

临床检验测量不确定度的进一步认识*

张诗诗^{1,2}, 王 薇¹, 赵海建¹, 王治国^{1,2} (1. 北京医院 国家老年医学中心
卫生部临床检验中心/北京市临床检验工程技术研究中心, 北京 100730;
2. 北京协和医学院研究生院, 北京 100730)

摘要:在临床检验中,测量不确定度(measurement uncertainty, MU)是测量系统检验结果的固有属性。作为 ISO 15189 中的一个重要组成部分, MU 是临床实验室进行方法确认或性能验证时必须确定的,且需要对其不断进行检查。虽然临床实验室目前提供的检验报告中通常并未包含 MU,但是部分实验室已经在常规工作中普遍进行了 MU 的计算。MU 的确定和监测可以辅助临床实验室提供更准确的结果并为临床医生提供客观工具用于结果解释。结果解释一般通过结果比较来进行,主要的结果比较对象有 3 种,分别是相同患者之前的结果、人群参考区间,以及临床决定点。由于每种类型比较中“真”值的含义以及对 MU 评定有贡献的组分不同,因此 MU 的最佳评定方式也有所不同,最终会导致 MU 的不同并影响临床决策的制定。显然,应该依照检测的实际临床用途选择合适的比较对象来进行结果解释并确定 MU 的最佳评定方式。对于相同检测的不同临床用途(诊断或监测),需要采用不同的 MU 估计方式来获取合理的 MU。该文通过解释 MU 的概念和特征、评定和应用,并说明三种主要的结果比较方式如何获取 MU,旨在帮助临床实验室进一步认识 MU 的重要性并为实验室常规工作中的 MU 评定提供指导。

关键词:测量不确定度;总误差;结果解释

中图分类号:R446 **文献标志码:**A **文章编号:**1671-7414(2017)02-001-05

doi:10.3969/j.issn.1671-7414.2017.02.001

Further Understanding of Measurement Uncertainty in Clinical Laboratory Medicine

ZHANG Shi-shi^{1,2}, WANG Wei¹, ZHAO Hai-jian¹, WANG Zhi-guo^{1,2}

(1. National Center for Clinical Laboratories/Beijing Engineering Research Center of
Laboratory Medicine, Beijing Hospital, National Center of Gerontology, Beijing 100730, China;
2. Graduate School of Peking Union Medical College, Beijing 100730, China)

Abstract: In clinical laboratory medicine, measurement uncertainty (MU) is a fixed property of testing results in the measuring system. As an important part of ISO 15189, it is necessary for clinical laboratories to determine MU during the period of validation and verification for each measurement procedure and to review MU over time. Now, testing reports provided by clinical laboratories usually do not offer MU, but some clinical laboratories have already estimated MU in their routine work. Estimation and monitoring of MU can help clinical laboratories offering more accurate results and provide objective tools for clinicians used in result interpretation. Generally, result interpretation can be achieved by the result comparison with three main comparators, including a previous result from the same patient, a population reference interval and a clinical decision point. The means of “true value” and the components contributing to the estimation of MU are both different when the comparison is conducted between testing results and different comparators, so the optimum estimation method of MU is accordingly different, which will subsequently affect the MU value and the determination of clinical decisions. Obviously, depending on the actual clinical uses, laboratories can choose appropriate comparators to the result interpretation and the determination of optimum estimation method of MU. For different clinical uses (diagnosis or monitoring) of the same measurands, the adoption of different estimation methods should be used to acquire reasonable MU. By interpreting the concept, characteristics, estimation, and uses of MU, as well as explaining how three main comparison methods of results exploit their own traceable chain to get MU, this paper intends to help clinical laboratories get further understanding of the importance of MU and provide guidance for the MU estimation in routine work.

Keywords: measurement uncertainty; total error; result interpretation

在临床检验中,全球通用的质量管理标准 ISO 15189 已经被广大实验室很好地接受,其焦点包括:实验室管理、质量控制和结果变异性、结果解

释、参考区间、伦理学和实验室安全等。近年来,采用 ISO 15189^[1]对临床实验室进行认可使得测量不确定度(measurement uncertainty, MU)的评定

* 作者简介:张诗诗(1991—),女,在读硕士研究生,研究方向:实验室质量控制管理,E-mail: Athena0915@163.com。

通讯作者:王治国(1964—),男,硕士,研究员,主要从事临床检验质量控制方法和室内质量评价工作,E-mail: zgwang@nccl.org.cn。

成为了一个必需进行的活动。计量学中规定 MU 是测量结果本身的一部分^[2], VIM 条款 2.26 也将 MU 明确定义为“根据所用信息, 表征被测量量值离散性的非负参数”^[3]。MU 的使用可以帮助临床实验室和临床医生更全面充分地对实验室测量结果进行理解、认识 and 解释, 不仅有助于实验室选择可靠的测量程序并改进实验室质量, 同时也可以将实验室的测量结果恰当地应用于临床诊断和治疗。

1 临床检验测量不确定度的特征 测量不确定度表示指南(GUM)中的测量不确定度模型和 Westgard 总误差模型的使用一直存在诸多争议。然而, 二者并非完全相反或矛盾的两种观点, 只是对测量结果本质的表达有所差异。

MU 指的是围绕测量结果的区域, 在该区域内“真”值以规定的概率出现; 而总误差 (total error, TE) 指的是围绕“真”(参考) 值的区域, 在该区域内测量结果以规定的概率出现^[4]。MU 被定义为一个量值区间, 它与测量结果的测得量值的可控程度呈反比关系。原则上, MU 只有一个组分类型, 也就是无论其来源类型均可用同一方式处理所有组分, 这与 TE 是由系统误差和随机误差 2 种组分构成有着明显的不同。此外, MU 实际上是一个不可知的量值, 并且不能用来修正; 而 TE 中系统误差的估计值是已知且可以被纠正的。对于“真值”和“误差”不可知的本质, MU 并未关注于此, 而是将测量的各种要素组合在一起并以 MU 来表达量值不完整性的信息。此外, GUM 中认为, MU 和 TE 的主要区别在于一个重要的假设, 即是否在测量早期识别并纠正所有系统误差。根据 MU 模型假设在测量早期识别并纠正所有系统误差, 所以当分析方法存在显著偏倚时, 需从每个测量结果中移除偏倚并将纠正偏倚所引起的不确定度包含在 MU 的评定中。

2 临床检验测量不确定度的评定和应用 ISO 15189 条款 5.5.1.4 规定“实验室应为检验过程中用于报告患者样品被测量值的每个测量程序评定测量不确定度。实验室应规定每个测量程序的测量不确定度性能要求, 并定期评审测量不确定度的评定结果。”显然, 如何评定 MU 是实验室质量管理中非常重要的环节。

目前, 有 2 种方法可用于医学实验室 MU 的评定, 分别是: ①自下而上 (bottom-up) 的方法, 通常特指 GUM 方法或模型方法; ②自上而下 (top-down) 的方法。其中, 后者对于常规实验室而言, 是较为经济实用的方法^[5]。“自上而下”的 MU 评定方法是针对不精密度和偏倚分别进行评估, 它通过将有关于所有随机误差 (不精密度) 的不确定度

加上有关于所有系统误差 (偏倚) 纠正的不确定度, 可以对整个测量程序的不确定度进行表达并反映出相应的质量水平。对于大部分医学实验室的应用而言, 一般使用室内质量控制及其均值和标准差的计算来评估不精密度, 并且在直接的溯源性研究难以实现时采用能力验证计划中的数据来评估偏倚。然而, 正如 Westgard 的观点, 偏倚可由多种原因导致, 例如缺乏仪器和方法间的标准化和一致化等, 使得偏倚的检测成为许多医学实验室的重要问题^[6]。因此, 可以发展类似于检测“一致化”的策略将偏倚最小化并且可能无需合并入 MU 的确定中^[7]。

检验医学的目标是产生可用于临床决策以提高患者健康水平的信息^[8]。MU 作为测量结果的固有属性, 不仅可以提供准确的检测结果, 同时也可以有效地辅助结果解释, 特别是判断同一被测量在不同次测量结果之间是否有显著差异^[2]。对于相同的被测量, 源自不同次测量的结果可以基于它们的 MU 进行相互比较, 因为 MU 模型已经从每个测量结果中移除偏倚并且将纠正偏倚所引起的不确定度包含在 MU 的评定中。

此外, ISO 15189 中将医学实验室完整的检测过程分为检验前、检验中和检验后 3 个阶段。这 3 个阶段在理论上都存在着 MU 的来源。因此 ISO 15189 强调“医学实验室认可不能仅仅关注检验中阶段的 MU, 同时也应该关注检验前和检验后阶段的缺陷”。同时, 越来越多的证据也表明了检验前和检验后阶段的缺陷会对检测的质量水平产生影响^[9]。实验室应该以患者为中心, 通过采用有价值且一致化的质量指标来对检验前、后阶段的质量问题进行管理^[10,11]。但是, 如何将检验前、后阶段的 MU 整合入完整检测过程的 MU 评定依然有待研究和探讨。

3 检验结果如何通过结果比较获得测量不确定度

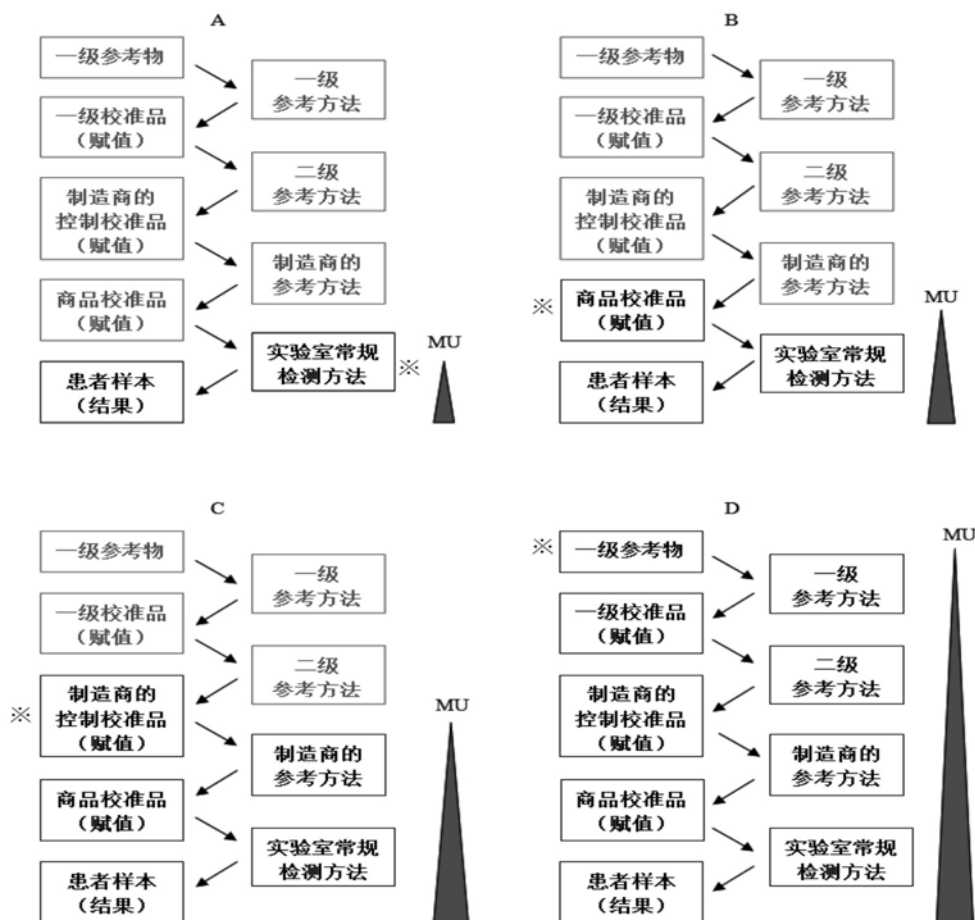
MU 的确定和监测有 2 个目的: 第一, 在实验室内部提供准确的检测结果; 第二, 为用户提供客观的工具用于结果解释^[7]。对于前者, 需要注意的是, 采用通用算法评定所有测量结果的 MU 是不合适的, MU 的评定实际上应该取决于“真”值是如何获得的; 对于后者, 这是非常值得深入研究的一方面, 结果解释大部分是通过比较进行, 正确的结果解释应该取决于结果的比较对象 (主要有 3 种, 分别是同一患者之前的结果、人群参考区间、临床决定点)^[12]。在比较过程中, 当偏倚保持内部一致并且与比较者相同时, 则偏倚在很大程度上是无关紧要的, 而 MU 和 TE 在本质上均成为不精密度的评价。同时, 在利用结果比较进行结果解释时, 必需

考虑检测结果如何用于患者临床决策的制定。因为若使用不同类型的比较对象或不同情况下的同类比较对象进行结果比较,则“真”值的含义、对MU有贡献的组分是不同的,均会影响测量结果的MU以及临床决策的制定。另外,对于每一种比较类型,用于测量比较对象的仪器如果不存在MU,则从该仪器获得的结果即为“真”值。

3.1 与同一患者之前的结果比较 这种比较最普遍且最简单的做法是在单一的实验室使用单一的仪器重复检测患者样本,监测某一分析物在同一患者体内的浓度是否增加、减少或保持稳定。其中,“真”值可用产生之前结果的方法来获得。在这种情况下,评定的是分析不精密度引入的MU,可以表示为标准MU(如分析标准偏差,SDa)、相对MU(如分析变异系数,CVa)、或扩展MU(含有覆盖因子,通常为2,表明此区间覆盖了大约95%的期望值)。由于SDa和CVa更易与个体内生物学变异结合用于判断同一患者的测量结果之间是否差异显著,因此一般不采用扩展MU^[13]。此外,利用这种比较来进行结果解释时,结果之间的测量时

间间隔会有所不同,导致制定合理临床决策的时限也有所不同。然而,在不同的时限内,检测方法、重新校准、试剂或校准品批号的改变等都会对MU产生影响,甚至可能要求在MU的评定中包含为校准品赋值而产生的MU。根据是否在同次校准内,我们可以将该类比较分成2种情况。

第一种情况是在同次校准内,使用源自同一分析仪的结果,监测来自同一患者的结果。此时,“真”值使用同次校准内分析仪的最佳性能来定义的,而最佳性能则是通过同次校准内重复分析样本而进行经验性地确定。在该情况下,结果的改变独立于校准、校准程序或校准品本身所产生的任何不确定度,因此在判断两个结果之间是否存在显著变化时必须考虑两次测量的MU。例如,在8 h内进行连续监测血清肌钙蛋白的浓度,判断患者肌钙蛋白的浓度水平是否出现显著的变化,从而对急性心肌梗死进行定义^[14]。在此种情况下,仅同次校准内的组分包含在MU的评定中,所有MU组分的溯源链见图1A所示。



注:在每张图中用※标明评定MU的“真值”,并且用灰色三角形、黑色加粗字体和黑色方框指出对MU有贡献的组分,对制定相关临床决策的MU没有贡献的组分用灰色加粗字体和灰色方框表示。

图1 测量结果与不同比较对象在各种情况下比较的溯源链

第二种情况是在使用相同校准品进行重新校准时,采用源自同一分析仪的结果监测来自同一患者的结果。在这种情况下,两次测量通常间隔1天或者少数几天,一般是在使用同一批号校准品进行不同次校准的一段时间内。例如,血清肌酐的检测,要求对分析仪进行日间校准,并按照 RIFLE 标准确定急性肾损伤^[15]。另外,若该情况用于评价患者在一段长时间内(例如,年)的改变,则 MU 也应该包括为校准品赋值的不确定度。包含该情况下所有 MU 组分的溯源链,见图 1B 所示。

然而,对于以上任意一种情况,若已知分析存在偏倚,在进行偏倚纠正(即使是对所有结果均采用相同的纠正)后将纠正偏倚产生的不确定度与相关的分析 SDa 或 CVa 相结合,会导致每个结果有更大的不确定度,并因此产生更大的参考变化值,导致错误临床决策的制定。因此,为了避免影响测量结果,该类型的比较不适合将固定的不确定度与分析不精密度相结合。

3.2 与人群参考区间比较 对于几乎所有分析物的检测结果,检测报告中提供的参考区间是其最初的比较对象,而参考区间由实验室自行建立或者由使用实验室进行确认^[16]。如果实验室自行建立参考区间,则比较的“真”值可以使用产生参考区间的方法来决定。如果用于建立参考区间的方法和实验室常规检测方法之间的偏倚是已知且可纠正的,则纠正过程产生的不确定度应该包括在 MU 的计算内。

如果实验室使用制造商指定的参考区间,则必须明确制造商指定的参考区间是在不同实验室内于不同时间采用相同方法进行一个或多个研究并基于此获得的。该情况下,“真”值的最佳估计是由制造商为校准品赋值时所采用的方法而决定的,产生结果相关的 MU 包括长期的分析不精密度、制造商控制校准品赋值的不确定度等。因此,包含该情况下所有 MU 组分的溯源链见图 1C 所示。然而,在实验室中,若使用不同于制造商的方法(可溯源到更高级的参考物质)建立参考区间,则“真”值是参考物质的赋值。

3.3 与临床决定点的比较 在这类比较中,临床决定点可能来自于国际或国家指南、临床研究或临床医生的经验。其中,“真”值是通过确定临床决定点的临床研究中所采用的方法来获得的。对于许多已发布的临床决定点,已声明可溯源到更高级的参考物质和参考方法。对于建立临床决定点所用的方法,“真”值位于其溯源链的顶端。例如,使用血浆葡萄糖的诊断决定点进行糖尿病的诊断^[17]。这些决定点以多个研究的评价为基础,且这些研究

均使用可溯源至一级参考物质(已认证)的方法来进行,则“真”值是由一级参考物质决定的。因此,包含该情况下所有 MU 组分的溯源链见图 1D 所示。

4 总结 在临床检验的质量管理中, MU 是一个非常重要的概念和工具。在医学实验室方面, MU 可用于改进检验质量、选择测量程序、加强临床沟通,也可以设定目标不确定度并通过判断检验结果的 MU 是否符合其目标不确定度的要求来说明实验室的检验结果是否达到期望的质量水平;在临床方面, MU 可以辅助医生更科学合理地解释检验结果并作出更准确的临床决策。然而,如何评定 MU 一直是一个令人困扰的问题。为此, ISO 15189 中并未提出 MU 的具体评定方法,仅 GUM 中提供了评定 MU 的基本框架。目前,对常规实验室最实用的“自上而下”的 MU 评定方法是将不精密度和偏倚分别进行估计,但是偏倚可能由许多原因导致。因此,需要发展诸如检测“一致化”的替代策略来使偏倚最小化。

此外,必须明确的是, MU 的其中一个重要目的是用于结果解释,一般通过结果比较来进行。本文中给出 3 种不同的结果比较对象在不同情况下的“真”值含义、对 MU 评定有贡献的组分以及实例,共同表明 MU 的最佳确定方式应该是依照检测的临床用途,也就是必须考虑检测结果如何用于患者临床决策的制定。因为在使用检测结果制定临床决策时,需要对检测结果进行解释,而解释结果所用的不同类型(3 种)的比较对象,会导致 MU 的评定方式有所不同。总之,基于检测的一种临床用途而计算的 MU,可能并不适用于该检测的其他临床用途并且可能会导致错误的临床决策。

参考文献:

- [1] ISO 15189: 2012. Medical laboratories-requirements for quality and competence[S]. Geneva, Switzerland: International Organization for Standardization, 2012.
- [2] Joint Committee for Guides in Metrology. Evaluation of measurement data-guide to the expression of uncertainty in measurement[S]. JCGM100, 2008.
- [3] Joint Committee for Guides in Metrology. International vocabulary of metrology-basic and general concepts and associated terms (VIM)[S]. JCGM200, 2008.
- [4] 王治国. 临床检验质量控制技术[M]. 3 版. 北京: 人民卫生出版社, 2014: 675-704.
Wang ZG. Quality control technology in clinical laboratory[M]. 3th Ed. Beijing: People's Medical Publishing House, 2014: 675-704.

(下转 9 页)

分子标志物的潜能。另外,术后血清中 miR-483-5p 的水平可以帮助预测病人的预后,后期研究中需进一步增加研究的病例数,延长随访时间,以便更全面评估术后 miR-483-5p 水平对病人预后的判断。

参考文献:

- [1] Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012[J]. CA Cancer J Clin, 2015, 65(2): 87-108.
- [2] Chen X, Ba Y, Ma L, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases[J]. Cell Res, 2008, 18(10): 997-1006.
- [3] Zhang C, Wang C, Chen X, et al. Expression profile of microRNAs in serum: A fingerprint for esophageal squamous cell carcinoma[J]. Clin Chem, 2010, 56(12): 1871-1879.
- [4] Liu R, Zhang C, Hu Z, et al. A five-microRNA signature identifies from genome-wide serum microRNA expression profiling serves as a fingerprint for gastric cancer diagnosis[J]. Eur J Cancer, 2011, 47(5): 784-791.
- [5] Chen X, Liang H, Guan D, et al. A combination of let-7d, let-7g and let-7i serves as a stable reference for normalization of serum microRNAs[J]. PLoS One, 2013, 8(11): e79652.
- [6] Xu H, Yang Y, Zhao H, et al. Serum miR-483-5p: a novel diagnostic and prognostic biomarker for patients with oral squamous cell carcinoma[J]. Tumour Biol, 2016, 37(1): 447-453.
- [7] Xu X, Chen Z, Zhao X, et al. MicroRNA-25 promotes cell migration and invasion in esophageal squamous cell carcinoma[J]. Biochem Biophys Res Commun, 2012, 421(4): 640-645.
- [8] Meng F, Qian L, Lv L, et al. MiR-193a-3p regulation of chemoradiation resistance in oesophageal cancer cells via the PSEN1 gene[J]. Gene, 2016, 579(2): 139-145.
- [9] 唐海霞, 王成, 路美玲, 等. 肾透明细胞癌患者血清 miR-193a-3p 水平增加及早期诊断价值[J]. 现代检验医学杂志, 2015, 30(3): 43-45, 49.
Tang HX, Wang C, Lu ML, et al. Elevated serum miR-193a-3p level in renal clear cell carcinoma and its early clinical value[J]. Journal of Modern Laboratory Medicine, 2015, 30(3): 43-45, 49.
- [10] Hamdi K, Goerlitz D, Stambouli N, et al. miRNAs in sera of tunisian patients discriminate between inflammatory breast cancer and non-inflammatory breast cancer[J]. Springerplus, 2014(3): 636.

收稿日期: 2017-01-20

修回日期: 2017-02-10

(上接 4 页)

- [5] CNAS-TRL-001: 2012. 医学实验室—测量不确定度的评定与表达[J]. 北京: 中国合格评定国家认可委员会, 2012.
- [6] Farrance I, Badrick T, Sikaris KA. Uncertainty in measurement and total error—are they so incompatible? [J]. Clin Chem Lab Med, 2016, 54(8): 1309-1311.
- [7] Tate JR, Plebani M. Measurement uncertainty—a revised understanding of its calculation of its calculation and use[J]. Clin Chem Lab Med, 2016, 54(8): 1277-1279.
- [8] Jones GR. Laboratory analytical quality—the process continues[J]. Clin Chem Lab Med, 2016, 54(8): 1275-1276.
- [9] Plebani M. The detection and prevention of errors in laboratory medicine[J]. Ann Clin Biochem, 2010, 47(2): 101-110.
- [10] Plebani M, Sciacovelli L, Aita A, et al. Performance criteria and quality indicators for the preanalytical phase[J]. Clin Chem Lab Med, 2015, 53(6): 943-948.
- [11] Sciacovelli L, Aita A, Padoan A, et al. Performance criteria and quality indicators for the post-analytical phase[J]. Clin Chem Lab Med, 2016, 54(7): 1169-1176.
- [12] Jones GR. Measurement uncertainty for clinical laboratories—a revision of the concept[J]. Clin Chem Lab Med, 2016, 54(8): 1303-1307.
- [13] Fraser CG. Reference change values[J]. Clin Chem Lab Med, 2011, 50(5): 807-812.
- [14] Thygesen K, Alpert JS, Jaffe AS, et al. Third universal definition of myocardial infarction[J]. Circulation, 2012, 126(16): 2020-2035.
- [15] Bellomo R, Ronco C, Kellum JA, et al. Acute renal failure—definition, outcome measures, animal models, fluid therapy and information technology needs: the Second International Consensus Conference of the Acute Dialysis Quality Initiative (ADQI) Group [J]. Crit Care (London England), 2004, 8(4): R204-212.
- [16] Jones GR. Validating common reference intervals in routine laboratories[J]. Clin Chim Acta, 2014, 432(5): 119-121.
- [17] American Diabetes Association. Erratum. Classification and diagnosis of diabetes. Sec. 2. In Standards of Medical Care in Diabetes-2015 [J]. Diabetes Care, 2015, 38(Suppl 1): S8-16.

收稿日期: 2016-09-26

修回日期: 2017-01-02