- [2] Simmons BP, Swiontkowski MF, Evans RW, et al (1999) Outcomes assessment in the information age: available instruments, data collection, and utilization of data. *Instr Course Lect*; 48:667–85, 1999.
- [3] Piccirillo JF (1994) Outcomes research and otolaryngology. *Otolaryngol Head Neck Surg*; 111:764–769.
- [4] Piccirillo JF, Stewart MG, Gliklich RE et al Outcomes research primer. *Otolaryngol Head Neck Surg*; 717:380–387.
- [5] The outcomes of Cancer Outcomes Research Focusing on the National Cancer Institute’s Quality-of-Care initiative. (2002) *Med Care*; 40:3–10.
- [6] Anderson C (1994) Measuring what works in health care. *Science*; 3:1080.
- [7] Epstein RS, Sherwood LM (1996) From outcomes research to disease management: a guide for the perplexed. *Ann Intern Med*; 124:832–837.

## 3 高效的评定量表

当选择和评估一个“高质量的结果评定量表”时，需要考虑许多因素，包括内容、方法和临床应用。对这些因素的说明是本章的主要着眼点，并且以本手册特有的清晰、前后一致和准确的方式呈现。本章讨论的概念是第9章和第10章中每个评定量表总体评估的基础。

在评估一个评定量表的总体质量时，应该考虑3个主要组成部分：内容、方法学和临床应用。

### 3.1 内容

简而言之，内容就是评定量表试图评估的东西。评定量表的问题是否与患者群体相关？评定量表的评分说明是否清楚？评定量表的内容可以被分为3个主要类目：类型、评分和说明。

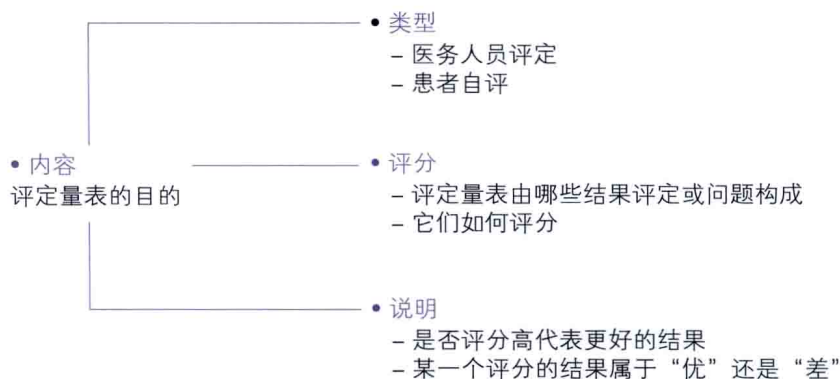


图1 评定量表内容的3个类目

### 3.2 方法学

量表的方法学可以被分为3个主要类目：效度、信度和敏感度。

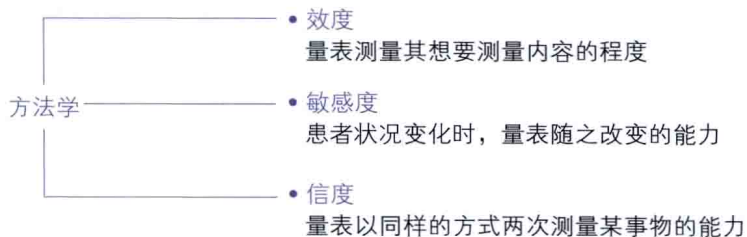


图2 概述评定量表方法学的3个主要类目

- 效度通常定义为量表测量其想要测量内容的程度。
  - 例如，在射箭运动中，效度表示射击的准确性——每次射击接近靶心的平均程度。
- 信度与评定量表的一致性有关。换言之，这是量表以同样的方式两次测量某事物的能力。
  - 以射箭为例，信度应该以集合连续射击如何相互接近表示，而无论它们是否击中目标。理想状态下，所有的射击命中靶心（有效并可信）。一个射手连续未命中靶心，表现为一组连续射

击，击中目标周围。这与评定量表可信、但是无效类似。

从事创伤骨科的住院医师必须首先学会成功地使骨折复位，之后不断学习使之熟练。这与评定量表的效度和信度类似。

- 敏感度指的是评定量表随着患者状况而改变的能力。如果患者在术后6个月重返工作时有一些功能障碍，而在12个月后进展为完全工作、没有功能障碍，那么测量患者这一功能状况的量表应该足够敏感，以反映这一改善情况。

当我们说一个评定量表有效时意味着什么？

因为效度不是一个固定的测量，应该在与一个具体目的或一组目的有关时<sup>[1]</sup>，以及在特定患者群体中使用时考虑评定量表的有效性。例如：

不能机械地认为评定膝关节骨性关节炎患者进行全膝关节置换术后功能障碍有效的评定量表用于股骨远端骨折患者有效。

此外，效度不能由一个概念概括，还可以细分为下述概念：内容效度、标准效度和结构效度。

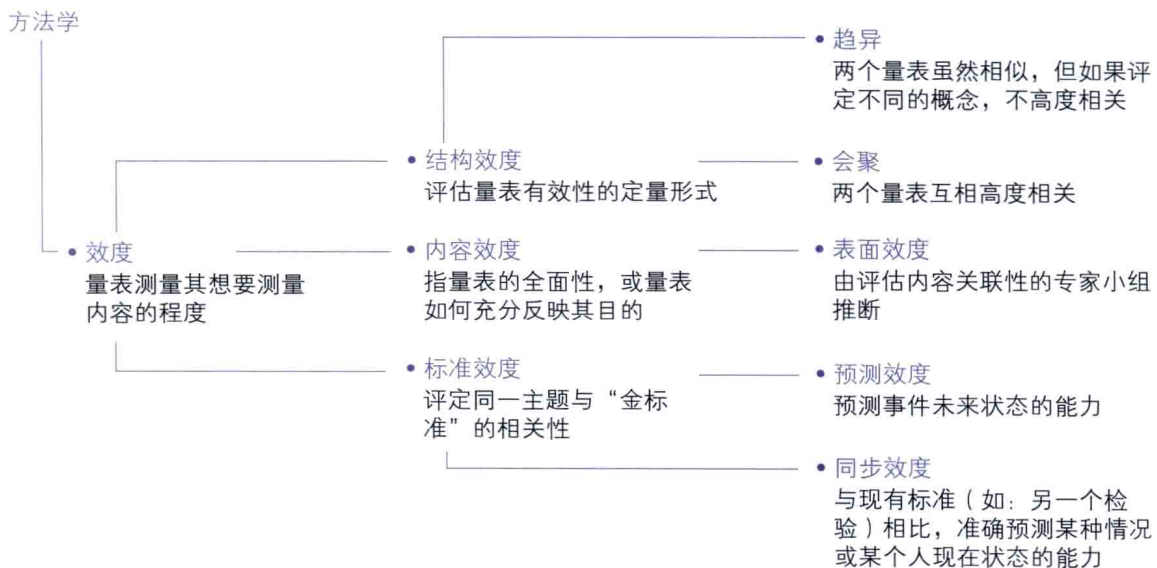


图3 效度可以进一步细分为3个主要概念

### 内容效度

“内容效度”指的是评定量表的全面性，或评定量表中包含的问题（即内容）如何充分地反映其总体目的。这被描述为常识和专业心理测量特性的混合<sup>[2]</sup>。

• 例如：

在一个设计用于年轻患者膝关节损伤的由患者自评的评定量表中，询问患者的日常生活活动，但是忽略了体育活动，该量表可能缺少充分的内容效度，导致结论不准确。

- 内容效度通常由专家制订。
  - 对于包括X线结果的医务人员评定，医师或医师小组通常是这一临床评估的专家。
  - 另一方面，对于致力于测量健康相关生活质量的患者自评，患者可能是专家。

尽管可以检验内容效度，但是很少正式进行。而评定量表的表面效度或临床可靠性通常由评估内容相关性的专家组进行推断<sup>[1]</sup>。以正式的方式证实所选的项目或问题代表所有的相关项目很难，也许是不可能的<sup>[3]</sup>。Guyatt及其同事通过下述语句辨别内容效度和表面效度：“表面效度检验评定量表是否看起来用于测量其想要测量的内容。内容效度检验评定量表中的项目或问题在目标领域中全面抽样的程度<sup>[4]</sup>。”它们共同用于评定量表中的问题是否解决了预期主题，以及是否充分涵盖了事物的全范围<sup>[1]</sup>。

表面效度和内容效度最好通过检查评定量表进行判定。判定内容效度的另一个方式是首先判定评定量表是如何制订的。具有相关临床和方法学经验的人有多少参与了内容的提出<sup>[5]</sup>？可能更为重要的是，有健康问题的患者在何种程度上参与了评定量表内容的创建并使其充实<sup>[6]</sup>？

### 标准效度

“标准效度”考察测量同一主题，评定量表是否与“金标准”高度相关\*<sup>[1]</sup>。

标准效度可以被进一步细分为同步效度和预测效度。

- 同步效度指的是评定量表准确预测某个人现在状态的能力。评定量表与某些已经存在的标准（如：另一个检验）进行比较。标准效度是效度中最基本的类型，并且是最常在传统医学研究中考察的类型<sup>[2]</sup>。此外，如果有一个外部的标准，应该用于验证评定量表。
  - 例如，一个新制订的用于年轻人创伤性股骨近端骨折的髋关节评分，可以与非关节炎的髋关节评分进行比较<sup>[7]</sup>。
- 预测效度指的是评定量表预测某个人未来状态的能力。
  - 例如，损伤严重程度评分中的某个特定临界值可能与最初创伤后30天内的死亡率高度相关。

### 结构效度

“结构效度”是评估评定量表效度时更加量化的形式。一个组成成分是一个项目或概念，如疼痛或功能障碍。经受更严重疼痛的患者预期将服用更多止痛药物，更严重功能障碍的患者预期将有更严重的关节活动度或肌力受限。通过比较一个组成成分与另一个变量，如疼痛和止痛药物使用，评估结构效度。

例如，如果你要在桡骨远端骨折患者身上检测上肢功能障碍（DASH）问卷的效度，你所期望的是DASH躯体功能评分与SF-36躯体功能评分最为相关，SF-36是躯体、社会和心理健康的—般性评定量表。相反，躯体功能的测量预期应与SF-36之中的情绪测量较少相关。

这突出了结构效度的2个方面：

- 会聚效度表示2个量表测量相似的概念时互相高度相关。

\* 对于疼痛、生活质量和抑郁的评定，难于证明其标准效度。因为假设存在完美的“金标准”，也很少与新的评定量表进行对照以检验其效度，效度检测更具挑战性。然而，有许多不同的和更为间接的方式被推荐用于判定评定量表的效度<sup>[8]</sup>。Fitzpatrick及其同事认为：“表面效度、内容效度和结构效度是在临床试验中使用患者自评测量时最相关的内容<sup>[1]</sup>。”

- 趋异效度表示2个量表虽然相似，但如果测量不同的概念时，不高度相关。

重要的是要认识到：

---

会聚效度和趋异效度共同起作用。如果可以证实同时存在会聚效度和趋异效度的证据，那么即存在结构效度的证据。但是仅有其中之一不足以确立结构效度。

证明会聚效度和趋异效度，需要在目标评定量表与组成成分相同或明显不同的其他评定量表之间形成假定关系。一旦建立假定关系，目标评定量表可以同时进行比较<sup>[9]</sup>。

例如，为了提供国际膝关节文献委员会（International Knee Documentation Committee, IKDC）膝关节主观评估表会聚效度和趋异效度的证据，Irrgang及其同事将这一量表与SF-36同时进行记录<sup>[9]</sup>。因为IKDC膝关节主观评估表是对躯体功能的测量，他们假定与同时测量的SF-36中的心理功能相比，其将会与SF-36中的躯体功能更加相关。

### 当我们说评定量表可信时意味着什么？

信度评估评定量表中的错误。在与射箭的类比中，信度由一组射击表示。因为信度可能与“可靠”混淆，其他人称之为“一致性”。如果评定量表反复产生错误的结果，这将是适合的<sup>[10]</sup>。

---

在评定干预措施的影响时，信度是评定量表非常重要的特性。如果结果不可信，在患者中观察到的变化可能不是一定归因于干预措施，而是评定量表自身的问题所致。

与效度一样，信度也不是固定的特性，而是有赖于研究的群体所处的环境<sup>[11]</sup>。可重复性和内部一致性是信度最好的代表：

### 可重复性

可重复性有2种形式：观察者间信度和重测信度。

- 观察者间可重复性。无论何时使用人作为评定过程的一部分，就不得不怀疑得到的结果是否可以重复，人们之间的不一致是众人皆知的。对同一患者使用同一量表，观察者1与观察者2之间的一致程度如何，是观察者间可重复性的本质。可重复性的这一形式适用于由医务人员判定的结果（在第4章医务人员评定中讨论）。
- 重测可重复性。当我们对同一患者在两个不同的场合下使用同一个评定量表时，估计重测可重复性。这一方法假定测量的组成成分在两个场合之间没有本质的变化。两次测量之间允许的时间间隔是关键。我们知道如果两次测量同一个事物，两次观察之间的相关性将会部分取决于两次观察的场合之间经过多少时间。时间差距越短，相关度越高；时间差距越长，相关度越低。

### 内部一致性

内部一致性评定量表中问题的同质性和一致性如何，以及测量同一个事物的程度如何。

---

多数评定量表使用几个问题或项目来评估一个单一的组成成分或维度（如：疼



痛、功能障碍)。这是因为几个相关的观察比一个通常能产生更可靠的判断<sup>[1]</sup>。为了实现这一点，所有的问题都需要类似，以评定单一属性的各个方面<sup>[11]</sup>。

最终结果是各个问题应该相互之间高度相关，并且与同一个量表中各个项目的总分高度相关。

- 例如，研究人员设计一个问卷，用于在交通事故造成严重的胫骨骨折之后重返工作的年轻患者，了解其完成工作的情况。如果关于需要双下肢肌力和反复使用双下肢活动的类似问题评分差异很大（有的高，有的低），由此表现为内部一致性差，你将不得不质疑评定量表的信度。另一方面，如果患者在有关下肢功能的工作活动相关问题中均一致的评分低，那么就这一组成成分而言，该评定量表可能是可信的。
- 内部一致性帮助解决可重复性固有的局限。通常不能在同一情况下对患者2次使用同一评定量表。患者的回答有可能不同，可能因为在其所处的环境中发生的事情，或其自己在2次评估之间改变回答；患者也有可能从之前的评估中记住他们给出的答案。

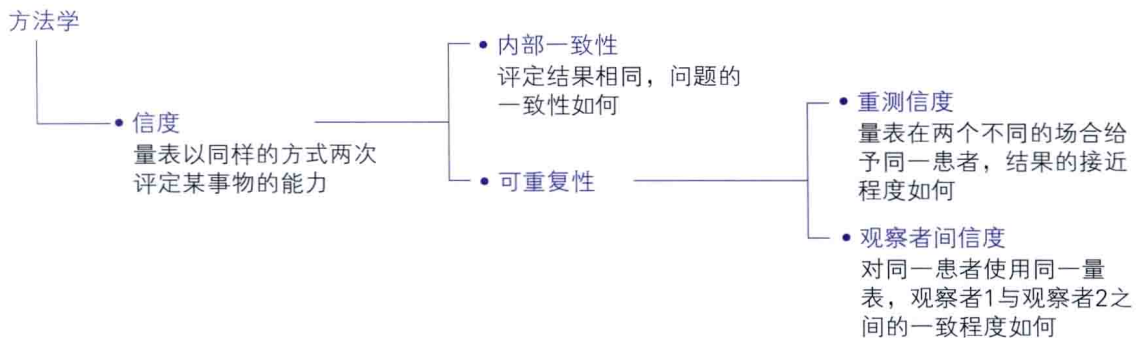


图4 信度可以进一步细分为2个主要概念

### 当我们说评定量表敏感时意味着什么？

敏感度，也称为“对变化的敏感度”，评定量表如何很好地发现作为干预措施结果的变化<sup>[2]</sup>。对于评定量表，有可能是有效、可信，但是不敏感的。当应用这一量表评估患者的进展或一种特殊治疗的效果时，将会出现问题。

一个不能在患者病情恶化或改善时反映变化的评定量表，即使是有效和可信的，也几乎没有临床或科研价值。

### 什么是最小临床重要差异？

有证据<sup>[12-16]</sup>表明，源自经过验证的评定量表中的有统计学意义的评分变化，不一定意味着这一变化在临床上是重要的。统计学意义基于许多事物，最重要的是样本大小。因此，如果临床研究中的样本量足够大，评估膝关节置换手术后患者中的少量变化，可能具有统计学意义；否则这一变化可能不能表示患者个体足够大的临床获益。从临床角度，在做出循证治疗决策时，我们需要评估治疗结果是否是值得的和重要的<sup>[17]</sup>。出于研究目的，为了估计将来研究所需的样本量大小，我们需要知道对临床重要的评分变化的大小。

作为结果，越来越多的人认识到敏感度应该包括“测量一个有意义的或重要的临床状态变化”的能力<sup>[18]</sup>。提倡“最小临床重要差异（minimal clinically important difference, MCID）”的概念，用做