

· 讲座 ·

Rasch 分析在康复医学科研中的应用(一)

卢成皆 许涛 黄铭聪

编者按:近年来,Rasch 分析在医学上的应用愈来愈广泛。为使康复医学工作者对这个分析方法有基本认识,我们特约请澳大利亚悉尼大学的医学统计学专家卢成皆教授等人撰文介绍它的概念及如何应用,并且在讨论及说明的过程中尽量避免使用复杂的数学公式,以方便读者了解。

前 言

康复医学工作以提高患者的整体功能水平为最终目标,但对于如何评价患者的功能能力,还有很多问题在困扰着我们。譬如,在不同的康复机构使用不同的评定量表,缺乏统一的标准去比较患者的功能疗效;同一个患者在疾病的不同阶段,可能用不同的量表在评定,因此对一个患者某种功能能力改善的追踪观察也受到影响;还有,在同一个机构中,对于不同能力水平的患者,均使用同一种量表去评定,可能难以精确地获取患者的信息,等等。如何使我们对功能能力的评定走向标准化,使我们的评定工作既准确有效,应用自如,又符合一个通用的标准而有助于结果的比较研究呢? Rasch 分析在这些方面的应用颇有价值。

计量模型原理

在讨论 Rasch 分析前,让我们先介绍“计量模型”(measurement model)的一些原理,这样可以帮助我们较容易了解 Rasch 分析。举一个例(表 1)来说,假设我们要使用 5 个项目(item)来评估 7 名患者的“功能状态”(functional status),若患者能够做到任何 1 个项目(如:穿衣服、煮食等),其该项目的得分便是 1,若做不到便得 0 分;每个项目都如是。因此,每个患者最高可得 5 分,最低则为 0 分;分数愈高代表功能状态愈高。从表 1 中显然可以看到 7 名患者功能状态的高低,同时,每个项目的难易程度(如:“煮食”这个项目便比“穿衣服”来得困难)也可从表 1 观察得到。愈少患者能够做到的项目,其困难程度便会比很多患者都能够做到的项目来得高。若以表 1 的资料为例,7 号患者的功能状态最高,所有项目都能够做到,得到 5 分;而 1 号患者的功能状态则最低,只能做到项目一,故只得 1 分。另一方面,项目一最容易,所有患者都能做得到,总分为 7 分;而项目五则最难,只有 1 名患者(7 号)做得到,总分只有 1 分。

表 1 内的“1”及“0”称为“观察值”(observations),亦可称为“响应值”(responses)。因为表 1 内共有 7 名患者及 5 个项目,所以这一个表共有 35 个响应值。每个响应值为“1”(做得到)或“0”(做不到),主要受到两个因素(即“该名患者的功能状态的高低”及“该项目的难易程度”)的影响。所谓“计量模型”,即是以“响应值”、“患者功能状态的高低”及“项目难易程度”(item difficulty)三者所建立的数学模型。更精确地说,是“响应

值”如何受到患者功能状态的高低及项目难易程度的影响。若以数学模型来表达,则为:

$$y_{ij} = f(B_i, D_j) \quad (一)$$

在公式(一)中, y_{ij} 为第 i 名患者第 j 个项目的响应值; B_i 为第 i 名患者的功能状态; D_j 为第 j 个项目的难易程度; $f(B_i, D_j)$ 念作“function of B_i and D_j ”,用以表示是一种数学算式,至于是那一种数学算式则没有明确说明。

表 1 各患者不同项目的得分情况

患者	项目					总和
	一	二	三	四	五	
1	1	0	0	0	0	1
2	1	1	0	0	0	2
3	1	0	1	0	0	2
4	1	1	0	1	0	3
5	1	1	1	0	0	3
6	1	1	1	1	0	4
7	1	1	1	1	1	5
总和	7	5	4	3	1	20

当然,从表 1 的资料,我们也希望可以估计(estimate)一下各患者的功能状态及各项项目的难易程度。公式(一)内的 $f(B_i, D_j)$ 没有明言是什么样的数学算式,那么使用那一个算式会较为合理呢? 有专家提出使用 B_i 与 D_j 的差异(或差距,difference),即以 $(B_i - D_j)$ 为基础的算式。他们的论点是,当一个功能状态很高的患者去做一件很简单的项目时,我们会期望(expect)他能够做到;当然,期望他做到并不表明他一定能够做到,因为除了患者的功能状态及项目的困难度外,还有其它的因素(如:患者当时的心情、一些其它客观的环境因素、随机误差等)会影响到该名患者当时是否能够真的做到该项目。所以,当患者的功能状态强而项目不太困难(即 $B_i > D_j$ 或 $\{B_i - D_j\} > 0$) 时,我们也只能说他能够做到该项目的机会较大(即 $\text{prob}\{y_{ij} = 1\} > 0.5$)。反过来说,当一个功能状态很低的患者去做一件很困难的项目时,我们会觉得他能够成功做到的机会很小。用数学公式来说,即当 $B_i < D_j$ 或 $\{B_i - D_j\} < 0$ 时,则 $\text{prob}\{y_{ij} = 1\} < 0.5$ 。综合上述,我们便可把公式(一)进一步发展为

$$\text{若 } \{B_i - D_j\} > 0, \text{ prob } \{y_{ij} = 1\} > 0.5 \quad (二甲)$$

$$\text{若 } \{B_i - D_j\} < 0, \text{ prob } \{y_{ij} = 1\} < 0.5 \quad (二乙)$$

$$\text{若 } B_i - D_j = 0, \text{ prob } \{y_{ij} = 1\} = 0.5 \quad (二丙)$$

在 1960 年,丹麦数学家 Rasch 更提出了使用 $(B_i - D_j)$ 的“指数”(exponent)作为公式(一)内未明言的数学公式。更明确地说,Rasch 提出了以下数学模型:

$$\text{prob } \{y_{ij}\} = \frac{\exp [y_{ij}(B_i - D_j)]}{1 + \exp (B_i - D_j)} \quad (三)$$

公式(三)内之 y_{ij} 不是“1”就是“0”,当 $\{y_{ij} = 1\}$ (即 i 患者能够做到 j 项目)时

作者单位:悉尼,澳大利亚悉尼大学国际卫生学院(卢成皆);同济医院康复医学科(许涛);香港中文大学流行病及临床试验中心(黄铭聪)

$$\text{prob}\{y_{ij} = 1\} = \frac{\exp(B_i - D_j)}{1 + \exp(B_i - D_j)} \quad (\text{四甲})$$

而当 $\{y_{ij} = 0\}$ (即 i 患者做不到 j 项目) 时,

$$\text{prob}\{y_{ij} = 0\} = \frac{1}{1 + \exp(B_i - D_j)} \quad (\text{四乙})$$

公式(三)便是所谓的 Rasch 模型。站在数学的立场,这个模型相等于(is equivalent to)“单参数”(1-parameter) logistic 模型。它是一个 logistic 模型,因为它符合了 $\text{probability} = \frac{\exp(x)}{1 + \exp(x)}$ 的模式。但读者可能会问,公式(三)中看起来有两个参数,即 B_i 及 D_j ,为什么它是一个单参数模型呢?原因有二。从应用的层面来说,使用 Rasch 分析的主要原因在了解量表中各项目是否适合,同时也希望在分析各项目是否过于容易或困难的过程中,不会受到所选取的样本中各患者能力高低的影响,因此,整个模型的重点在项目的难易度(即 D_j)。其实公式(三)若改写成公式(三乙),读者可能更容易理解它只有一个参数。

$$\text{prob}\{y_{ij} = 1\} = \frac{\exp[(b - D_j)]}{1 + \exp(b - D_j)} \quad (\text{三乙})$$

即把 B_i 改为 b ,而在阐释的时候,我们则说“对于所有功能状态为 b 的患者,他们能够做到或做不到第 j 个项目的几率是依循公式(三乙)的”。从以上的说明,我们已经可以清楚看到所谓单参数是指公式中的 D_j 。此外,从数学的层面来说,我们有兴趣的并非是单独的 B_i 或 D_j ,而是 B_i 与 D_j 的差异(即 $B_i - D_j$),这一点从公式(二甲)至(二丙)已可见端倪。我们可以看到三个公式都是以 $(B_i - D_j)$ 大于、小于或等于零作为考量,即模型是把 $(B_i - D_j)$ 作为一个参数来建立的。

那么 D_j 表示什么意义呢?简单来说,它是在座标图的 X 轴上的一点,而这一点刚好对应于 Y 轴的 0.5;亦即是说,若项目的困难度为 D_j 的话,则平均来说,只有一半的患者能做到这样的项目;从另外一个角度来看,亦可以说在所有患者中随机抽取一人,那么他做到这样困难度的项目的几率为 0.5。再引伸一点,若 D_j 愈大,便代表该项目愈困难。因此,在进行 Rasch 分析时,我们通过比较各 D_j 的大小,便可知道那一个项目较易,那一个项目较难了。若在分析时把资料先进行转换(transform),使平均值为 0,标准差为 1 的话,那么, D_j 一般都会在约为 -3 至 +3 之间。很多软件都会自行进行转换这个过程,使用者不必费心于如何转换。

早在 1920 年,数学家便已经使用 logistic 函数了,那么为什么公式(三)会被称为 Rasch 模型呢?因为在 Rasch 以前,研究者并未将此公式应用于“计量模型”这个领域内;更重要的是,Rasch 首先发现了这条公式应用在“计量模型”时的一些特质与重要性(psychometric properties and significance),下面我们逐一说明。

从公式(一)到公式(四), D_j 代表了第 j 个项目的难易度,那么利用如表 1 的资料,如何去估计每个项目的难易度呢?若使用传统统计方法,第五个项目只有 1 名患者做到,故其困难度为 $(1 - 1/7) = 6/7 = 0.86$ (或 86%)。其余四个项目的困难度分别为:项目一,0%;项目二,29%;项目三,43%;项目四,57%。困难度的百分比愈高,代表该项目愈困难。这种传统统计方法的计算步骤很简单,却有一个颇为严重的缺失,在于表 1 是由一个有 7 名患者的样本所产生的资料,利用这一个资料所计算出

来的只是难易度的估值。很明显,若找另外 7 名患者,不见得他们全都可以做到项目一。再想深一层,当某项目困难度的估值为 100% (即所有患者都做不到)时,究竟是该项目真的太难呢?抑或只是这个样本内的患者的能力太差?上述的传统计算方式,在估计项目的困难度时,完全没有把患者的平均能力及标准差放入计算公式内,亦即没有考虑患者的平均能力及人与人之间的差异。故这样的计算方式似乎有欠完整。若使用 Rasch 分析法,以项目五为例,项目困难度的估计值为:

$$M + (1 + \sigma^2/2.89)^{1/2} \ln[1 - (1/7)/(1/7)] \quad (\text{五})$$

公式(五)内的 M 及 σ 分别为患者能力的平均值及标准差。由此可见,Rasch 分析考虑到了患者的平均能力及人与人之间的差异,因此较为合理。(注: M 及 σ 应如何计算,在此未予交待。因为这篇论文的对象是康复医学界专业人员,把数学公式弄得太深、太难可能更增加读者的疑惑。有兴趣进一步了解者,可与作者直接联系。)

以上有关项目困难度的论点及计算方式也可引伸至患者的功能状态高低。如第一名患者在 5 个项目中只做到 1 项,故使用传统统计方法,他的“功能状态”的估值为 $(1/5) = 0.2$ (即 20%)。我们试想,去评估一个患者的功能状态可以有很多项目,现在使用的这 5 个项目是否实在太难,致使这名患者只能做到其中 1 个项目呢?若选用另外 5 个项目,他的表现是否会较佳呢?抑或他的功能状态的确偏低,用任何 5 个项目他都只能做到 1 个?单利用以上的传统计算方式,我们无法得知,因为计算程序是估值 = $1/5$,其中的“1”是“该名患者做到了多少个项目”,“5”则是项目总数,两者都与现在使用的这 5 个项目是否实在太难无什么关系。但若使用 Rasch 分析法,第一名患者的功能状态的估值为

$$H + (1 + \sigma^2/2.89)^{1/2} \ln[(1 - 1/5)/(1/5)] \quad (\text{六})$$

公式(六)中的 H 及 σ 分别为各项目的平均困难度及标准差。在评估患者的功能状态的过程中,考虑并校正项目的难度,Rasch 分析自然较传统分析法更具说服力。

再举一例子说明传统方法的缺失:假设另有一量表,共有 40 个项目,每一项目都是做到的得 1 分,做不到得 0 分。现有两组患者,每组各 10 人,他们的总得分及项目一的得分均列于表 1。举例来说:甲组的第一名患者在 40 个项目中只做到 8 个,得 8 分,而他做不到第一个项目。若使用传统方法的话,则第一个项目的“项目困难度”可从做得到这个项目的人数百分比来估计。甲组的 10 名患者中,只有 2 人做到了项目一,故这个项目的“项目困难度”在甲组来说为 0.2;同样道理,项目一的“项目困难度”在乙组来说则为 0.8。至于“项目鉴别度”(item discrimination),传统方式是以计算该项目的得分与整个量表的总分之相关来估计的。以表 1 的资料为例,项目得分与量表总分的相关(正确来说应是 point-biserial correlation,数学上等于 Pearson correlation)分别为甲组 $r = 0.62$,乙组 $r = 0.34$ 。从这个例子我们可看到传统方法的缺失:项目困难度与项目鉴别度不应该受到测试样本的影响,但甲、乙两组的项目困难度与鉴别度系数都相差很远,很明显受到测试样本(即甲组为能力较低之一组,而乙组为能力较强之一组)的影响。

此外,若某项目困难度的估值(以传统方法计算)为 100% 或 0% 时,则这个估值的标准误差(standard error)为 0;当难度的估值为 50% 时标准误差会最大。简单来说,是愈极端(如 100%

或 0%) 的标准误差愈低,愈靠近中央的(如 50%)标准误差便愈大。这个现象,似乎在说愈多患者做到或做不到的项目愈好(因为标准误差会低,标准误差当然愈低愈好),而有一半人做到的项目最不好。但我们试想,若全部患者都做到或做不到,这个项目不是就没有“鉴别能力”(discriminating power;即把患者功能状态的高低分别出来)了吗?这样的项目怎会适合呢?相反,有些患者做到、有些患者做不到的项目才会有“鉴别能力”,才会较好。这又反映出传统方法的另一种缺失。使用 Rasch 模型则不一样,愈多患者做到或做不到的项目,它们的标准误差会较大;相反,当只有一半患者做到时,该项目的标准误差会最小。这个现象完全反映了项目的“鉴别能力”。因此 Rasch 模型的确比较合理。同样道理,使用传统方法,当一名患者大部分的项目都做到或做不到时,其标准误差会比只有一半项目做到时为低;但若使用 Rasch 模型,结果会反过来,标准误差在只做到一半项目时为最低。

Rasch 模型还有一个很合理的现象。假设患者的功能状态(即公式(三)中的 B_i)为 0~4 分,而项目的困难度(即公式(三)中的 D_j)也是 0~4 分,表 2 显示了 $(B_i - D_j)$ 及 $\text{prob}\{y_{ij} = 1\}$ [即第 i 名患者能够做到 j 项目的几率,计算程序请参看公式(四甲)]各数值。表 2 内最值得注意的是最右面一行,即

$$\text{prob}\{y_{ij} = 1\} \times [1 - \text{prob}\{y_{ij} = 1\}] \quad (七)$$

这个数值称为“反映信息”(information in a response),数值愈大代表它愈能反映我们希望得到的信息。从表中可见,“反映信息”最强是当 $(B_i - D_j) = 0$ 时。它的意思是说,当我们设计量表时,能设计一个困难度刚好配合患者的能力(即功能状态高低)的项目是最好的。这个理念很多研究者都知道,只是 Rasch 分析能把它量化,用一个数学模型来表达其意义而已。

表 2 患者功能状态、项目困难度及相应的反映信息值

患者功能 状态 B_i	项目困难度 D_j	$(B_i - D_j)$	$\text{prob}\{y_{ij} = 1\}$ 患者做到 项目几率	$\text{prob}\{y_{ij} = 1\} \times$ $[1 - \text{prob}\{y_{ij} = 1\}]$
4	0	4	0.98	0.02
3	0	3	0.95	0.05
2	0	2	0.88	0.11
1	0	1	0.73	0.20
4	1	3	0.95	0.05
3	1	2	0.88	0.11
2	1	1	0.73	0.20
1	1	0	0.5	0.25
...	...	...	...	...
0	0	0	0.5	0.25
...	...	...	...	...
1	2	-1	0.27	0.20
1	3	-2	0.12	0.11
1	4	-3	0.05	0.05
0	1	-1	0.27	0.20
0	2	-2	0.12	0.11
0	3	-3	0.05	0.05
0	4	-4	0.02	0.02

站在计量模型分析的立场,Rasch 模型可说是一种“项目反映理论”(item response theory, IRT)模型,项目反映理论与泛化理论(generalizability theory)及经典的心理测量理论是三种基本的测量模型。项目反映理论(又称为潜在特性理论,latent trait theory)关注的是测量的对象(如患者)与测试项目的难度之间的关系,它包含了以下两个重点。

1. 受试者对量表内任何一个项目的表现如何是可以利用某些“因子”(factors)预估(predict)的,而这些因子便称为潜在特性(latent trait)。举一个简单的例子来说,假设现有一量表,内有 10 个项目(如:第一个项目与穿衣服有关;第二个项目与煮食有关等),研究者希望透过这 10 个项目,来评估受试者的功能状态,那么,功能状态便是一个潜在特性。功能状态的高低便会影响到受试者能否自己穿衣服、能否自己煮食等等。

2. 受试者对量表内任何一个项目的表现与潜在特性之间的关系可以利用项目特性曲线来描述。例如,图 1 是一个典型的项目特性曲线(item characteristic curve, ICC),很多项目特性曲线都是呈 S 型的,它的 Y 轴显示了“能够做到某个项目的几率或百分比”,X 轴则是潜在特性的得分(latent trait score)。图 1 代表了潜在特性愈高,能够做到项目的几率也愈高。读者这时可能会问:如果潜在特性是“老人痴呆”,量表中的项目则是数学及方向感等题目,那么潜在特性愈高,答对题目几率便愈低,是否项目特性曲线便会呈反过来的 S 型呢?理论上可以是这样,但数学模型此时便会更改相关的参数(parameter),使潜在特性愈高代表老人痴呆的程度愈低,那么答对题目几率便愈高,而项目特性曲线的形状便仍然保持如图 1 的形状了。

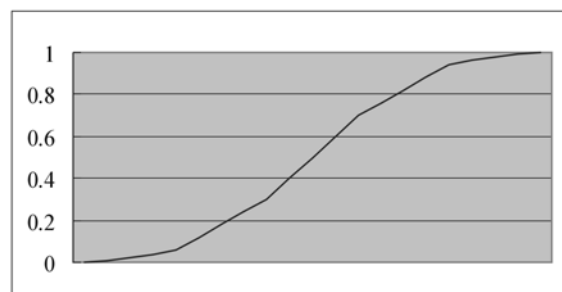


图 1 项目特性曲线图

那么,为什么 Rasch 模型是项目反映理论中的一个数学模型呢?显而易见:公式(三)中在等号左边的几率相等于图 1 的 Y 轴,因为这个几率反映了患者在第 j 个项目的表现;而在公式中等号右边的部分,因为包含了患者的功能状态与项目的难易程度两者之关系,而这关系构成了一种潜在特性,其意义自然与图 1 的 X 轴相若。(待续)

(收稿日期:2002-07-19)

(本文编辑:郭铁成)