

自评一般健康的信度和效度分析

社会
2014 · 6
CJS
第 34 卷

齐亚强

摘要:本文利用 2008 年“中国流动与健康调查”(IMHC)数据,分析了自评一般健康指标的信度和效度。研究发现,自评一般健康具有较好的信度,被访者先后两次回答的结果高度一致,两次回答结果的微小变动表现为随机性的波动,而非系统性偏差。该指标在一定程度上会受到调查中题目次序效应的影响。关于自评一般健康指标效度的分析发现,该指标能够有效反映被访者自我感知的各种健康状态和个体既有的关于自身健康的知识,但不能很好地反映个体无法感知的机体变化等健康问题。自评一般健康存在较为复杂的回答偏误问题,受不同年龄、社会经济地位群体关于健康的评价标准、期望与认知差异的影响,其回答结果在不同人群中的可比性值得商榷。

关键词:自评一般健康 信度 效度 回答偏误

Reliability and Validity of Self-Rated General Health

Qi Yaqiang

Abstract: Using data from the 2008 Survey of Internal Migration and Health in China, this study examines the reliability and validity of self-rated general health for the Chinese population. Results show that self-rated general health is a highly reliable measure of individual health. Two repeated measures of self-rated general health in the survey are quite consistent and the difference between the two answers reflects random variations rather than any systematic biases. Nonetheless, there is also some evidence that self-rated general health is likely to be affected by question orders in a survey.

In addition, this study examines the validity and potential reporting bias of

* 作者:齐亚强 中国人民大学社会学系(Author: QI Yaqiang, Department of Sociology, Renmin University of China) E-mail: qiyang@ruc.edu.cn

** 本研究受中国人民大学科学研究基金(中央高校基本科研业务费专项资金资助)项目(13XNJ034)资助。[This research is supported by the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China (13XNJ034).]

self-rated general health by fitting Hopit models. Results show that self-rated general health is a valid summary measure of individual's self-perceived and known health conditions, although it does not reflect bodily functional changes that can hardly be perceived. The response of self-rated general health is strongly correlated with respondent's chronic medical conditions, the occurrence of acute illness, self-perceived pains/discomfort, insomnia symptoms and depression; however, it is only weakly correlated with objective biometrics such as blood pressure and lung capacity.

Finally, it is worth noting that there exists complicated reporting heterogeneity in self-rated general health among different social groups. Due to differences in rating standard, expectation and cognitive capability regarding health, different social groups respond to the question of self-rated general health differently. After controlling for all the specific measures of individual health in the data, older respondents tend to underestimate their true health status, while those better educated and respondents with higher family incomes tend to overestimate their true health conditions. The existence of reporting heterogeneity is likely to hamper the cross-population comparability of self-rated general health.

Keywords: self-rated general health, reliability, validity, reporting heterogeneity

自 20 世纪 80 年代以来,健康不平等及其社会影响因素日益成为社会科学研究热点之一。¹受研究成本、技术和手段的限制(如问卷调查方法),自评一般健康(self-rated general health)指标成为社会研究者分析健康问题的首选工具。与其他健康指标相比,自评一般健康通过询问被访者一个单一的问题来测量其健康状况,²操作简便,成本很低,目前已经在许多大型社会调查中得到普遍应用,并通过对该指标的分析也产生了大量的研究成果。不过,自评一般健康指标的信度和效度问题也一直是学术界关注和争论的焦点(Jylhä, 2009; Mathers, 2003; Sadana, *et al.*, 2002; Thomas and Frankenberg, 2002)。这样一个简单问题能够在多大程度上反映被访者的真实健康状况?其

1. 相关评述可参见:王曲、刘民权(2005)。

2. 自评一般健康是指在问卷调查中询问被访者以下问题:“总的来说,您认为您的健康状况怎么样?”回答选项一般由 3-5 个定序类别组成,不同调查的具体措辞可能存在细微差别。

测量的稳定性如何?对于这些问题,学术界还未达成共识,仍有待于在相应领域做进一步的深入探讨。

到目前为止,国际学术界关于自评一般健康指标的信度和效度的研究成果已相当丰富,但在中国相关问题所受到的关注仍非常有限。考虑到大量国际研究主要使用西方发达国家的数据资料,其研究结论的普适性仍有待检验。本文旨在利用中国的调查数据,就该问题提供新的证据资料,以期为推动健康问题的社会理论发展贡献新的知识。具体而言,本文的主要研究问题包括:第一,自评一般健康的测量是否具有稳定性?换句话说,如果针对同一调查对象重复进行测量,所得到的结果是否一致?如果不一致,是哪些因素导致了这种变化?第二,自评一般健康是否能够反映被访者的真实健康状况?它与其他健康指标之间的关系如何?使用自评一般健康指标来考察不同社会群体的健康差异时是否会导致系统性偏误?本文接下来将首先简单回顾和梳理针对自评一般健康信度和效度问题的既有研究成果与相关争论,再介绍本研究所使用的主要数据以及分析方法。在第三部分,笔者将利用较新的中国调查数据分析自评一般健康的信度、效度和可能的回答偏误问题。最后,对本研究的主要发现进行简单的总结和讨论。

一、既有研究评述

(一) 健康的概念及其测量

按照世界卫生组织(WHO,1946)的定义,健康是指“一种身体、精神和社会的完全安宁状态,不仅仅是没有疾病或体质虚弱”。由此可见,健康是一个复杂、多维的概念,在实际研究中很难通过某项单一的指标来反映个体健康的所有维度。穆雷和陈(Murray and Chen,1992)认为,个体健康状况实际上包括三个不同的组成部分:能够自我感知但通过外部手段难以观测的疼痛和不适感、既能自我感知又能观测到的机体变化,以及无法自我感知但却可以测量到的机体变化。其中,前两个部分的信息可以通过个体的自我感知获得,后两个部分的信息可以通过医学检测手段获知。因此,要想全面了解个体的健康状况,就有必要同时收集自评健康指标和客观体测结果。当然,自我感知的健康状况与医学观察结果并不必然一致,这是因为它们所反映的健康维度本身就存在差异。

自我感知的健康状况既取决于机体病理负担,还受个体自身的社会和文化背景的影响。健康自评取决于被访者与参照群体或者自认的理想健康状态进行比较的结果,这两者都可能因为在社会文化方面的差异而不同。尽管如此,自评健康指标在健康研究中仍具有非常重要的价值,个人对疾病负担的主观评价与认知对个体生活质量的影响甚至比一些客观的健康指标更为关键(Murray and Chen,1992),因此,关于疾病的自我感知情况本身就是值得研究的社会现象。

(二) 关于自评一般健康指标的争论

在现实应用中,由于成本和可操作性等方面的限制,大型的社会调查一般只收集被访者自报或自评的健康状况信息,其中的自评一般健康已经成为社会调查研究中使用最广泛的一项健康指标。

自评一般健康能否有效测量个体的真实健康状况,尤其是在不同群体之间的自评一般健康是否具有可比性,对于这些问题的回答,以往的研究充满了争议。一方面,有研究发现,自评一般健康对个体死亡风险具有很强的预测效力(Bailis, *et al.*, 2003; Benjamins, *et al.*, 2004; Benyamini and Idler, 1999; Ferraro and Kelley-Moore, 2001; Frankenberg and Jones, 2004; Idler and Benyamini, 1997; Miilunpalo, *et al.*, 1997; Van Doorslaer and Gerdtham, 2003);另一方面,也有研究指出,自评健康的回答行为在不同社会群体之间存在明显的异质性(reporting heterogeneity),因而该指标在不同人群之间缺乏直接的可比性(Bound, 1991; Groot, 2000; Krause and Jay, 1994; Lindeboom and Van Doorslaer, 2004; Sadana, *et al.*, 2002; Salomon, *et al.*, 2004; Sen, 2002; Tandon, *et al.*, 2002; Thomas and Frankenberg, 2002)。

伊德勒和本亚米尼(Idler and Benyamini, 1997)综述了 27 项有关个体自评一般健康与死亡风险的研究成果,发现超过 85% (23 项)的研究表明自评一般健康对个体的死亡风险具有独立的显著效应,自评一般健康较差的对象在跟踪期内死亡的风险明显更高。这一效应即便在控制了个体的社会人口特征、社会经济地位、社会网络、健康行为,甚至客观健康体测指标后依然显著。因此,他们认为,自评一般健康是反映个体健康状况的非常有效的综合性指标。与其他指标相比,自评一般健康反映的信息更加综合和更为全面,它包括了那些通过医学手段无法检测到的机体感觉和变化,代表着主体对疾病严重程度的综合判断

和预测,甚至隐含着被访者家庭病史等与个体健康变化轨迹密切相关的重要内容。在后续研究中,本亚米尼和伊德勒(Benyamini and Idler, 1999)进一步回顾了在1995-1998年间发表的另外19篇相关文献后发现,尽管较新的研究使用了更为成熟的调查数据和分析方法,但这些研究同样印证了个体自评一般健康与死亡风险的强相关关系。

费拉罗和凯利摩尔(Ferraro and Kelley-Moore, 2001)认为,以往研究中仅使用调查基期时的自评一般健康状况,实际上导致了对个体自评一般健康和死亡风险之间关系的低估。他们利用美国健康与营养监测调查(the National Health and Nutrition Examination Survey)数据研究发现,当只考察被访者基期的自评一般健康状况时,该指标对死亡风险的效应只在白人样本中显著,在黑人样本中并不显著。但当把自评一般健康作为一个时变变量(time-varying covariate)纳入模型后,无论是白人群体还是黑人群体,对死亡风险均有显著的预测效力。本杰明斯等(Benjamins, *et al.*, 2004)进一步检验了美国成年人自评一般健康与死因别死亡风险(cause-specific mortality risk)之间的关系,结果表明,在控制社会人口特征、基期健康状况和健康行为后,自评一般健康状况与死于糖尿病、传染性疾病和呼吸道疾病的风险高度相关,与死于心脏病、中风和癌症等的风险也具有较强的相关关系,但与死于意外事故、凶杀、自杀等的风险不存在显著关系。这些结果进一步印证了自评一般健康包含关于被访者真实健康状况的重要信息。弗兰肯伯格和琼斯(Frankenbergh and Jones, 2004)对印度尼西亚家庭生活调查(the Indonesia Family Life Survey, IFLS)的三期跟踪数据的分析表明,自评一般健康为“差”的个体在调查跟踪期内死亡的概率明显高于自评一般健康为“好”的个体。相应效应在控制了调查对象营养状况、生理和机体活动能力、高血压以及抑郁等众多健康指标后仍然显著。这一研究从发展中国家的数据出发,为自评健康指标的有效性提供了进一步的支持。

尽管有大量研究支持自评一般健康与死亡风险密切相关,但这些证据并没有减少人们对使用自评健康指标考察健康差异的质疑,尤其是自评健康指标可能带来的回答偏误(reporting biases)问题,在不同人群之间的跨人群可比性仍是悬而未决的重要议题(Mathers, 2003)。³

3. 值得指出的是,相应争议不仅出现在关于自评健康指标的应用中,还广泛存在于其他针对主观性调查指标的研究中,如幸福感、满意度等(王广州、王军, 2013)。

与自评一般健康和死亡风险的一致性关系的研究结果相对照,不少研究考察了自评健康指标在测量个体真实健康状况时可能导致的偏误。例如,托马斯和弗兰肯伯格(Thomas and Frankenberg, 2002)指出,在评估自评健康指标的效度时,最大的困难在于我们对每个被访者的“真实”健康状况知之甚少。他们利用美国健康与营养监测调查数据,比较了个体自报的身高、体重与客观测量的身高和体重之间的关系,由于后者只受随机测量误差的影响,在一定程度上可以视为“真值”。分析结果显示,与客观测量结果相比,自报的身高和体重具有系统性的测量误差,这些测量误差与个体特征之间存在非常复杂的关系,因而难以有效修正。他们还利用印度尼西亚家庭生活调查(IFLS)数据,分析了自评一般健康与其他生理健康指标之间的关系,发现自评一般健康确实包含了关于个体健康状况的重要信息,但同时也受到了个体特征与生活经历的影响,如社会经济地位、健康服务利用和关于怎样才算是“健康”的个人标准等,都不同程度地影响着被访者关于自评一般健康的回答结果。

萨达纳等(Sadana, *et al.*, 2002)认为,对健康的界定标准、期望和认知过程存在差异严重影响了自评健康指标的跨人群可比性。基于对46个国家调查数据的比较分析,他们指出,男性被访者的自评一般健康往往好于女性;在部分非洲国家,不同年龄群体的健康自评结果甚至不存在显著差异,这明显与常识相悖。这些分析结果表明,不同性别和年龄群体关于自评健康的回答可能存在明显的异质性,由此严重影响了自评一般健康在不同人群之间的可比性。即便在同一国家内部,不同文化群体之间的健康自评也可能存在系统差异,例如,在以色列,尽管阿拉伯人比犹太人的预期寿命明显更低,各种疾病发生率和死亡率普遍更高,但他们的自评健康状况却好于后者(Baron-Epel, *et al.*, 2005)。

于尔海(Jylhä, 2009)从健康自我评价的社会心理过程出发,对这种回答行为的异质性现象进行解释。他指出,从问答心理的角度看,被访者在回答自评一般健康问题时,首先需要界定什么是“健康”,在不同社会文化和时代背景下,这一定义可能存在差异。其次,被访者会根据自己的界定从记忆中搜集回答问题所需要的相关信息,如疾病史、机体功能、肢体感觉等。当被访者根据这些信息来判断自己的健康状况时,往往将自身情况与一个参照群体做比较,而到底选择怎样的参照群体

可能会因年龄、性别、生活经历而异。最后,当被访者形成了自己的最终判断并依据题目所给出的选项进行回答时,在表达上的文化差异也可能影响被访者倾向于选择哪一个选项。

如何有效辨识和调整自评健康指标可能存在的回答偏误问题,目前已成为国际健康比较研究中的热点和难点之一。其中,一种可行的策略是,在调查中加入辅助的固定情境(anchoring vignettes)问题,即要求被访者在评价自身健康的同时,还根据假想的情境描述评价一系列特定的健康状况。这样,可以通过这些辅助的虚拟情境问题识别每个被访者的回答模式,并据此调整自评一般健康回答行为的异质性以及由此造成的偏误(刘宝、胡善联,2005;Jones, *et al.*, 2007; King, *et al.*, 2004; Murray, *et al.*, 2002; Salomon, *et al.*, 2004; Tandon, *et al.*, 2002)。

另一种考察自评一般健康的回答偏误问题的方法是,在调查数据中同时收集其他能够有效反映被访者真实健康状况的详细信息(Iburg, *et al.*, 2002; Jürges, 2007; Lindeboom and Van Doorslaer, 2004)。该方法的基本思路是,如果我们能够在一定程度上掌握被访者的真实健康状况,那么将自评一般健康状况与之相比,其中的差异自然反映了自评一般健康指标所包含的回答偏误。例如,林德布姆和范杜尔斯勒(Lindeboom and Van Doorslaer, 2004)利用麦克马斯特健康效用指数(McMaster Health Utility Index, HUI)测度被访者的“真实”健康状况,分析了加拿大居民关于自评一般健康回答行为的异质性问题。

为了进一步探讨和检验自评一般健康的效度,本文利用中国较新的健康调查数据,综合调查数据收集的不同维度的主客观健康指标,以模拟个体的“真实”健康状况,从而借助上述方法来考察自评一般健康与“真实”健康状况的一致性。

二、数据与方法

(一) 数据

本研究的数据取自2008年“中国流动与健康调查”(The Survey of Internal Migration and Health in China, IMHC)。该项目是由美国加州大学洛杉矶分校和中国首都医科大学联合完成的全国性调查项目。⁴

4. 有关该项目的进一步介绍,请参见项目网站:<http://www.ccpr.ucla.edu/IM-China>。

调查采用分层整群随机抽样,在中国 28 个省、自治区、直辖市随机选取的 150 个镇级单位,共调查了 3 000 名 18-64 岁的被访者。调查问卷的内容包括被访者的基本社会人口信息、迁移流动史、教育和职业经历、婚姻和家庭状况、被访者的心理健康状况等。在由专业调查员完成问卷调查的同时,经过培训的社区医务人员继续询问了被访者的既往病史和生理健康状况,并为被访者进行了简单的体验。

IMHC 调查先后两次询问了被访者的自评一般健康状况。第一次是在整个调查中作为关于健康的第一个问题,由专业调查员完成提问;第二次是临近调查的末尾,由社区医务人员询问了被访者详尽的疾病史和具体健康状况等问题后,要求被访者重新评价自己的一般健康状况。这两次测量结果和调查收集的其他主客观健康信息为本文分析自评一般健康的信度和效度提供了重要的数据基础。表 1 列出了本研究所使用的主要健康指标及其界定标准。

表 1:本研究所使用的健康指标及其具体定义

健康指标	界定方法
自评/自报指标	
自评一般健康	您认为您身体的健康状况是:(1=很好;2=好;3=一般;4=不好)
慢性病	过去被确诊过以下慢性病(11种)的种类数:高血压、糖尿病、高血脂、心脏病、中风、肾结石或其他肾疾病、支气管炎/肺气肿/肺炎/哮喘或其他肺部疾病、消化道溃疡/胃食管返流性疾病/胃炎/肠炎或其他胃肠道疾病、癌症或恶性肿瘤、肝炎/胆结石或其他肝胆疾病、肺结核
急性病	在过去 30 天中被访者是否患过以下疾病:感冒、流感、急性肺炎、耳部感染、伴随呕吐或腹泻的胃部或食道疾病(1=是;0=否)
疼痛和不适感	在过去三个月内经常出现下列症状(10种)的种类数:头痛、头晕、眼睛酸胀、喉咙痛、肌肉和关节僵硬、肩颈或腰部酸痛、走路时感到双腿沉重、静息时感到胸闷或气短、心区不适、胃部不适
抑郁	CES-D 抑郁量表,共包括 20 个项目,得分大于等于 16 被视为抑郁
失眠	下列五个问题中回答结果为括号内所给出选项的个数:“你是否经常感到入睡困难”(经常);“你是否经常在半夜醒来”(经常);“你是否经常醒得太早而且无法再入睡”(经常);“你是否经常在白天或傍晚感到困倦而需要打个盹儿”(经常);“你是否经常在早上醒来时感到休息得很充分”(极少/从不)
客观体测指标	
体质指数(BMI)	体质指数等于所测量的体重(公斤)除以身高(米)的平方,其取值小于 18.5 为体重偏轻、18.5-25 为正常、25-30 为超重、大于等于 30 为肥胖
高血压	取三次测量的平均值,收缩压 ≥ 140 或舒张压 ≥ 90 被定义为高血压
低肺活量	取三次测量的平均值,呼吸峰流速 < 250 ml 被定义为肺活量偏低

注:CES-D 抑郁量表的详细介绍请参见:Radloff(1977)。

(二) 方法

本文的分析主要包括两个部分。首先,对 IMHC 调查中自评一般健康的两次回答结果进行交叉列联表分析,考察自评一般健康的信度。在此基础上,拟合定序 Probit 模型,探讨影响两次测量结果之间一致性的可能因素。其次,本文借鉴林德布姆和范杜尔斯勒(Lindeboom and Van Doorslaer, 2004)的方法,通过拟合层级定序 Probit 模型(Hierarchical Ordered Probit Model, HOPIT),以检验自评一般健康的效度,以及可能存在的回答偏误问题。与普通的定序模型相比,该模型允许结果变量的各回答类别之间的切点(cut points)可以随被访者个体特征的变动而变动。通过对比不同被访者自评健康切点位置的差异,可以有效辨识不同群体回答行为的异质性,从而考察自评健康指标的跨人群可比性问题。

在定序 Probit 模型中,假定存在一个关于真实健康状况的连续性潜变量(H^*),将自评一般健康(H^s)视为对该潜变量的一种粗糙测量,二者之间的关系可表示为:

$$H^s = i \Leftrightarrow c_{i-1} < H^* \leq c_i, \quad i = 1, \dots, k \quad (1)$$

其中, k 表示自评一般健康所包括的类别数, c_i 为相应的切点, 并有 $c_0 = -\infty$ 和 $c_k = +\infty$ 。在 Hopit 模型中, 允许切点 c_i 随着个体特征(X)如年龄、性别、社会经济地位等的变动而变动, 也即:

$$c_i = g_i(X; \beta_i), \quad i = 1, \dots, k-1 \quad (2)$$

由于真实健康状况(H^*)往往是无法直接观测的, 本文将真实健康状况定义为 IMHC 调查数据中所收集的一系列多维度主客观健康测量指标(H^0)的函数。这些测量指标包括慢性病史、急性病状况、疼痛和不适感、心理健康状况和一系列客观的体测指标。这样就有:

$$H^* = f(H^0, \epsilon; \alpha) \quad (3)$$

将式(2)、(3)带入式(1), 本文所构建的 Hopit 模型可表示为:

$$H^s = i \Leftrightarrow g_{i-1}(X; \beta_{i-1}) < f(H^0, \epsilon; \alpha) \leq g_i(X; \beta_i) \quad (4)$$

式(4)是对 Hopit 模型一般形式的简化。其一般形式不仅允许真实健康状况(H^*)表示为其他健康测量指标(H^0)的函数, 而且真实健康状况也可以随着其他个体特征变量(X)的变化而变化, 即:

$$H^* = f(H^0, X, \epsilon; \alpha, \gamma) \quad (5)$$

以及

$$H^s = i \Leftrightarrow g_{i-1}(X; \beta_{i-1}) < f(H^0, X, \epsilon; \alpha, \gamma) \leq g_i(X; \beta_i) \quad (6)$$

由模型的待估计参数个数与模型约束条件可知,除非数据中包含固定情境变量等额外信息或者对模型参数设定人为的约束条件,当相同的个体特征变量既出现在关于自评一般健康的切点(c_i)的函数中,又出现在关于真实健康状况的函数中时,模型(6)无法拟合。相比之下,它的简化形式模型(4)是可以拟合的,其隐含假定是, H^0 反映了所有基于真实健康状况导致的自评一般健康的差异,因而自评一般健康的其他残余变异都是由被访者回答行为的异质性引起的,这些异质性可以由个体特征来预测和解释。

在定序 Probit 模型的框架下,式(4)可以进一步表示为:

$$\Pr(H^s = i) = \Phi \left[\frac{g_i(X; \beta_i) - f(H^0; \alpha)}{\sigma} \right] - \Phi \left[\frac{g_{i-1}(X; \beta_{i-1}) - f(H^0; \alpha)}{\sigma} \right] \quad (7)$$

其中, $\Phi(\cdot)$ 表示标准正态分布累积分布函数。

三、主要分析结果

(一) 自评一般健康的信度分析

如前所述,在 IMHC 调查中,先后两次询问了被访者的自评一般健康状况。表 2 列出了被访者针对同一问题的两次回答结果的交叉列联表。总体来说,这两次测量的结果具有很强的正相关关系($\gamma = 0.768$, $SE = 0.013$)。表 2 中主对角线上的数值(即两次回答结果一致)明显高于其他数值,这表明,自评一般健康在相当程度上是稳定和一致的健康测量指标。不过,如表 2 所示,在两次测量中,仍然有部分被访者改变了对自己的健康评价结果,这种改变主要发生在相邻类别之间。例如,在第一次自评健康为“很好”的被访者中,有 30% 的人第二次的回答结果为“好”;第一次回答“差”的被访者中,有 33% 的人第二次回答结果为“一般”。相比之下,在两次测量中评价结果从“很好”变成了“差”(或是相反)的被访者比例很低(分别约为 1% 和 3%)。

将两次关于自评一般健康的回答结果进行对比,利用第二次回答结果的值减去第一次的值,可以产生一个表示两次回答差异的新变量。根据两次回答差异的方向,笔者将这个新变量划分为以下三个类别:变

表 2: 被访者关于自评一般健康问题的两次回答
结果的交叉列联表 (%)

第二次的回答	第一次的回答			
	很好	好	一般	差
很好	61	18	7	3
好	30	59	22	8
一般	8	22	66	33
差	1	1	5	56
小计	100	100	100	100
N	468	854	1 302	366

注： $\gamma = 0.768$ (SE = 0.013)。

好(23%)、保持一致(62%)和变差(15%)。通过对该变量拟合定序 Probit 模型,分析两次回答结果之间的变化究竟受到哪些因素影响。

这里所考察的可能影响因素包括被访者的年龄和性别等基本社会人口特征,反映被访者社会经济地位的居住地、教育程度和家庭收入等变量,以及与被访者具体健康状况和健康行为有关的变量。表 3 给出了相应变量的样本分布情况。

表 4 展示了针对自评一般健康两次回答的差异拟合的定序 Probit 模型结果。其中,模型 1 只考虑了被访者的年龄、性别、居住地、受教育程度和家庭收入等基本社会人口特征和社会经济变量,数据结果显示,这些变量对因变量的预测能力有限,所有的模型系数均不显著。这表明,两次关于自评一般健康状况的回答差异跟被访者的人口和社会经济特征无关,而更多地表现为一种随机波动。

模型 2 在模型 1 的基础上加入了多项与被访者健康状况有关的变量,包括慢性病史、急性病状况、身体的疼痛或不适感、抑郁和失眠等自报/自评健康指标,锻炼身体、吸烟和饮酒等健康行为指标,以及体质指数(BMI)、血压和肺活量等客观体测指标。在调查实施过程中,上述自报/自评健康指标和健康行为指标是在两次询问自评一般健康状况之间进行调查的。因此,模型 2 也旨在检验自评一般健康在调查中的题目次序效应(order effect)。模型 2 显示,被访者对自评一般健康问题的回答确实受到了题目次序效应的影响。具体而言,那些在过去一个月内患过急性病、过去三个月间有较多身体疼痛或不适感和失眠症状更严重的被访者,其自评一般健康的第二次回答结果明显比第一次变差(相应回归系数分别在 0.1、0.05 和 0.01 的统计水平上显著)。这表

表 3:相关变量的样本统计分布情况(未加权)

变量名	均值/比例	标准差	最小值	最大值	N
年龄	42.907	12.505	18	64	3 000
性别(男性)	0.395		0	1	3 000
居住地(农村)	0.343		0	1	3 000
教育程度					
小学及以下	0.289				2 999
初中或高中	0.584				2 999
大专及以上	0.127				2 999
家庭收入(千元)	37.586	60.557	0.8	37.5	2 809
低收入组	0.294				
中等收入组	0.393				
高收入组	0.314				
自评一般健康(第一次)					2 999
很好	0.157				
好	0.285				
一般	0.436				
差	0.123				
自评一般健康(第二次)					2 991
很好	0.181				
好	0.323				
一般	0.402				
差	0.094				
慢性病	0.713	1.032	0	7	2 982
急性病	0.387		0	1	2 990
疼痛感	0.583	1.337	0	10	2 983
抑郁	0.241		0	1	3 000
失眠	0.734	1.020	0	5	2 984
体质指数(BMI)	23.823	3.603	14.808	40.859	2 852
偏轻(≤ 18.5)	0.045				
正常($18.5 - 25$)	0.623				
超重($25 - 30$)	0.276				
肥胖(≥ 30)	0.056				
高血压	0.239		0	1	2 853
低肺活量	0.279		0	1	2 847
锻炼身体					2 992
从不	0.547				
每周一次或更少	0.186				
每周两次及以上	0.267				
吸烟(当前仍吸烟者)	0.228		0	1	2 990
饮酒(每周多于一次)	0.147		0	1	2 987

注:有关健康指标的具体定义参见表 1。

明,详细的健康问询可能对被访者具有提醒效应,使之能更全面地评价自己的健康状况。与之相反,有抑郁症状的被访者第二次关于自评一般健康状况的回答结果反倒比第一次的回答显著变好。这可能是因

表 4:关于两次自评一般健康回答结果差异的定序 Probit 模型结果

变量名	模型 1		模型 2	
	回归系数	标准误	回归系数	标准误
年龄	0.00300	0.01200	0.00020	0.01300
年龄的平方	- 0.00008	0.00010	- 0.00009	0.00020
男性	- 0.01600	0.04500	- 0.04900	0.06800
农村居民	- 0.05500	0.04800	- 0.04300	0.05100
教育程度				
小学及以下	-	-	-	-
中学	0.01300	0.05400	0.00040	0.05700
大专及以上	- 0.03600	0.08500	- 0.02600	0.09200
家庭收入				
低收入组	-	-	-	-
中等收入组	0.03200	0.05400	0.04800	0.05700
高收入组	0.06400	0.05800	0.06600	0.06200
慢性病			0.03700	0.02400
急性病			0.08200 [#]	0.04700
疼痛感			0.03400 [#]	0.01900
抑郁			- 0.17100 ^{**}	0.05500
失眠			0.05000 [*]	0.02400
高血压			0.08500	0.05800
体质指数				
≤18.5			0.01700	0.11300
18.5 - 25			-	-
25 - 30			0.05100	0.05300
≥30			- 0.07000	0.10100
低肺活量			- 0.00300	0.05300
锻炼身体				
从不			-	-
每周一次或更少			0.03300	0.06400
至少每周两次			0.01700	0.05600
吸烟			0.06900	0.07100
饮酒			- 0.00070	0.05200
切点 1	- 0.75000	0.25500	- 0.76700	0.27800
切点 2	1.05100	0.25600	1.06500	0.27800
对数似然值(自由度)	- 2 544.4(8)		- 2 367.8(22) ^{**}	
N	2 638		2 798	

注: [#] p<0.1; ^{*} p<0.05; ^{**} p<0.01。

为,调查开始时与抑郁症状相关的消极情绪使这些被访者对很多问题的评价要比实际情况更为负面。但是,随着具体健康状况问题的展开,这些被访者在详细回顾了自身的各项健康指标后,倾向于对开始过于负面的评价予以修正。模型 2 中的其他变量(包括健康行为、客观体测指标等)的回归系数都不显著,表明这些变量对两次自评一般健康回答

结果的一致性不存在明显的效应。

(二) 自评一般健康的效度分析

在关于自评一般健康指标的信度分析之外,本文以下部分对该指标的效度进行了考察。具体而言,本研究结合使用 IMHC 调查收集的一系列主客观健康指标(见表 1),假设被访者的“真实”健康状况可表示为这些健康指标的函数,从而拟合 Hopit 模型以考察自评一般健康指标的有效性。在本文的 Hopit 模型中,分别以两次自评一般健康的回答结果作为因变量,模型允许自评一般健康的切点可以随被访者的年龄、性别、居住地、受教育程度和家庭收入的变动而变动。

表 5 给出了具体的模型拟合结果。由表 5 可见,自评一般健康的两次测量结果与其他自评或自报健康指标均表现出强相关关系,不过,它们与客观体测指标的关系明显较弱。具体而言,两次自评一般健康的回答结果与被访者自报的慢性病史、急性病状况、抑郁和疼痛感具有显著的相关关系。有自报“曾被确诊过的慢性病数量越多”、“在过去 30 天内患过至少一种急性病”和“有较多疼痛或不适感”的被访者的自评一般健康状况明显较差。类似地,有抑郁症状的被访者自评一般健康状况也明显不如其他被访者。失眠症状与自评一般健康状况也存在一定的关系,不过,相应关系仅在第二次自评一般健康回答结果的模型中具有统计显著性,失眠症状越严重的被访者,其自评一般健康状况明显较差。

与上述结果形成明显对照的是,两次测量的自评一般健康状况与模型中所包括的客观体测指标只存在微弱的关系。事实上,只有体质指数(BMI)与两次自评一般健康结果具有显著的相关关系。与体重正常的被访者(BMI 为 18.5 - 25)相比,体重偏轻的被访者(BMI 在 18.5 以下)的自评一般健康状况显著更差。第二次测量的自评一般健康结果与高血压也存在一定的相关性,即临床诊断为高血压的被访者其自评一般健康状况较差,相应回归系数仅在 0.1 的统计水平上显著。

概括而言,本文的分析结果表明,自评一般健康可以很好地反映特定维度的个体健康状况,尤其是那些能够自我感知的身心状况和已知的疾病经历,如既往病史、疼痛感和心理疾病等,但它在反映那些个体不易察觉的机体变化时其效度值得怀疑(Murray and Chen,1992),具体表现为自评一般健康指标与客观体测指标之间的弱相关关系。

表 5:关于自评一般健康的 Hopit 模型结果 (N=2 046)

变量名	第一次自评一般健康		第二次自评一般健康	
	回归系数	标准误	回归系数	标准误
慢性病状况	0.2470 ***	0.0240	0.2920 ***	0.0240
急性病状况	0.2770 ***	0.0460	0.3590 ***	0.0460
抑郁	0.4320 ***	0.0540	0.2810 ***	0.0530
失眠	0.0290	0.0230	0.0870 ***	0.0230
疼痛感	0.1350 ***	0.0190	0.1820 ***	0.0190
高血压	0.0100	0.0560	0.0970 #	0.0560
体质指数				
≤18.5	0.3240 *	0.1090	0.3950 ***	0.1090
18.5 - 25	-	-	-	-
25 - 30	-0.0880	0.0510	-0.0590	0.0510
≥30	0.0320	0.0970	-0.0850	0.0970
低肺活量	0.0790	0.0520	0.0540	0.0520
切点 1				
年龄	-0.0630 ***	0.0180	-0.0790 ***	0.0160
年龄的平方	0.0004 *	0.0002	0.0008 ***	0.0002
男性	0.0390	0.0670	0.1210 #	0.0640
农村居民	-0.0820	0.0720	-0.0220	0.0670
教育程度				
小学及以下	-	-	-	-
中学	-0.0570	0.0850	-0.1080	0.0780
大专及以上	-0.1360	0.1240	-0.1220	0.1150
家庭收入				
低收入	-	-	-	-
中等收入	0.1210	0.0850	0.1110	0.0800
高收入	0.2040 *	0.0890	0.1520 #	0.0840
截距项	1.0610 **	0.3560	1.2410 ***	0.3390
切点 2				
年龄	-0.0530 ***	0.0150	-0.0710 ***	0.0150
年龄的平方	0.0004 *	0.0002	0.0007 ***	0.0002
男性	-0.0220	0.0560	0.0630	0.0560
农村居民	0.0130	0.0570	0.0180	0.0580
教育程度				
小学及以下	-	-	-	-
中学	-0.0220	0.0650	0.0410	0.0650
大专及以上	-0.0110	0.1020	0.0900	0.1030
家庭收入				
低收入	-	-	-	-
中等收入	0.2710 ***	0.0670	0.1470 *	0.0660
高收入	0.4810 ***	0.0720	0.3210 ***	0.0720
截距项	1.4330 ***	0.3140	2.0260 ***	0.3200
切点 3				
年龄	-0.0610 **	0.0220	-0.0680 **	0.0250
年龄的平方	0.0007 **	0.0003	0.0008 **	0.0003
男性	0.1300 #	0.0760	0.0590	0.0840
农村居民	-0.1210	0.0750	-0.1660 *	0.0840
教育程度				
小学及以下	-	-	-	-
中学	0.0810	0.0810	0.0940	0.0920
大专及以上	0.7000 ***	0.1880	0.5400 **	0.1990
家庭收入				
低收入	-	-	-	-
中等收入	0.1350	0.0830	0.0960	0.0930
高收入	0.1610 #	0.0930	0.0220	0.1030
截距项	2.9080 ***	0.4840	3.4420 ***	0.5660
对数似然值	- 2 939.1000 ***		- 2 913.4000 ***	

注：#p<0.1；* p<0.05；** p<0.01；*** p<0.001。

在控制了每个被访者的“真实”健康状况后,Hopit 模型将自评一般健康指标所有残余的变异归结于不同个体回答行为的异质性,也即切点位移(cut-point shifts)现象。由模型结果可知,对于两次自评一般健康状况的回答结果,其各类别对应的切点均存在明显的年龄差异。模型中所有关于年龄的主效应的估计系数均为负,这表明年龄较大的群体自评一般健康的切点小于年轻群体,也即,给定同样的“真实”健康状况,老年人对自己的健康评价似乎更加悲观。由于切点的取值较小,在“真实”健康状况相同的情况下,老年人在回答自评一般健康问题时,选择任何一个较好类别的概率均较低(如选择“很好”而不是“好/一般/差”、“很好/好”而不是“一般/差”、或者“很好/好/一般”而不是“差”的可能性更小)。模型中年龄的二次项效应为正,表明这种差异随着年龄的上升有逐渐缩小的趋势。

模型结果还显示,在不同受教育程度和家庭收入水平的被访者之间,也存在一定的切点位移问题。与受教育程度为“小学及以下”的被访者相比,“大专及以上”的被访者自评一般健康状况的第三个切点取值明显更大。这表明,给定“同等”的健康状况,受过高等教育的被访者比教育程度为“小学及以下”的被访者自评一般健康回答为“差”的可能性更小。相比之下,家庭收入的效应更多反映在前两个切点上。模型中家庭收入的效应均为正,表明给定“真实”健康水平,家庭富裕的被访者更有可能对自己的健康状况评价较高。

此外,在两个模型中,性别对各切点的效应在 0.05 的统计水平上均不显著。这显示,两性之间自评一般健康的差异主要反映了他们关于既往病史的了解和身体对病痛感知敏感程度等方面的差异,而不是两者回答行为的不同。城乡居民关于自评一般健康的回答行为的差异主要反映在第二次测量的第三个切点,即回答结果为“很好/好/一般”和“差”之间。与城镇居民相比,农村居民的相应切点取值更小,表明在控制了模型中所有已测量的健康差异后,农村居民回答自评一般健康为“差”的可能性显著高于城镇居民。

通过以上分析不难看出,自评一般健康能够有效反映被访者对身心健康状况的自我感知和综合认知。与此同时,该指标也存在相当复杂的回答偏误问题。这些回答偏误主要是由于不同年龄和不同社会经济地位的被访者对健康的评价标准、期望和认知水平等方面存在差异,

由此影响了他们对自评一般健康的回答结果及其可比性。

四、结论与讨论

本文利用“中国流动与健康调查”数据,考察了自评一般健康指标的信度和效度。研究发现,尽管自评一般健康仅仅是一项单一的主观性指标,但仍然具有较好的信度,调查中被访者对该问题先后两次回答的结果高度一致。通过对两次回答结果中变动的情况拟合定序 Probit 模型发现,这些变动具有很强的随机性,被访者的个体社会经济特征等因素对相应变动并不具有显著的影响。被访者对自评一般健康的回答在一定程度上会受到调查中题目次序的影响。在提问自评一般健康问题之前先对被访者提问一系列具体的健康问题,有助于被访者更加全面和客观地评价自己的综合健康状况。例如,本文的模型结果显示,在回答了一系列具体健康问题之后,那些近期曾患过急性病、存在经常性身体疼痛或不适感和失眠严重的被访者均调低了对自己一般健康状况的评价。与之相反,有抑郁症状的被访者反倒调高了对自己健康状况的评价,这可能反映了与抑郁相关的消极情绪和过度负面心理对自评健康存在负向影响,通过整理和回答具体的健康问题有助于减缓这种负面心理的影响,从而使被访者能够更加客观地看待自己的健康状况。

此外,本文还通过拟合 Hopit 模型分析了自评一般健康的效度和回答偏误问题。分析结果表明,自评一般健康状况与被访者的其他自评或自报健康指标具有高度相关关系,但与客观健康体测指标的关系则相对较弱。被访者的慢性病史、近期是否患过急性病、身体疼痛或不适感、失眠、抑郁等状况都对自评一般健康具有很强的效应,这表明自评一般健康能够较为综合有效地反映个体自我感知的健康状况。但是,自评一般健康与本文所使用的客观体测指标关系微弱,这说明它难以反映那些个体不易觉察的机体变化,如血压、肺活量等。由此可见,自评一般健康能够有效反映被访者个人感知的身体和精神状况,但不能反映健康的所有维度。有鉴于此,在相关健康调查研究中同时收集各个维度主客观的健康信息,对于全面、深刻地认识健康问题及其社会影响因素必不可少。

值得注意的是,本文的分析结果还表明,自评一般健康存在较为复杂的回答偏误问题。受到不同被访者关于健康的评价标准、期望和认

知水平差异的影响,自评一般健康的回答结果显示出了明显的回答行为异质性或切点位移问题。即便在控制了本文所考察的这些具体健康差异后,老年群体仍然更倾向于“低估”自己的健康状况,而受教育程度较高、家庭收入较高的群体则倾向于更“乐观”地评价自己的健康状况。这在一定程度上会影响自评一般健康指标的跨人群可比性。

尽管如此,自评一般健康仍然是反映个体健康状况的最重要的指标之一。虽然不同群体关于自评一般健康的回答模式可能具有一定的系统差异,针对自评一般健康进行的比较研究仍能在一定程度上有效反映不同人群在许多方面的健康差异,包括慢性病史、急性病状况、身体疼痛感和心理健康状况等。而这些方面的健康问题或者难以通过客观体测指标进行反映,或者极有可能受严重测量误差的干扰(Bound, 1991; Murray and Chen, 1992)。因此,基于健康概念本身的复杂性和多维性,一方面,我们不能寄希望通过任何一项单一的指标就能反映个体健康状况的所有维度,另一方面,本文的研究结论也表明,在众多健康指标中,自评一般健康仍不失为一项较为综合有效的测度指标。

最后,尽管本文所拟合的 Hopit 模型为检验自评一般健康的回答偏误问题提供了依据,但这一方法仍存在一定的局限性。由于该方法假设能够准确测量个体的真实健康状况,任何与“真实”健康测量不同的残余健康差异都归结为回答行为的异质性,因此,利用该方法的分析结果对“真实”健康状况的测量高度敏感。在现实中,很难构建一个或一组健康指标来全面反映个体的真实健康状况,我们对此只能通过尽可能广泛、有效的健康指标来与之近似。在本文的分析中,笔者利用了“中国流动与健康调查”中所收集的丰富的主客观健康信息来对被访者的“真实”健康状况进行近似,但自评一般健康所反映的信息仍然有可能超出了这些可测量的健康指标的范畴。因此,本文关于自评健康回答行为异质性的结论具有一定的尝试性,其准确性和稳健性仍有待后续研究的进一步检验。更有效地解答相关研究问题,还有待于我们在健康理论和研究方法上新的突破。

参考文献(References)

- 刘宝、胡善联. 2005. 人群自报健康资料切点位移偏倚的校正——健康情景方法[J]. 中国卫生统计(5):314-317.

- 王广州、王军. 2013. 中国家庭幸福感测量[J]. 社会 33(6):139 - 160.
- 王曲、刘民权. 2005. 健康的价值及若干决定因素:文献综述[J]. 经济学季刊(1):1 - 52.
- Bailis, Daniel, Alexander Segall, and Judith G. Chipperfield. 2003. "Two Views of Self-Rated General Health Status." *Social Science & Medicine* 56(2):203 - 217.
- Baron-Epel, Orna, Giora Kaplan, Amalia Haviv-Messika, Jalal Tarabeia, Manfred S. Green, and Dorit Nitzan Kaluski. 2005. "Self-Reported Health as a Cultural Health Determinant in Arab and Jewish Israelis." *Social Science & Medicine* 61(6):1256 - 1266.
- Benjamins, Maureen Reindl, Robert A. Hummer, Isaac W. Eberstein, and Charles B. Nam. 2004. "Self-Reported Health and Adult Mortality Risk: An Analysis of Cause-Specific Mortality." *Social Sciences & Medicine* 59(6):1297 - 1306.
- Benyamini, Yael and Ellen L. Idler. 1999. "Community Studies Reporting Association between Self-Rated Health and Mortality: Additional Studies, 1995 to 1998." *Research on Aging* 21(3):392 - 401.
- Bound, John. 1991. "Self-Reported Versus Objective Measures of Health in Retirement Models." *The Journal of Human Resources* 26(1):106 - 138.
- Ferraro, Kenneth F. and Jessica A. Kelley-Moore. 2001. "Self-Rated Health and Mortality among Black and White Adults: Examining the Dynamic Evaluation Thesis." *Journal of Gerontology: Social Sciences* 56B(4):S195 - S205.
- Frankenberg, Elizabeth and Nathan R. Jones. 2004. "Self-Rated Health and Mortality: Does the Relationship Extend to a Low Income Setting?" *Journal of Health and Social Behavior* 45(4):441 - 452.
- Groot, Wim. 2000. "Adaptation and Scale of Reference Bias in Self-Assessments of Quality of Life." *Journal of Health Economics* 19(3):403 - 420.
- Iburg, Kim Moesgaard, Joshua A. Salomon, Ajay Tandon, and Christopher J. L. Murray. 2002. "Cross-Population Comparability of Physician-Assessed and Self-Reported Measures of Health." In *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*, edited by Christopher J. L. Murray, Joshua A. Salomon, Colin D. Mathers, and Alan D. Lopez. World Health Organization: Geneva:433 - 448.
- Idler, Ellen L. and Yael Benyamini. 1997. "Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies." *Journal of Health and Social Behavior* 38(1):21 - 37.
- Jones, Andrew M., Nigel Rice, Teresa Bago d'Uva, and Silvia Balia. 2007. *Applied Health Economics*. London and New York: Routledge.
- Jürges, Hedrik. 2007. "True Health vs. Response Styles: Exploring Cross-Country Differences in Self-Reported Health." *Health Economics* 16(2):163 - 178.
- Jylhä, Marja, J. M. Guralnik, L. Ferrucci, J. Jokela, and E. Heikkinen. 1998. "Is Self-Rated Health Comparable across Cultures and Genders?" *Journal of Gerontology Series B: Psychological Sciences and Social Sciences* 53(3):S144 - S152.
- Jylhä, Marja. 2009. "What is Self-Rated Health and Why does it Predict Mortality? Towards a Unified Conceptual Model." *Social Science & Medicine* 69(3):307 - 316.
- King, Gary, Christopher J. L. Murray, Joshua A. Salomon, and Ajay Tandon. 2004. "Enhancing the Validity and Cross - Cultural Comparability of Measurement in Survey Research." *American Political Science Review* 98(1):191 - 207.
- Krause, Neal M. and Gina M. Jay. 1994. "What do Global Self-Rated Health Items Measure?" *Medical Care* 32(9):930 - 942.
- Lindeboom, Maarten and Eddy van Doorslaer. 2004. "Cut-Point Shift and Index Shift in

- Self-Reported Health." *Journal of Health Economics* 23(6):1083 - 1099.
- Mathers,Colin D. 2003. "Commentary: Towards Valid and Comparable Measurement of Population Health." *Bulletin of the World Health Organization* 81(11):787 - 788.
- Miilunpalo,Seppo,Iikka Vuori,Pekka Oja,Matti Pasanen, and Helka Urponen. 1997. "Self-Rated Health Status as a Health Measure: The Predictive Value of Self-Reported Health Status on the Use of Physician Services and on Mortality in the Working-Age Population." *Journal of Clinical Epidemiology* 50(5):517 - 528.
- Murray,Christopher J. L. and Lincoln C. Chen. 1992. "Understanding Morbidity Change." *Population and Development Review* 18(3):481 - 503.
- Murray,Christopher J. L. ,Ajay Tandon,Joshua A. Salomon,Colin D. Mathers, and Ritu Sadana. 2002. "New Approaches to Enhance Cross-Population Comparability of Survey Results." In *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*, edited by Christopher J. L. Murray, Joshua A. Salomon,Colin D. Mathers, and Alan D. Lopez. World Health Organization; Geneva: 421 - 431.
- Radloff,Lenore Sawyer. 1977. "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population." *Applied Psychometric Measurement* 1(3):385 - 401.
- Sadana, Ritu, Colin D. Mathers, Alan D. Lopez, Christopher J. L. Murray, and Kim Moesgarrrd Iburg. 2002. "Comparative Analysis of More than 50 Household Surveys on Health Status." In *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*, edited by Christopher J. L. Murray, Joshua A. Salomon,Colin D. Mathers, and Alan D. Lopez. World Health Organization; Geneva: 370 - 386.
- Salomon,Joshua A. ,Ajay Tandon, and Christopher JL Murray. 2004. "Comparability of Self-Rated Health:Cross-Sectional Multi-Country Survey Using Anchoring Vignettes." *British Medical Journal*, doi:10.1136/bmj.37963.691632.44.
- Sen,Amartya. 2002. "Health: Perception Versus Observation." *British Medical Journal* 324(7342):860 - 861.
- Tandon,Ajay, Christopher J. L. Murray,Joshua A. Salomon, and Gary King. 2002. "Statistical Models for Enhancing Cross-Population Comparability." *Global Programme on Evidence for Health Policy Discussion Paper* 42.
- Thomas,Duncan and Elizabeth Frankenberg. 2002. "The Measurement and Interpretation of Health in Social Surveys." In *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*, edited by Christopher J. L. Murray,Joshua A. Salomon,Colin D. Mathers, and Alan D. Lopez. World Health Organization; Geneva:387 - 420.
- Van Doorslaer,Eddy, and Ulf-G. Gerdtham. 2003. "Does Inequality in Self-Assessed Health Predict Inequality in Survival by Income? Evidence form Swedish Data." *Social Science & Medicine* 57(9):1621 - 1629.
- World Health Organization. 1946. *WHO Definition of Health*. Preamble to the Constitution of the World Health Organization as Adopted by the International Health Conference, New York, 19 - 22 June.

责任编辑:张 军