

我国骨质疏松症诊断标准的制定原则及方法探讨

聂四平

中图分类号: R195.2 ;R681 文献标识码: A 文章编号: 1006-7108(2008)04-0270-15

摘要: 对标准的含义、制定方法和影响因素、制定目的以及新医学模式对标准制定的影响等方面进行比较详细的讨论,强调了在目前制定理想标准的重要性和可行性。在此基础上讨论了遗传因素、环境因素对我国各民族的骨质疏松症诊断标准制定的影响,指出我国只应制定 1 或 2 个全国统一的标准而不是更多的标准。对骨质疏松症诊断标准的制定过程中所涉及的参照人群、评价方法进行了具体的讨论,并首次对骨质疏松症的诊断界值如“理想正常值”、“亚疾病值”和“疾病值”的确定方法在统计学上通过具体的实例进行了详细的推论。还指出了目前我国的骨质疏松症诊断标准在统计学及具体制定过程中存在的主要问题,同时对标准的制定及相关研究也提出了进一步的展望。

关键词: 中国;骨质疏松症;标准;骨骼发育

A Probe into the making principle and methods of diagnostic criterion of osteoporosis in China NIE

Siping. Department of Nutrition and Food Hygiene, School of Public Health, Guiyang Medical College, Guiyang 550004, China

Abstract: This article discussed the definition of criterion, methods of making criterion and influential factors, purpose of making the criterion, and the influence of new medical mode on criterion making, emphasizing the importance and feasibility of making ideal criteria. Based on the discussion, it further discussed the influences of hereditary and environmental factors on the diagnosis criteria of osteoporosis among different ethnic groups in China, pointing out that only one or two national criterion(a) is(are) needed. Statistically specific discussion and deduction have been made by case study about the method to define control subjects, evaluation methods, and relevant diagnosis values such as ideal normal value, sub-illness value, and illness value, which are all involved in making the diagnosis criteria for osteoporosis. In addition, this article also pointed out the major problems in making diagnosis criteria of osteoporosis in China and proposed prospect for relevant researches.

Key words: China; Osteoporosis; Criterion; Bone growth

我国第 5 次人口普查结果显示 2000 年全国总人口为 129 533 万人,其中 65 岁以上人口达 8 811 万人,占总人口 6.96%。而骨质疏松症(Osteoporosis, OP)的诊断标准作为骨质疏松症的诊断、防治的重要依据及相关研究的常用工具,其制定是否正确而合理,对于庞大的老年人群而言,其重要性不言而喻。但目前骨质疏松症诊断标准在制定原则和方法上尚不统一,显然不利于骨质疏松症的诊断、防治及相关研究。笔者试图通过相关学科的一些概念及研

究,对骨质疏松症的诊断标准的制定原则和方法等进行探讨。

1 标准的概念

1.1 标准的含义、制定方法和主要影响因素

标准是评价某同一总体人群中的个体和/或群体的某种生理状态的统一尺度。

标准可客观地描述同一总体人群中个体或群体的某种状态的水平(该水平也可理解为是在该同一总体人群中的位置或等级),采用同一标准评价同一总体中的个体或群体所得到的评价结果间可相互比较。

标准的制定通常是采用标准化的测量方法,在

作者单位: 550004 贵阳,贵阳医学院公共卫生学院营养与食品卫生学教研室

通讯作者: 聂四平, Email: niesiping_nsp@163.com

某同一总体人群中选取一定量的、有代表性的样本进行横断面调查(有条件时最好应结合纵向调查),获得某种状态各项指标的测量数据。在以往的研究及实际医学意义的基础上,并按照标准制定的目的、要求对数据进行必要的分类(或分组,如按性别、年龄等划分),再进行合理的统计学处理,根据其数据的分布类型计算出所需指标的均数、标准差、百分位数等诸多参数。最后将这些参数进行综合后所得到的规范化的、可用于评价个体“正常”或“异常”的数据、量表、评价图等资料即为实用的标准。

上述的同一总体人群是指具有相同特征(如种族、性别、年龄等)的同质个体的集合。被选取的样本称为参照人群,参照人群的选择根据研究目的可以在该同一总体中随机抽样而得(如用于制定现状标准),也可以人为指定(如用于制定理想标准)。

标准的制定受主客观因素的影响。客观因素主要是指标准的制定需建立在相关领域大量客观科学研究的基础上;主观因素主要是指不仅在这些客观科学研究中,而且在标准制定过程中(如研究的目的、原则、对象、仪器、抽样方法、统计分析方法等的选择),由于对这些客观科学研究(或标准)的主观认识不同,不可避免均有人为因素参与。

主客观因素是相互影响的。客观科学是主观认识的基础,它可以影响研究者乃至人群对健康概念等的认识,而研究者(或人群)对客观科学研究的理解所产生的主观认识又反过来会对进一步的客观科学研究产生影响。因此,不同时期的客观科学的水平以及相应时期的社会、经济发展水平等因素对研究者所产生的有关健康的主观认识均是不同的。二者相互影响的结果决定了标准的制定及其应用评价均是一个建立在客观的科学依据基础上的、适应当时的科学技术、经济、社会发展水平的一个人造的、暂时的和相对的过程。

随着科技的发展,影响标准制定的主客观因素也在不断改变,导致标准在概念上也将不断地更新,具体值也将不断地进行修订,标准的使用年限也一般限定在若干年内。

1.2 标准制定的目的

标准制定及其应用评价过程的最终目的是要通过标准来在医学上正确地、客观地评价个体/群体某种生理状态是否“正常”及其程度(如是否“充分”),并直观地反映出来,再据此决定是否应进一步调查研究其“正常”是否“合理”或造成其“正常”或“异常”生理状态的生活环境影响因素,以及是否应采取相

应的措施来积极进行相应的预防和治疗干预,以促进或维持其生理状态达到最佳状态,达到促进健康的最终目的。

1.3 新医学模式对标准制定的影响——现状标准与理想标准

但何为“正常”?又如何通过标准来界定“正常”呢?

实际上,何为“正常”并无一个特定的定义。在不同的生理状态下有不同的定义。以往“正常”的概念主要体现为“健康”,即“无病就是健康”。但随着社会、经济发展、科技进步和生活条件的极大改善,人口老龄化进程加速以及疾病谱从传染病向慢性非传染性疾病转变。人们对健康的认识也发生了很大变化,已不满足于仅仅是有病就医,而是健康长寿。WHO 已提出:“健康是身体上、精神上和社会适应上的完好状态,而不仅仅是没有疾病和虚弱”。这标志着“无病就是健康”的观念业已结束,也标志着“正常”的含义更加广泛,即要达到“完好状态”的健康。这种“完好状态”的健康也可理解为是一种“合理”的“最优”的“正常”状态,是人们需要追求的目标。

“正常”概念的转变也促使医学模式从注重临床治疗的传统生物医学模式转变为有更广泛含义的、不仅注重临床治疗而且更注重预防的、新的生物-心理-社会医学模式。这进而也必然影响标准的制定,使标准的制定要更多地考虑预防的要求。表现为从制定现状标准逐渐转变为制定理想标准。

现状标准是在某一总体人群中随机抽取一定数量的“无病”个体作为“正常”参照人群而制定的标准。

理想标准是在某一总体人群中有目的地选择整体生理状态“最优”且“合理”的一定数量个体作为“正常”参照人群而制定的标准。

这种“最优”且“合理”的生理状态也称为“理想水平”。“理想水平”是个体或群体经过后天环境因素的干预后能达到也应达到的水平。这些环境因素包括合理的营养、合理的体育运动以及合理的社会、医疗保障等,而且为达到该“理想水平”的干预条件也是在目前的社会、经济水平下已部分达到且在将来能更普及的条件,是可行的。当然,随着社会、经济等条件及研究水平不断的发展,将来这种“理想水平”的定义还会有所调整。

应当指出的是,这里的“理想水平”不一定是在某些特殊的生理和/或环境条件下该种生理状态所能达到的最大值(极端值),因为能达到这种最大值

的某些特殊的生理和/或环境条件对机体正常的整体生理状态反而可能有害,是一些一般人群较少或没有必要达到的条件。这种可能不合理的最大值不一定有必要去追求。

因此,“理想水平”具有“最优性”、“合理性”和“可实现性”的特点,是一种有条件的、相对的“最佳状态”。

由此可见,由于现状标准对参照人群未设定较严格的条件,其参照的“正常”人群实际所具备的水平在多数情况下有低于“理想水平”的趋势。因而现状标准用于评价个体的结果主要是反映其在所抽样群体中所处的位置,是一种相对的评价。

很显然,如果采用水平可能偏低的现状标准来进行评价,就可能导致被评价个体或群体的生理状态被高估(随标准偏低的程度而异),例如将“正常”高估为“良好”,甚至可能将疾病状态高估为“正常”。由于在一般医务人员和被检查者的概念中,“正常”即“健康或达标”、“良好”即“已”、“不错”或“很棒”。以致会产生错觉,造成医务人员不能及时深入调查和/或提出相应建议;而不懂医学的被检查者也可能觉得自己的某种生理状态并不理想(如营养较差、运动较少等),但因为医生的结论是“正常”或更好的情况,而因此放松对疾病防治的重视。从预防医学的角度而言,这种偏低的现状标准对促进个体或群体健康的意义显然是较为消极和有限的。再考虑到通常一个标准要使用数年或更长,而目前的生活条件和健康水平在不断地改善和提高,对疾病的防治重点也从重治疗转向重预防,则偏低的、滞后的现状标准的消极作用会更大。

而理想标准对参照人群的要求更严格,能更好地判断个体生理状态与“理想水平”的差距,在理论上能够最大限度地降低将生理状态不理想的个体(或人群)高估的可能性。其使用能更有助于达到标准制定所应达到的目的,是符合新医学模式的标准。也有人将这种代表“理想水平”的理想标准所作的评价称为绝对评价。

需要强调的是,“理想标准”是一种可以达到也应当达到的“高水平”的“标准”,是个体或群体需要追求的目标。但采用这种“高水平”的“标准”并不意味着要求所有的人群在短期内要达到该标准的“理想水平”。由于社会、文化、经济等多种因素的影响,不少地区的人群在目前甚至可能在一个较长的时期内都达不到相应时期的“理想水平”。但正因为达不到,才提示在该人群中还存在有环境等影响因素,需

要进一步的研究分析或根据具体情况立即作相应地干预或制定措施为将来的干预做准备。

因此,这种“高水平”的“标准”除了可以进行临床诊断外,还具有更积极的预防作用,即可以不断地“提醒”那些达不到“理想水平”的人群作相应地改进。所以滥用“本地区达不到”标准的理由来制定单独的、符合“本地区情况”(或“国情”)的标准就容易掩盖那些可变因素被研究发现并被改进的可能,不利于疾病的预防和治疗。

因而,在一些尚不够了解的医学领域的标准制定中,现状标准作为暂时的评价体系仍然有其存在的依据,但是随着科技的进步均应逐渐以理想标准取而代之。

2 骨骼发育标准的制定

2.1 骨质疏松症的定义及制定骨骼发育“理想标准”的意义

原发性骨质疏松症是以骨量减少(osteopenia)、骨组织显微结构退化(松质骨骨小梁变细、断裂、数量减少;皮质骨多孔、变薄)为特征,以致骨的脆性增高及骨折危险性增加的一种全身骨病^[1]。从该定义可见,骨质疏松症的发生是一个渐进而长期的过程,而骨折只是其最严重的后果之一。

理论上,个体的骨骼发育在达到骨密度(bone mineral density, BMD)高峰后(峰值骨密度, peak bone mass, PBM)的退化(密度下降)即骨质疏松症的发生最终是不可避免的,但骨质疏松症研究的最终目的显然不是在严重骨质疏松症发生或将要发生时才采取措施,因为此时治疗效果较差。

从理论上推论,假设骨密度高峰年龄为 35 岁(此时骨密度均值为 PBM_{35} , 标准差为 SD_{35}), 此时同时有甲、乙两个体, 甲个体的骨密度值正好在 PBM_{35} 均值水平, 而乙个体的骨密度值正好在 $PBM_{35} - 1SD_{35}$ 的水平, 那么在不改变甲、乙两个体生活条件的情况下, 甲、乙两个体的骨密度可能会一直分别按其以后各年龄的骨密度均值水平的趋势发展, 即将来在某年龄如 60 岁时, 甲个体的骨密度水平可能仍然会在 60 岁时骨密度的均值水平(此时骨密度均值为 M_{60} , 标准差为 SD_{60} , 以下类推); 同理, 乙个体的骨密度水平也可能仍然会在 60 岁时骨密度的 $M_{60} - 1SD_{60}$ 水平, 即乙个体可能会较甲个体要更早地发生骨质疏松症。但如果乙个体在 35 岁时及时进行干预, 其骨密度水平就有可能恢复至 PBM_{35} 以后各年

龄的均值水平甚至更高(如在 40 岁时即已恢复至均值 M_{40} 甚至更高),将来在 60 岁时,其骨密度水平就有可能早已恢复至 M_{60} 甚至更高,而不是保持在 $M_{60} - 1 SD_{60}$ 。此时乙个体发生骨质疏松症的概率可能与甲个体一样甚至更低。实际上,这种推论正是骨质疏松这种慢性疾病需要重视及早预防的基础。

WHO 推荐的骨质疏松诊断标准的依据主要是所选地区人种骨密度正常分布数据^[2],我国目前也按照这种依据来进行标准的制定研究。由于这种未按规定条件的“正常分布”所得到的标准实际上也是一种现状标准,因此,目前以临床治疗为主要目的而制定的骨质疏松症诊断标准的含义显然已经较狭窄,未能完全符合新医学模式的要求。因而有必要进一步扩展骨质疏松症诊断标准的含义,将其扩展为“? ~ ? 岁中国人骨骼发育(或狭义的骨密度)正常值(理想值)及其异常范围(本含义仅为笔者建议,更好的含义名称及年龄范围的确定应由全国相关的研究同仁确定)”。这个含义更广泛的“高水平”的“理想标准”对于骨质疏松症的防治十分必要。它不仅符合新医学模式的要求,能满足制定标准的目的,即促进或维持个体/群体的骨发育达到“理想水平”的最佳状态及延缓高峰以后骨骼退化的过程,延缓骨质疏松发生和/或不发生严重骨质疏松,而且对于进一步丰富和完善骨骼发育及标准的相关理论也有重要意义。

因此,在目前对骨质疏松已有较多了解、各种条件已有较大改善及医学模式转变的情况下,不仅已有制定骨骼发育“理想标准”的必要,而且也逐渐具备了制定的可能。

2.2 制定骨骼发育标准的有关概念

2.2.1 不同种族间的遗传差异与标准的制定

:与在生命特征等表象上及在疾病发生上均存在种族间的遗传差异一样,骨骼发育(包括退化,下同)在潜力大小、发育方式以及在骨质量上均存在种族间的遗传差异。

大量的研究已表明,与骨骼发育有关的体型、身高、躯干和四肢长度的比例存在种族差异,骨龄研究证实种族间的骨骼发育方式不同,如亚洲儿童的骨龄在青春期前一直落后于非洲裔和欧洲裔美国儿童,但青春期阶段骨的干骺愈合速度显著超过后者^[3],骨密度测定也表明,黑种人 > 白种人 > 黄种人,还有研究表明,BMD 的遗传力为 0.6 ~ 0.8,不同人种的骨量存在显著差别,骨折发生率以白种人为最高,亚洲人种居中,以黑种人为最低,对一些影响

骨代谢的遗传基因多态性研究也均表明不同种族人群在钙吸收、各年龄骨密度及骨质疏松发生的不同可能与基因差异有关。

因此,由于种族间骨骼发育存在着遗传差异,此时如果将某一种族的骨骼发育标准作为统一标准来评价所有不同种族的骨骼发育水平,显然其结果会有误差,结论也容易对疾病的防治或相关研究产生误导。如果该标准在制定过程中再存在错误的话,则问题更严重。

所以,用本种族人群作为参照人群所制定的标准用以判断本种族人群的骨骼发育水平或状态才能得到最客观、科学和准确的结果。

需指出的是,强调标准的本种族化并不是否定有关研究与国际的“接轨”。但笔者认为和国际“接轨”的应主要是研究方法,如某骨密度的研究方法或标准制定方法就可以是国际上统一的。至于在统计学的理论及方法上,可能是统一的(如标准差法或百分位法),也可能是不统一的(如具体采用几个标准差和多少百分位,这也是目前争论较大之处),但无论是否统一,均需通过理论的合理推导及实际应用的检验才能确定。

如果在有关研究的早期,国内外在研究方法 & 疾病发生的了解均不完善时,采用某个相对较好的标准作为诊断标准或研究参考还说得过去的话,那么在研究方法日益成熟及疾病的发生有更多了解的今天,仍采用国际统一的标准,这无论是从理论研究还是实际防治的角度,显然是欠妥当的。

2.2.2 同一种族的生活环境差异与标准的制定

:需要强调的是,遗传背景相同的种族人群在不同生活环境间(如国家间、行政区域间、山区与平原间、城市与乡村间以及家庭间等)所体现出的骨骼发育水平的差异(如某地区的骨质疏松症发生率较高和/或发生年龄较早),并不表明该种族的骨骼发育方式在遗传上发生了什么变化,而是体现了各种后天环境因素的影响。

因此,采用统一的“理想标准”评价遗传背景相同、但生活环境不同的个体或人群的意义不仅表现在可以了解这些个体或群体与“理想水平”间的差距,而且通过研究造成这些差距以及个体或群体间的差异的原因可进一步了解环境因素对骨骼发育的影响,进而有针对性地进行相应的预防和治疗。

所以,为了体现这一意义,理论上,遗传背景相同的种族人群只需制定一个统一的标准即可。

2.2.3 我国各民族的遗传差异与标准的制定

(1) 汉族的遗传差异与标准的制定:

理论上,如果汉族的遗传背景完全相同,则只需制定一个统一的标准即可。

但我国以往的研究表明在生长发育水平上北方地区(通常的概念是新疆、青海、甘肃、内蒙古、宁夏、陕西、山西、河北、北京、天津、黑龙江、吉林、辽宁、河南、山东这 15 个省(市),其余为南方地区)在整体上要高于南方,北方人体形高大,南方人则较瘦小。近年来大量的遗传学研究^[4]表明,汉族确实存在遗传背景有明显差异的南北两大人群,不过却表明两大人群的地域划分应大致以北纬 30 度即长江附近为界。如考虑具体的行政省(市)划分,从遗传学^[4]、体质形态学^[5]以及能影响人群间大规模相互迁移融合的因素之一的地理学综合来看,安徽、江苏、上海等三个以往被认为是南方地区的省(市)明显应划归北方汉族省(市)。这样划分后可见在以往的与骨骼发育有关的儿童青少年身高发育研究中,无论是北方还是南方汉族省(市)中,身高发育水平较高的多数是经济较发达的地区。但南方汉族省(市)中无论经济水平如何,人群最高的身高发育水平也大约只与北方汉族省(市)中人群最低的相当(本文仅列出 2000 年的体质调研结果,其他的调查结果未列出,但趋势基本一致。见图 1、图 2)^[6]。提示除了经济等后天环境因素可影响身高发育外,南北方汉族的骨骼发育在遗传上也可能有较明显的差异。

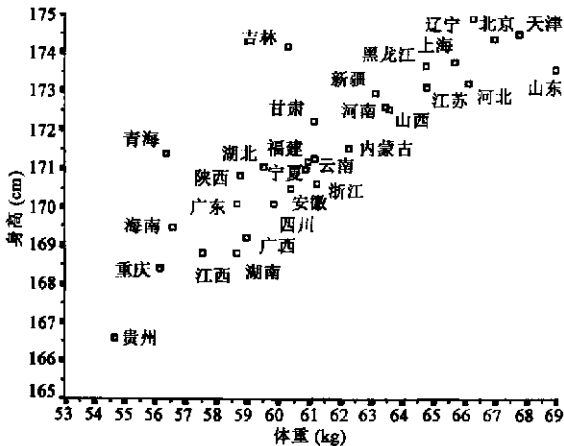


图 1 2000 年中国 30 个省(市)城市 18 岁汉族男青年身高、体重分布图

同时,国内的有关研究可能从流行病学角度给我们一些提示。李宁华等^[7]的研究提示北京市和上海市的骨质疏松症发生率有低于吉林省的趋势应可被视为是在同一遗传背景下经济等因素的影响;

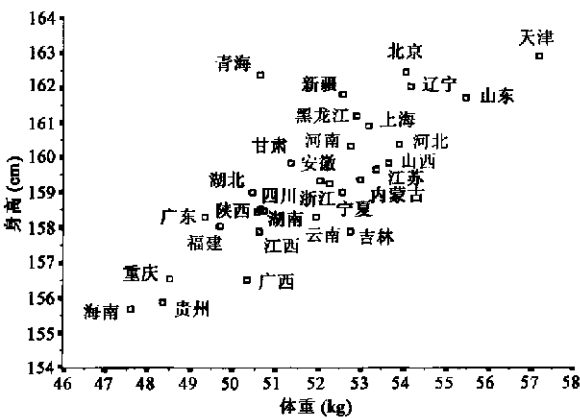


图 2 2000 年中国 30 个省(市)城市 18 岁汉族女青年身高、体重分布图

而广东省和四川省的发生率均低于上述 3 省(市)则可能与两大汉族人群的遗传背景差异有关,因为广东的经济、体育和环境方面与吉林的差距应不会像其骨质疏松症发生率那样高;至于四川的发生率最高,则可能与遗传和经济两方面的因素均有作用有关。

因此,尽管南方北方汉族这两大人群的遗传背景差异不如汉族与其他人种那么大,但是否应分别制定标准也可能应作考虑,否则用北方汉族“可能”的“高标准”来评价南方汉族其结果可能会存在一定的误差,反之亦然。当然实际有无必要还需要作进一步的研究尤其是与骨质发育有关的有关遗传基因在南北方人群中差异的研究并结合实际情况才能确定。

(2) 少数民族的遗传差异与标准的制定:

如果我国的各少数民族人群的遗传背景不同,理论上则需要制定各自的统一标准。

但是从遗传学^[4]、体质形态学^[5]来看,除维吾尔族、哈萨克族等几个少数民族均与南北方汉族有较大差异外,其他少数民族与相邻的南方(或北方)汉族间的差异均远小于南北方汉族间的差异,表明这些少数民族在骨骼发育潜力上与相邻的南方(或北方)汉族有相似甚至相同之处,同时少数民族较少有同一民族分南北两个遗传背景不同人群的情况,因而少数民族可适用于其所在地区的汉族标准。如蒙古族、回族、朝鲜族、藏族等可适用于北方汉族标准;壮族、黎族、苗族、侗族等可适用于南方汉族标准;而维吾尔族、哈萨克族等几个北方少数民族由于其相互间的遗传差异较小,可考虑制定一个统一的标准。不过考虑到这几个少数民族虽然在遗传上与白种人

更接近,但其经济条件等和生长发育水平均较低(相对于北方汉族人群而言),其所制定的单独的标准可能会有偏低的可能,是否应当暂时采用全国或北方汉族人群的标准则有待于进一步的研究探讨。

很显然,无论是制定我国统一的标准还是分别制定南北方的标准,即使是加上维吾尔族、哈萨克族等几个遗传背景基本相同的少数民族可能单独制定的标准,我国的人群在理论上均只需要 1~3 个标准即可。

因此,我国目前不少环境不同的地区的研究因为其研究对象的骨密度与其他地区有差异而考虑制定其“本地区”(或“本民族”)的“标准”(或“正常参考值”,下同)就很容易掩盖骨骼发育的环境因素的影响,不利于研究探索造成这些差异的原因。至于在目前骨密度测量中采用同一种方法(如 DEXA)但厂家不同而使其测量仪器内置的标准不同(主要来自国外)的情况是不合理的。

2.2.4 统计学理论与标准的制定:

(1) 参照人群的选择:

由于理想标准代表的是骨骼发育“理想”的“理想水平”,显然其“理想”参照人群也必须是来自于骨骼发育“理想”且“合理”的人群。因而首先在国内筛选出骨骼发育“理想”且“合理”的人群非常重要。

而要初步筛选出骨骼发育“理想”且“合理”的人群,理论上不外乎有在全国进行普查和抽样调查两种方法。普查由于人力、物力投入巨大,基本不具有实际操作可行性。而抽样调查的主要问题是抽样范围及人数,即抽样范围越大,样本代表性相对越好,但投入的人力、物力越大。抽样范围一般概念上常见的是各大地区(主要以行政区域划分,如华北等)或各省(市)(主要以省会城市为主)。但就短期内制定标准而言,在现有条件下,即使是做一种方法(如 DEXA)的标准,如果既要严格按照统计学的要求,又要保证较低的人力、物力投入,则在各省(市)的 32 个省会城市进行抽样调查的可行性很差,而只在各大行政区域进行抽样则代表性可能不理想。

那么有没有既能保证有较高的质量、又能较快地、同时还能使投入的人力、物力尽量最少的制定骨骼发育标准的方法呢?笔者试图通过同样与骨骼发育有关的儿童青少年身高发育情况进行探讨。

大量研究已表明,合理的营养、合理的体育运动和健康的生活方式等措施,不仅能提高儿童青少年的身高和体质水平,还能提高峰值骨密度、促进长期的骨骼健康和降低老年期发生骨质疏松的危险。即

骨质疏松作为一种可以终生预防的疾病,在儿童青少年时期的初级预防也特别重要。

理论上,骨骼发育是遗传因素和后天环境因素综合作用的结果。因此,在遗传背景相同的情况下,作为生活环境中各种因素综合作用的结果,如果某一人群的骨骼发育水平“最高”且能证明该“最高”的骨骼发育水平不是主要由一些非正常因素所导致,则已间接表明其影响因素中的有利因素可能较不利因素要多,进而可初步认为该人群的骨骼发育是“理想”且“合理”的。这种“理想”且“合理”的骨骼发育不仅有利于身高发育,而且有利于骨密度的提高。

进一步的推理可以认为,如果通过研究在全国找出某个人群的儿童青少年的身高发育水平“理想”且“合理”,那么该人群中的成年人的骨密度水平也可能是“理想”且“合理”的。可以建议在该人群中筛选“理想”参照人群。这时再在该人群中人为设定统一的“合理”条件进一步筛选“理想”参照人群,使“理想”参照人群更加“合理”。

这种得到骨骼发育“理想”且“合理”的人群以及进一步“优中选优”得到“理想”参照人群的方法,不仅在理论上而且在实际制定中都具有一定的合理性、客观性、科学性和可行性。因为该初步筛选的过程可利用现有的全国儿童青少年的历年监测数据,大大节省人力、物力,同时,由于在一个人群中集中了有限的人力、物力等资源,不仅可快速、高效地制定出相关的骨骼发育标准,还可以使所制定的骨骼发育标准在统计学、质量控制等方面达到高水平,成为一个高质量的标准。

而就全国汉族儿童青少年的身高、体重发育水平来看,一般是城市高于其邻近的农村,在城市中从近几年的资料来看又以北京、天津、辽宁、山东等环渤海省(市)的男女儿童青少年相对较高(见图 1、图 2)。同时上述环渤海省(市)无论是在过去还是在现在,基本上均是我国的经济较发达的地区,也基本上是体育大省。众所周知,较好的营养和生活环境、较广泛的群众性全民健身等有利于骨骼发育的条件/因素与经济是密切相关的,而且这些条件随着近年来经济的进一步发展还在不断改善。这些有利的因素都使得这几个省(市)有条件成为骨骼发育是“理想”且“合理”的人群所在地。前述李宁华等^[7]的研究中北京、上海、广东、四川等省(市)所表现出的骨质疏松发生率的差异也间接支持了以上推理。

但从人力、物力上以及理论上只需要在一个地区的人群来筛选“理想”参照人群即可的角度而言,

还需要在上述省(市)中筛选出一个省(市)的城市来作为“最适宜”地区。笔者建议将北京作为“最适宜”地区比较合适。因为北京基本上满足上述作为“最适宜”地区的各种条件,具有较好的代表性。

当然,问题的关键在于身高虽然是一个能综合反映骨骼发育且与成年后的骨骼发育有关联的指标,但其并不绝对与成年后的骨密度成正比,也不意味着作为“最适宜”地区的人群所处环境的各项影响因素均是合理(或最佳)的。此时,不能推理具有有利于骨骼发育的一些因素,而经济、身高发育水平低于上述省(市)的地区(如有较高牛奶摄入量)的某些人群的骨密度就不高,也不能推理一些运动量更大,而经济、身高发育水平低于城市的农村人群的骨密度就不高。但正因为这些人群的身高与骨密度水平可能不成正比,说明就综合而言其有利因素可能不如北京多,不利因素则反之。而李宁华等^[7]、区品中等^[8]、安珍等^[9]的研究中所见到的农村男女性人群的骨质疏松症发生率均有高于城市的趋势也从侧面说明了这点。

因此,要使得这些被筛选出来的“理想”参照人群及其骨密度能够有尽量好的代表性,其关键在于如何为“优中选优”设定有代表性的筛选条件。这里的为“优中选优”所设定的筛选条件应将一些已知的、主要的、能影响骨骼发育的正常因素(有利因素,在此可理解为入选条件)和非正常因素(主要指不利因素,但也可能包括部分虽有利于骨骼发育但对机体的整体生理状态有害的、不容易达到也不一定有必要达到的因素,如高强度活动量、不合理的高钙摄入等。在此可理解为排除条件)包括进来。

不过由于在实际生活中很难完全避免非正常因素的影响,因而设定“优中选优”的筛选条件主要目的是为了尽量减少非正常因素对骨骼发育的影响,使得骨骼发育在尽量“合理”的条件下能得到“最大限度”的促进,所筛选的“理想”参照人群更加“合理”、更加有“代表性”。依此推理,这里的“理想”骨骼发育水平也是在“合理”条件下的“最高”水平,不一定是在某些特殊条件下所能达到的最高水平。

从发展的角度而言,无论是正常还是非正常因素,其条件水平的设定应是目前已部分达到而将来更完善的水平,需有一定的超前性。例如有利因素牛奶(或钙)的摄入水平应考虑以理想膳食中和健康角度综合考虑后的一个较高的“理想水平”,而不能以目前较低的实际摄入量为筛选水平。即如果经研究证实牛奶对健康有利的“合理”摄入量在约 250 g/

天,且该量在目前已可达到并在将来会更普及,则无论目前的摄入量是多还是少,仍然应考虑用约 250 g/天。其他条件的设定原则也是如此。

因此,由于这种设定的筛选条件实际上可能已经包括了各省(市)已知的、主要的、可能影响骨骼发育的因素(尤其是有利因素),所以在各省(市)进行普查或抽样调查的必要性相对而言实际上也就进一步降低了。当然,这只是在希望尽快制定骨骼发育标准的前提下而言的,并不能否定在多地区进行相关调查的重要性。但本文作者建议这种全国性的大规模抽样调查研究不如放在将来标准制定后再陆续开展,以用于标准的修订和相关理论的丰富。

(2)标准评价方法的选择——标准差法或百分率法的选择:

目前国内外骨质疏松症诊断标准采用的评价方法主要有标准差法和百分率法两种。采用哪种方法较好呢?

从统计学而言,如果同一人群的骨骼发育水平数据的分布不是偏态的,那么其某 SD 下所占的面积与相应的百分位的意义是一样的,如 +1.96 SD 下的面积与指定第 97.5 百分位下的面积是相同的,即采用哪一种都是等效的,如果该数据虽然是偏态的,但经数据转换后仍能呈正态分布,则仍可用两种方法,如果未经转换和不便转换,则以采用百分位法或采用其他更合适的方法为宜。另外百分位法在应用中要更直观,容易为一般人理解些(尤其在制成百分位图形式时)。因此,在将来统一的标准中,只要标准所用的数据符合标准差法的要求,两种方法均应采用,以满足科研和应用的需要。

由此可见,全国无论是采用一个统一的还是南北方各采用一个标准,其统计学上的意义表现在,由于采用了统一的标准值和标准差(或等效的统一的百分位,下同)就避免了目前各地在探讨制定自己的“标准值”时,“标准值”代表性较差,理论上就不适合作为“标准”的弊病,同时也避免了由于“标准值”及其“标准差”均不同以至于采用这些“标准”所评价的结果不能够相互间进行比较的弊病。

因此,无论是何地区、采用何厂家的机型,只要是同一类型(如 DEXA)的、且经过了国家统一校正的,那么其测量值均可采用该统一的标准进行判断,其有关研究中的某“发生率”就可进行相互间的比较。至于其他一些调查“本地正常参考值”的研究所得到的骨骼发育值及其标准差仅仅意味着该调查所涉及人群的水平(或本底值)及离散程度,其结果主

要是用作探讨有关影响因素及标准将来的修订参考。

(3) 标准差或百分位的确定原则:

与一些生命现象中“正常”与“疾病”的界限较明显(如一些急性传染性疾病)不同,骨质疏松等慢性非传染性疾病作为一类病因复杂、起病隐匿、随年龄增长逐渐的或进行性的器官功能降低且长期存在的、治疗效果相对较差、具有健康损害和社会危害严重等特点的疾病状态,其“正常”与“疾病”间往往缺乏明显的界限。由于对“疾病”(此处为狭义的“疾病”,仅指已出现症状或已处于发病条件但未出现症状的个体,下同)的诊断范围通过研究确定相对较容易,但对接近发病条件却无症状或在该诊断范围之外的个体如果均诊断为“正常”显然对疾病的防治不利。

因此,对于骨质疏松等慢性非传染性疾病的标准而言,标准所采用的标准差倍数(或等效的百分位,下同)的大小应满足的要求是:除了能界定“疾病”以及“最优”而“合理”的“正常”范围外,还要合理地划分出介于“正常”与“疾病”间的“亚疾病”(亚健康状态)范围,以利于骨质疏松等慢性非传染性疾病的提前预防和治疗。同时,对该三个范围的“诊断”还要能达到尽量小的误诊率和漏诊率。

目前在标准和相关研究中的标准差倍数常采用 1 SD、2 SD、2.5 SD 或 3 SD。选择这些倍数的标准差一方面是为了反映大多数生命现象的发育规律,即“正常”个体占多数,“异常”个体占少数;另一方面在对某种生命现象的性质、比例、范围等尚不明了时,人为确定这些简化的、常见倍数的标准差就能以一种易理解、易计算的方式初步反映其规律,便于研究间的相互比较及进一步的研究。

但显然这些整数倍的标准差不可能准确反映所有生命现象的规律,同时,做研究的目的不仅要深入了解疾病的各种规律,而且要将这些规律以简便的方式进行应用。具体到标准而言,标准的制定实际上也是一个研究的过程,在对某种生命现象的性质、比例、范围等已有较多了解后,就应当在客观科学依据的基础上,使其反映疾病规律的各项指标能够尽量的精确,以利于疾病的防治;不过在其表达形式上,即最后呈现给临床的一般专业人员或普通大众等非专业人员时则应简单易懂,而不是在研究阶段就过分简化。因此,在科学研究的基础上,如果有必要,对标准差倍数应进行相应地调整,有利于准确地反映个体的生理状况,满足预防和治疗的需要。百

分率的确定也是如此。

(4) 与标准差或百分位的确定相关的基础值:

以下就骨骼发育标准中与“正常”、“疾病”和“亚疾病”这三种状况的标准差倍数确定相关的基础值进行探讨。

为推论方便,笔者在文献[10]中选择年龄划分较细的女性腰椎(L_{1-4})的数据采用统计软件进行绘图。文献[10]中表4的数据绘制的曲线最光滑,同时考虑到该数据与目前多数研究的结果中 $PBM \approx 1.100 \text{ g/cm}^2$, $SD \approx 0.100 \text{ g/cm}^2$ 、PBM 年龄 $\approx 30 \sim 35$ 岁接近。因而选取该曲线作为以下骨骼发育“理想标准”中的三种状态值的示意图(见图3)。

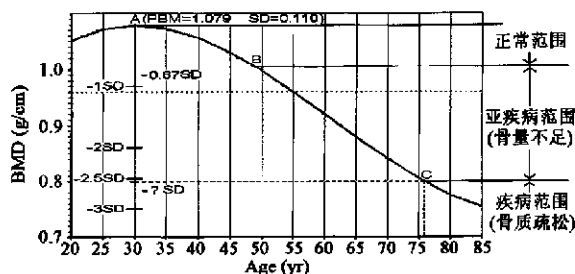


图3 骨骼发育理想标准的三类值示意图(百分位法以此类推)

①“理想人群”各年龄骨骼发育值(“理想人群值”)、峰值骨密度值及发生年龄的确定

在现有研究基础上确定“理想人群”的条件并进行抽样调查可得到各年龄组的骨密度值($M_{理想}$, $SD_{理想}$)、峰值骨密度值($PBM_{理想}$, $SD_{PBM_{理想}}$, 如图3中A点)及发生年龄。“理想人群值”是确定“理想标准”的基础。

②“疾病值”范围的确定

“疾病值”是指对骨质疏松的临床诊断值范围。在此范围内个体发生骨痛和/或骨折(非暴力性、骨质疏松性骨折,下同)症状的“可能性”(“危险性”)较高。该诊断值范围的界值主要通过研究“理想人群”中已发生症状个体的骨密度值($M_{理想症状}$, $SD_{理想症状}$)以及结合这些个体在各年龄的发生概率来确定。

这里“疾病值”应从“理想人群”中获得的原因是因为在理论上,一方面“理想人群”骨骼的退化也是不可避免的,终会退化到可能出现症状的骨密度水平;另一方面,“理想人群值”是从“理想人群”中获得的,因而其“疾病值”也应当从“理想人群”中获得。

需要注意的是,在理论上,由于骨骼只有退化到一定的水平才可能出现症状,因而“一般人群”在出现症状时的骨密度值($M_{一般症状}$, $SD_{一般症状}$)应当是与

“理想人群”相等或接近的,但这两种人群出现症状时的骨密度水平在标准制定中的意义是不同的。即在理论上,由于“一般人群”的骨骼退化可能要比“理想人群”快一些,因而在同样达到可能出现症状的骨密度条件下,“一般人群”出现症状的年龄可能也会早一些,进一步表现为在同年龄时出现症状的概率也要高于“理想人群”。例如假设“一般人群”在65岁时骨骼就已退化到可能出现症状的水平,而“理想人群”则可能要在70岁后才退化到该水平,则同样在70岁时,“一般人群”出现症状的概率可能高于“理想人群”。

正是由于这种意义上的差异,“一般人群”在出现症状时的骨密度值及在各年龄的发生概率只能作为“理想标准”中“理想人群”的“疾病值”制定的参考或临时的过渡,而不能用来代替“理想人群”的“疾病值”。

但即使在得到“理想人群”的已出现症状的骨密度值($M_{\text{理想症状}}, SD_{\text{理想症状}}$)后,不宜直接用该骨密度值作为“疾病值”上限界值(如图3示意图中的C点)。因为理论上在该均数上尚有约50%的已出现症状个体未被包括进来,即漏诊率可能较高;也不宜取 $M_{\text{理想症状}} + 2 SD_{\text{理想症状}}$ (即至少包括95%以上的骨折个体)或更高的值以将大部分的骨折个体都包括进来。这样虽然可大幅降低漏诊率,但又可能较大幅度提高误诊率。

例如,在目前较少的有关一般人群骨折时的骨密度测定(DEXA法)研究中,可见其腰椎(L_{2-4})骨密度值($M_{\text{一般骨折}}$)波动在 $0.600 \sim 0.800 (g/cm^2)$ 间,而假设平均取 $SD_{\text{一般骨折}} = 0.100 (g/cm^2)$ 时,如果取 $M_{\text{一般骨折}} + 2 SD_{\text{一般骨折}}$,则其上限值约在 $0.800 \sim 1.000 (g/cm^2)$ 间。显然,从图3可见,此时 $0.800 (g/cm^2)$ 的值还可接受,而 $1.000 (g/cm^2)$ 则不仅已包括了 $PBM_{\text{理想}} - 1 SD_{\text{理想}} \sim PBM_{\text{理想}} - 2 SD_{\text{理想}}$ 的“骨量不足”范围,且涵盖了相当部分“正常”的范围。这显然不妥。而如果取中间值 $0.700 (g/cm^2)$ 时为 $0.900 (g/cm^2)$ 是否合适则尚待探讨。但从其大约相当于65岁时的骨密度水平来看,似乎偏高。

研究“理想人群”中已出现症状的人群在各年龄的发生概率可以判断其危险度。其意义是提示在何年龄组的危险度开始值得在临床上关注,进而提示该年龄组是否可作为漏诊率和误诊率间的合适的平衡点年龄,为进一步确定“疾病值”的 $SD_{\text{理想症状}}$ 倍数提供依据。

目前的WHO推荐的骨质疏松诊断标准也是依据生存期(life-time)骨折危险性的数据而定。但国内还尚未见到有关的生存期骨折危险性的报道^[2]。不过需要注意的是,如果只研究发生骨折而不研究出现骨痛症状的人群的骨密度值以及在各年龄的发生概率容易缩小“疾病值”的范围。

值得注意的是,无论“理想标准”中的“疾病值”最终定在何水平,都只是表明在该水平下容易出现症状,是处于“疾病”状态,无论何年龄的中老年个体,只要其骨密度低于“疾病值”的水平,就可诊断为“骨质疏松”。此时在“理想人群”的“理想值”曲线上,等于或低于该“疾病值”的年龄只是表明“理想人群”在该年龄及以后发生“疾病”的可能性较大,是“高发年龄”。而不应理解为只能在该年龄及以后的个体才能作“疾病”诊断。但可以认为“该个体的骨密度”相当于“该岁数或以后的水平”。

③“理想正常值”范围的确定

“理想正常值”是以“理想值”为基础的一个“合理”的“正常”范围值。在该值范围内,理论上应无患病的危险性或危险性极小。

但什么是骨骼发育“正常”而“合理”,目前还不能用某个骨密度水平来表达其确切范围。不过由于在理论上“理想正常值”范围内无患病的危险性或危险性极小,因而通过研究所得到的“理想人群”在各年龄的无或极少出现症状的发生概率以及其相应的骨密度值($M_{\text{理想正常}}, SD_{\text{理想正常}}$)可作为进一步确定“理想正常值”范围的骨密度值下限的基础。但显然与“疾病值”一样,“理想正常值”可能不宜直接用其均值或 $M_{\text{理想正常}} - 1 SD_{\text{理想正常}}$ (或 $2 SD_{\text{理想正常}}$)来作为其下限的确定依据。

④“亚疾病值”(骨量减少)范围的确定

“亚疾病值”是界于“理想正常值”和“疾病值”间的一个范围值。是“正常”个体与“疾病”个体的过渡阶段。在该值范围内,理论上应理解为已有发生疾病的可能。但与“疾病值”相比,一方面患病的危险性要低,另一方面该范围的个体的骨密度回复正常的可能性(可逆性)也要更大一些。正是由于这种危险性较低以及可逆性的存在,“亚疾病值”在预防上可作为“提醒”人群作预防性干预或早期治疗的范围值。

⑤“理想正常值”、“疾病值”及“亚疾病值”的确定原则建议

从预防及临床医学的角度而言,“理想标准”要能对“正常”、“疾病”及“亚疾病”三个范围的“诊断”

达到尽量小的误诊率和漏诊率 ,以利于对骨质疏松进行尽早的预防 ,及时的治疗。但这种误诊率和漏诊率之间的平衡取决于对骨骼发育的各种状态在预防和治理上的不同认识。笔者建议在这三个值的漏诊率和误诊率的平衡上采取即使适当增加误诊也不希望漏诊的确定原则。

由于研究表明 ,骨质疏松尤其是严重骨质疏松的治疗效果已不是特别理想 ,此时“ 疾病值 ”的确定以及划分轻度、中度、严重等的等级主要是在临床上反映其病情的严重程度 ,在预防及治理上的意义已相对较小。但鉴于骨质疏松尤其是严重骨质疏松的骨折后果可能会较严重地影响其生活质量 ,因而适当增加误诊以尽早地诊断骨质疏松有利于“ 提醒 ”已有可能或已经出现症状的“ 患者 ”以降低和防止骨折发生的危险性。

在“ 疾病值 ”范围确定后 ,较窄的“ 亚疾病值 ”范围对于提前进行疾病的预防、治理应当也不是最理想的。因而适当增加误诊以适当扩大“ 亚疾病值 ”的范围并根据具体情况划分 2 ~ 3 个等级以表示其骨量不足的程度有利于提前进行预防性干预以防止或减缓骨质疏松的发生或进行早期治理。

表 1 调整前后的离差法(标准差法)及百分位数法等级评价标准的 5 级划分原则

评价等级	离差法标准		百分位数法标准		离差法标准	百分位数法标准	两法调整后理论 %
	传统范围	理论 %	传统范围	理论 %	调整后范围	调整后范围	
上等	$> \bar{x} + 2s$	2.30	$> P_{97}$	3	$> \bar{x} + 1.28s$	$> P_{90}$	10
中上等	$> \bar{x} + s$	13.55	$> P_{75}$	22	$> \bar{x} + 0.67s$	$> P_{75}$	15
中等	$\bar{x} \pm s$	68.30	$P_{25} \sim P_{75}$	50	$\bar{x} \pm 0.67s$	$P_{25} \sim P_{75}$	50
中下等	$< \bar{x} - s$	13.55	$< P_{25}$	22	$< \bar{x} - 0.67s$	$< P_{25}$	15
下等	$< \bar{x} - 2s$	2.30	$< P_3$	3	$< \bar{x} - 1.28s$	$< P_{10}$	10

注 摘自文献[5]

这种人为地扩大下等和中下等个体所占百分比的调整 ,虽然可能会“ 增加 ”误诊率 ,但从预防的角度就有利于提醒其发育不理想 ,有必要作进一步研究和/或进行干预 ,较之未调整前的比例 ,对促进其生长发育意义更加积极。

骨骼发育标准也有类似情况。目前采用峰值骨量丢失量(率)诊断骨质疏松的方法采用正常为 $\geq (PBM - 1 SD)$,骨量减少为 $(PBM - 1 SD) > \sim \geq (PBM - 2 SD)$ 。从表 1 可见 ,在统计学上此时 $\geq (PBM_{理想} - 1 SD_{理想})$ 的“ 正常 ”个体约占 84.15% ,而 $< (PBM_{理想} - 1 SD_{理想})$ 的个体所占比例约为 15.85%。其中骨量减少 $(PBM_{理想} - 1 SD_{理想} \sim - 2 SD_{理想})$ 所占的比例又为 13.55%。虽然前面已提到疾病或亚疾病的人群只占少数 ,但从尽早进行预防的角度而言 ,本文作者认为该 13.55% 的比例有偏

而适当增加误诊将使“ 理想正常值 ”的范围相应缩小 ,但相对较高的“ 理想正常值 ”水平有利于提高人们对骨骼发育的重视 ,以体现骨质疏松作为一种发病缓慢的慢性病 ,其预防更重于临床治疗的总原则。

(5) 生长发育标准中标准差和百分位的调整对“ 理想正常值 ”和“ 疾病值 ”界值确定的启示 :

如何体现即使适当增加误诊也不希望漏诊的确定原则呢 ? 在儿童青少年生长发育标准的制定原则中对传统的离差等级评价法(标准差法)的标准差倍数及百分位等级评价法进行调整的方法可能有一定的借鉴意义。

该标准的标准差倍数及百分位调整情况见表 1。调整的原因是因为在传统的 5 级评价标准中 ,各等级所占的正态曲线下面积的百分比 ,以中等范围最大(占 68.30%)这可能使发育差距较大的个体均被评价为同一等级 ,以致影响了评价的客观性、可靠性和准确性 ,同时两种方法在比例上调整为一也防止了以往同一组数据不同评价方法间结果不一致的问题^[5]。

低的可能。

即使是在“ 理想人群 ”中 ,如果某个体的骨密度在该“ 理想人群 ”的 $PBM_{理想} - 0.67 SD_{理想}$ (第 25% 位)水平及以下 ,其骨密度也不应当算是很理想的水平。因而在“ 理想人群 ”中适当增加误诊率也有利于骨质疏松的预防。

同时从图 3 可以见到 ,在 $PBM_{理想} - 0.67 SD_{理想}$ 的水平时(如图 3 中 B 点) ,其骨密度值接近 50 岁的骨密度值 ,而 50 岁左右基本上是我国城市妇女的自然绝经年龄^[11] [注 :在 1999 年中国老年学学会骨质疏松委员会诊断学科组提出的骨质疏松的诊断标准建议值中所附《中国人原发性骨质疏松症按生理年龄自我预诊法表》(见表 2)中也以 49 岁为平均绝经年龄。但该年龄是一般调查所得 ,不是“ 理想人群 ”的平均绝经年龄]。大量研究表明 ,女性在绝经后初

期由于雌激素下降明显,骨量丢失明显(每年丢失约2%~5%),并可持续10年。随着年龄的增长,骨量丢失速度趋缓(每年丢失约1%);在绝经后20年中,骨总量丢失可达20%~30%^[12]。而这种绝经后骨量的快速丢失正是骨质疏松的预防和治疗中所需要关注的。因此,如果假设一个女性个体在30~35岁时其骨密度值即已达到绝经年龄时的骨密度值,则在其生活方式不改变的前提下,其将来在绝经后通过干预的效果可能也会比较差。不过这也反过来提示在女性的标准制定中其“理想人群”中的女性的平均绝经年龄时的骨密度是确定“理想正常值”下限值的一个重要参考值。

因此,建议“理想正常值”的下限可大致定在 $PBM_{理想} - 0.67 SD_{理想}$ 的水平,这是一个在临床和预防上均可以接受的水平。

考虑到骨质疏松发生后治疗效果差及其骨折后果的严重性,“疾病值”上限值的确定也应与“理想正常值”一样采取宁愿适当增加误诊率不愿增加漏诊率的原则来调整 $SD_{理想症状}$ 的倍数。但与“理想正常值”的下限25%百分位略有不同,可考虑将“疾病值”的上限值制定得高一些,例如可以假设该疾病界值内包括所有已出现症状人群的90%,此时的 $SD_{理想症状}$ 倍数可定为1.28倍左右(即 $M_{理想症状} + 1.28 SD_{理想症状}$ 包括了所有已出现症状人群的90%)。从图3可见,如果仍然采用前述的骨折研究数据即腰椎(L₂₋₄)骨密度值($M_{骨折}$)波动约在0.600~0.800(g/cm²)间,假设这时取中间值 $M_{骨折} = 0.700$ (g/cm²),且假设平均取 $SD_{骨折} = 0.100$ (g/cm²)时,此时 $M_{骨折} + 1.28 SD_{骨折} = 0.828$ (g/cm²)。另外,考虑到这些研究中仅仅是部分研究中一般骨折个体的骨密度值,而如果考虑骨痛个体,该骨密度值还可能更高一些,即可能在0.850(g/cm²)左右甚至更高。同时也可见到,在0.850(g/cm²)左右时相应的年龄在70岁左右,接近骨折的高发年龄。提示0.850(g/cm²)左右的水平作为“疾病值”的上限有一定合理性,是可以接受的,而取1.28倍左右的标准差也是有一定可行性的。

而假设将“疾病值”上限值确定为0.850(g/cm²)左右时,此时如果要将该“疾病值”上限值转换为 $PBM_{理想} - ?SD_{理想}$ 的形式,则从图3可见,该标准差的倍数约在2~2.5左右。显然,这种通过研究已出现症状人群的骨密度值以及合理的推理所得到的“理想人群”的“疾病值”及标准差倍数是较为科学和

可行的。

除了在上述三个值的推断上可用调整的标准差外,在“理想人群”各年龄组内的“理想人群值”分布中建议也可采用调整的标准差(或百分位)。这样就不仅可以了解个体的骨密度值与 $PBM_{理想}$ 的差距,而且可以了解个体的骨密度值在某年龄时的位置。例如,某60岁的个体按 $PBM_{理想} - ?SD_{理想}$ 法诊断为“骨量不足”,但其骨密度值与60岁组的“理想人群值”相比居于第90%位以上,则说明该个体虽然不能避免骨骼的退化,但其骨骼的退化速度较慢。提示其生活方式可能较好,可以继续保持或进一步加强。反之,如果其骨密度值与60岁组的“理想人群值”相比居于第10%位以下,则说明该个体需要加强干预的过程。其他年龄也是如此。这种综合的判断较单独用 $PBM_{理想} - ?SD_{理想}$ 法诊断的结果显然要更客观。

当然,这里的“理想正常值”和“疾病值”的标准差倍数的确定只是一个推理的过程,其中建议采取0.67倍或1.28倍的标准差(或75%或90%位)的原因是如果在儿童的身高发育标准中采取这个倍数,那么在将来儿童身高与骨骼发育值的研究中就有可能与成人的骨骼发育值在比例上取得部分统一(如“理想正常值”部分),使得骨骼发育的研究呈现从童年直到老年的一致性。但这并不意味着只能采用该倍数(或百分位数),也许将来在研究中发现其他的倍数更加合理,那么不单在成人的骨骼发育标准中采用,也可能反过来会对儿童的身高发育的等级划分产生影响。但无论怎样变化,均建议能够考虑到多学科之间的协调乃至标准的协调,使标准间在科学的前提下尽量不重复、不矛盾。

与上述概念值相比较,在目前的标准中,“正常”、“骨量减少”及“骨质疏松”无论是在定义上还是在界值界定上均缺乏较充分的理论依据及相应推论的支持,因而根据这种“标准”作出的“诊断”来作为防治的依据,应当说还缺乏充分的有效性。

(6)参照人群年龄组(段)的选择及各年龄组(段)样本数的确定:

并不是每个标准都需要按1岁一个组(段)来设定标准值。但按1岁一个组(段)来设定标准值在理论上肯定会好一些。因此在理论上,如果能按每岁一个组和所设定的“理想条件”来调查从发育较快的儿童青少年至已明显退化的80~90岁老年人的骨骼发育水平,且每岁的人数应满足统计学的要求。那就不仅可以了解在“理想条件”下整个人生阶段的骨骼发育规律(包括“理想条件”下准确的峰值骨密

度及峰值骨密度前后的发育、退化规律),而且可以制定出较准确、实用的骨骼发育的“理想标准”。

而除了分布为不能转换的偏态分布情况外,如果要在同一个标准中同时采用等效的标准差法和百分位法表示,则各年龄组(段)的样本数需要同时满足两种方法中任意一种方法的统计学最低要求。

在统计学上,在制定标准时,如果需要采取百分位法,则对制定标准的参照人群的样本数要求较高,每个年龄组人数如不足 150 人,制成的标准曲线两端(如 P_3 、 P_{97})值波动较大,影响标准的精确度^[3]。而在制定全国性的标准时,参照人群的每个年龄分段的人数不宜少于 500 人,千人以上更理想^[5]。

因此,如果按最低要求即每个年龄分段的人数为 150 人,且每岁 1 组,测量 10~90 岁共 72 个年龄组,男女两个性别,则至少一共约需要调查 21 600 人。如果说在过去由于对骨质疏松的发生机制了解不多,且对诊断标准的含义理解有误,以及经济、社会等的原因,导致目前仍然未能制定出实用的标准的话,那么在目前各个方面的条件已大为改善以及对骨质疏松的发生机制已有较多了解和标准含义的深刻理解的情况下,在我国选取一个“最适宜”地区进行最低人数为 21 600 人或更多的、较为全面的调查是有实现可能性的。

但仅 21 600 人的最低人数是否能作为全国的标准呢?笔者认为在理论上是可行的,笔者认为以往的观点即制定全国的标准需要 500 人甚至千人以上主要是为了在统计上减少由于人数较少时个别一些偏大值和/或偏小值而引起的误差,使标准更有代表性。但从过去的一些标准实际的制定过程来看,这种大数量人群才能制定标准的观点往往被理解为是在全国各“有代表性”的地区进行“正常人群”抽样,然后再合并进行统计后制定标准。但由于对这些“正常人群”缺乏严格的定义,且各地区的抽样水平往往会不完全一致,因而从某种意义上也可以理解为是一种“现状标准”。

而如果按照本文所讨论的只确定某人群为唯一

的参照人群的抽样点后,实际上已经是将该参照人群视为全国的“代表人群”,因此这时只要研究立项的理论依据正确、设计周密,并能够根据现有的各种知识严格地定义“理想人群”并进行严格的筛选的话,这些“理想人群”每个年龄组的人数在 150 人是可以代表相应人群的水平。当然,如果条件允许,每个年龄组的人数更多些当然更理想。但显然不必很刻板地非要在 500 人至千人以上。当然,如果按每 5 岁一个年龄组,则该人数显然已超出 500 人的低限。

2.2.5 目前我国在制定骨质疏松症诊断标准中存在的一些问题:

(1)目前我国骨质疏松症诊断标准概况:

1994 年以前全世界都执行 WHO 1985 年提出的峰值骨量丢失 2 个标准差(S 或 SD)诊断为骨质疏松症的标准。1994 年 WHO 制定了欧美白人妇女骨质疏松症诊断标准。即正常为测量值在同性别青年人骨密度平均值(均值为 $\bar{x}$ 或 M)减去 1 个标准差以内($\bar{x} - 1s$);在($\bar{x} - 1s$)至($\bar{x} - 2.5s$)之间为骨量减少;小于($\bar{x} - 2.5s$)为骨质疏松,同时有 1 处或 1 处以上骨折者为严重骨质疏松症。

1996 年日本骨代谢学会制定了日本人群的骨质疏松诊断标准,骨密度在同性别青年人平均值的 70%~80% 之间为骨量减少。骨密度低于平均值的 70% 为骨质疏松^[10]。

1999 年中国老年学学会骨质疏松委员会诊断学科组提出了骨质疏松的诊断标准建议值。标准差法为参考 WHO 的标准,结合我国国情,以种族、性别、地区的峰值骨量为依据而制定,主要用于女性成人、男性参照执行。百分率法为参考日本 1996 年改动版的标准而制定,即在自己尚未作峰值骨密度调查,亦或自己作了一些调查,但 SD 不便应用时,可用骨量丢失百分率(%)诊断法。(见表 2、表 3。表 2、表 3 中个别符号根据统计学概念及其他研究的常见表达方式笔者作了建议性修改^[1]。)

表 2 运用峰值骨量诊断骨质疏松症的两种诊断方法

标准差法		百分率(%)法		诊断结论
原表述	建议表述	原表述	建议表述	
$> M - 1 SD$	$\geq M - 1 SD$	$> M - 12\%$	$\geq M - M \times 13\%$	正常
$M - 1 SD \sim 2 SD$	$< M - 1 SD$	$M - 13\% \sim 24\%$	$< M - M \times 13\%$	骨量减少
$< M - 2 SD$	$< M - 2 SD$	$< M - 25\%$	$< M - M \times 25\%$	骨质疏松症(根据诊治要求分为轻、中二级) 伴有一处或多处骨折,为严重骨质疏松症 为严重骨质疏松症
$< M - 2 SD$	$< M - 2 SD$	$< M - 25\%$	$< M - M \times 25\%$	
$< M - 3 SD$	$< M - 3 SD$	$< M - 37\%$	$< M - M \times 37\%$	

注 摘自文献 [1], 笔者作了建议性修改

表 3 中国人原发性骨质疏松症生理年龄自我参考预诊法

分类	临床表现				峰值骨密度 丢失百分率(%) (与峰值骨量比)		相应骨峰值 标准差(SD) 的骨丢失	
	征象	程度	男性(岁)	女性(岁)	原表述	建议表述	原表述	建议表述
初期	无	-	64±8	49±7	<12	≤13	<1	≤1
骨量减少(轻度)	骨髓疼痛	+	74±8	59±7	13~24	>13	1~2	>1
骨质疏松症(中度)	骨髓疼痛或骨折	++	84±8	69±7	25~36	>25	2~3	>2
骨质疏松症(重度)	严重骨髓疼痛或骨折	+++	94±8	79±7	≥37	>37	≥3	>3

注:1. 例如:女性 49 岁绝经,那么 52 岁到 66 岁可患骨量减少,62 岁到 76 岁可患骨质疏松,72 岁到 86 岁可患严重骨质疏松症。由于个体、环境、营养和运动等差异,我们给出女性加减 7 岁,男性加减 8 岁的变化幅度。最后确诊应到医院骨密度室检查

2. 摘自文献[1],笔者作了建议性修改

(2)在统计学上存在的主要问题:

①在标准差法和百分率法的概念上存在的统计学问题

在百分率法的采用原则中“即在自己尚未作峰值骨密度调查,可用骨量丢失百分率(%)诊断法”的提法是有误的。很明显,如果“自己尚未作峰值骨密度调查”,那就只有用别人的峰值骨密度数据来诊断,那么用谁的峰值骨密度数据才合适呢?如果“自己作了一些调查,但 SD 不便应用”,那就说明这个调查可能存在“代表性”较差的问题。

而且在目前的标准判定方法中,通过简单的计算还可见到,如果在某研究“峰值骨密度(PBM)及其标准差(SD_{PBM})”一定时,同样是诊断为“骨量减少”,那么采用标准差法计算(PBM-1 SD_{PBM})与采用百分率法(M-M×13%)的诊断界值所得到的结果是不同的(即不等效),其他的几个界值也是如此。这显然是不妥的。

如果同一个研究的数据采用标准差法与采用百分率法时其某诊断的结果都可以是不等效的,那么在那些“自己尚未作峰值骨密度调查”但却采用了其他地区的峰值骨密度数据并采用百分率法来作为“标准”进行诊断的结果自然也肯定是不等效的。以此类推,那些“峰值骨密度及其标准差”或相同或不相同的,并且或采用标准差法或采用百分率法的众多的“标准”不仅在“代表性”上有疑问,这些“标准”的诊断结果肯定是五花八门的,其诊断结果的科学性、有效性可想而知。更不要说这些众多的研究本身的代表性、科学性、有效性问题了。

因而在逻辑上这些提法就是说过去的。因此,这种将“正常参考值”“符合”标准差法就用标准差法,“不符合”就抛弃标准差法采用百分率法来制定“标准”的标准制定方法在统计学的理论上是不妥的。

②在制定过程中存在的统计学问题

显然,按照前面讨论的要求,对于与年龄密切相关的骨密度指标而言,目前一些省(市)地区正在探讨的“标准”(或“正常参考值”,下同)的研究中,不要说按一些标准的一岁一个组(段)来进行抽样,即使按照每 5 岁(甚至 10 岁)一个年龄段也较少能完全满足每个年龄段 150 人的要求,更不要说 500 人及以上了。同时各年龄段的人数还往往相差较大(多数只是在 20~40 岁间的人数较多)。而即使是在这种情况下,在一些调查中,也只列出每个年龄段的总人数,未列出每个年龄段内部具体每岁年龄组的人数或构成。由于 5 岁(或 10 岁)一个年龄段本来就是较粗的分组,再加上这个较粗的年龄段总人数较少,这样具体到每一岁年龄的人数就必然更少,此时如果抽样也不规范的话(多半是一般的“正常”人群的体检的“正常人群”而不是经过严格定义的“正常”或“理想”人群),就很容易造成结果的偏倚。这是这些研究中出现测量值及其标准差波动较大的原因。这使得这些研究的“正常参考值”的代表性差。此时,即使规定了统一倍数的标准差(或丢失百分率),这些研究的“正常参考值”如果作为一个地区的标准也仍然是有问题的。

③在统计学的表达上值得商榷的问题

一般来讲,在不少采用标准差法的判定原则中,“正常”(或中等,下同)的统计学描述常将 $\bar{x} \pm s$ 的界值包括在内,这样有利于对称划分不同等级(或水平或程度,下同)的人群。即表述为下图形式(见图 4)。



图 4 标准差法中各等级评价的常见表达方式示意图

或以文字形式表述为:下下等 < $\bar{x} - 3s \leq$ 下等 < $\bar{x} - 2s \leq$ 中下等 < $\bar{x} - s \leq$ 中等 $\leq \bar{x} + s <$ 中上等 $\leq$

$\bar{x} + 2s < \text{上等} \leq \bar{x} + 3s < \text{上上等}$ 。

或表述为更为简明易懂的形式(见表 4)。

表 4 常见标准差法等级评价表

评价结论				
$\bar{x} + 3s$	<	上上等		
$\bar{x} + 2s$	<	上等	$\leq$	$\bar{x} + 3s$
$\bar{x} + s$	<	中上等	$\leq$	$\bar{x} + 2s$
$\bar{x} - s$	$\leq$	中等	$\leq$	$\bar{x} + s$
$\bar{x} - 2s$	$\leq$	中下等	<	$\bar{x} - s$
$\bar{x} - 3s$	$\leq$	下等	<	$\bar{x} - 2s$
		下下等	<	$\bar{x} - 3s$

这样表 2、表 3 中的原表述就可进行简化。百分位法的表述也是如此,但百分位法的原表述中如果仅用 M - 13% 等易被误解为 M 减去一个常数。

虽然以上将“正常”个体的统计学描述中一般都将 $\bar{x} \pm s$ 的界值包括在内的表达不是一种硬性原则,但由于骨骼发育与整个人生有关,考虑到将来条件改善后有可能将儿童青少年的骨骼发育与成年后的骨骼发育统一起来,那趁目前的不少相关研究工作尚在起步阶段,现在先将一些原则统一较之将来再统一可能会更好,即将来的研究数据与现在的数据对比时有更好的可比性。

(3)在实际操作中存在的主要问题：

由上述可见,目前国内外的骨质疏松症诊断标准实际上只是一个规定了统一的制定方法(如“结合我国国情,以种族、性别、地区的峰值骨量为依据而制定标准”)及判定方法(标准差法或百分率法)的“标准制定原则”而已,并不是一种具体的、统一的骨质疏松症诊断标准值。而即使是这个“标准制定原则”在理论和实际应用中也尚有一些值得商榷之处。

在“结合我国国情”的大框架下,这里的“种族”、“地区”概念均给人以不仅要各民族、而且要以各地区来“制定自己的标准”的印象。这些不仅在理论上有所误(原因见前所述),而且也不具备实际操作的可行性。

因为在实际操作中首先遇到的问题是,这个“不同地区”的具体划分条件是什么?例如,是按行政区划(大至华北、华南、东北等大区,中至省、直辖市、自治区,小至各省内的各地区,细分则可至县市)来划分?还是按民族(在某种程度上往往被理解为“种族”)来划分?或者是按其他条件如城乡、生活习惯/条件等来划分?抑或是这些所有条件的综合?

在此对“标准”的数量不妨作一个简单的推论,如果按省(市)来划分,则至少要制定 32 个省(市)的标准,如果按民族来划分,则要制定 56 个民族的标准。

既然在汉族要制定 32 个省(市)的不同标准,那么,其他一些跨省(市)的、较大的少数民族当然也要制定其所在省(市)的标准。如果按城乡来划分,则上述标准的数目要翻倍。如果再加上生活习惯/条件(如何界定生活习惯/条件?)等的综合则更复杂。

实际上,无论怎样设定这些条件,在中国广袤的国土上,如果完全满足这些“独特”的条件来划分地区/人群的话,相应理论上这个标准的数量就可能是少则几十,多则上百甚至上千个(中国有 2000 多个县市)。面对这些数量很多的标准,不要说按照严格的统计学要求以及更细的年龄段划分来做,就是简单的按照现在的方式做,在数年甚至数十年内完成都几乎是不可能的。并且一种类型的测量机型尚且如此复杂,其他不同类型机型、不同部位的标准制定则工作量更大。

而如果不能在短期内完成这些标准,我们如何才能及时、有效地指导医生或普通老百姓进行骨质疏松的治疗和预防呢?而如果在短期内完成了这些标准,但这些标准又千差万别,不仅可能导致某个体在同一地区用同一标准但判断方法不同(如标准差法或百分率法)而得到不同的诊断结果,而且可能导致该个体在不同地区采用不同的标准得到不同的诊断结果,这些五花八门的标准在学术理论和/或实际应用中的价值可想而知。

尽管《中国人骨质疏松症建议诊断标准(第 2 稿)》^[1]中在今后的展望中提出在将来应“形成一个普遍的、统一的、有指导意义的、科学的和便于操作的诊断标准”、“应分别建立不同测量机器(如 US、DEXA、QCT、pQCT、pDEXA、SPA 等)的正常值”以及“在数据统计处理中,应以 5 岁为一个年龄组,防止因 10 岁一组造成年龄组过大,数据过于分散的现象”。但由于在“结合我国国情,以种族、性别、地区的峰值骨量为依据而制定标准”的概念主导下已给人造成了要以各民族、各地区来“制定自己的标准”的印象,而在数据统计处理中,应以 5 岁为一个年龄组,防止因 10 岁一组造成年龄组过大,数据过于分散的现象”则一方面提示了目前各地在探讨“本地正常参考值”时所存在的、应注意的问题,另一方面却恰恰进一步强化了各民族、各地应制定“本地正常参考值”的印象。此时易给人的印象是各地先制定自己的标准,然后再从中进行挑选以“形成一个普遍的、统一的、有指导意义的、科学的和便于操作的诊断标准”,这显然在理论上是不符合逻辑,在实际上也是不可行的。

可想而知,在此前提下,制定一种测量机器的“标准”都存在理论的缺陷及不具备实际操作的可行性,而同时“应分别建立不同测量机器的正常值”则更不现实了。

3 今后骨骼发育标准(骨质疏松症标准)的制定和相应研究的方向展望

综上所述,由于目前国内外在骨质疏松症诊断标准的定义、概念和方法等理论方面的不统一,导致目前国内“标准”(包括不同机型内置的各种标准或其他正在探讨的各种“正常参考值”)多、方法杂,甚至出现了几个研究者在研究同一地区的“正常参考值”。这些不规范的“标准”不仅造成了学术、应用上的混乱和人力、物力的浪费,最根本的是这些“标准”的应用不能很好地寻找出不同区域间的影响骨骼发育或退化的因素,不利于对急需及时进行正确诊断和有效防治的老年人群进行相应的治疗和预防,这是与当前的新医学模式相违背的。所以改变上述情况并结合实际情况尽快制定出更科学且适合我国人群的、统一的、新的骨质疏松系列诊断标准已势在必行。

本文列出儿童青少年的一些数据,一方面是因为儿童青少年的骨骼发育与成年后的骨骼发育有关联,另一方面更多的是通过这些数据对今后的骨质疏松症研究进行一些推论和建议。但由于儿童青少年时期的骨骼水平并不完全与成年后的骨骼水平成正比,而成年后的骨骼水平的种族间差异也可能并不像身高、体重的差异那么大。因而笔者建议今后的研究方向可能应首先综合目前各地已有的各种研究数据,提出制定理想标准的“最适宜”地区及其“理想人群”,再进行相应部位的不同仪器的标准制定(同一种方法、不同厂家的仪器可采用标准物进行标定校正)。而一旦制定了统一的理想标准后,各地的研究应主要集中在应用该理想标准来进行相应的诊断和相关影响因素的寻找和对理想标准的合理性进行不断的建议及修改方面,以促进骨骼发育及骨质疏松症的预防和治疗。同时,制定一个合理、科学的标准,不仅有利于相关疾病的治疗,也有利于统一研究的方向,使全国其他的相应研究更有针对性,相互间的结果也更有可比性。而在更远的将来还可能根据人类基因研究的结果来制定更准确的不同基因型人群的骨骼发育标准。

笔者主要是从遗传学、统计学和预防医学的角度来初步探讨骨骼发育的标准制定原则,未更多的讨论临床的诊断,但临床的诊断不应脱离预防的角度来单独进行临床的诊断标准制定,否则这种“标准”的意义是有限的。但由于涉及到骨骼发育的因素很多,不可能一一涉及,更多、更全面的有关标准制定的原则、方法等还有待于全国相关研究同仁的进一步探索。

【参考文献】

- [1] Research Section of the Diagnostic Criteria of Osteoporosis, Commission of Osteoporosis, Association of Chinese Geriatrics. The diagnostic criteria of osteoporosis for Chinese (second edition). Chinese Journal of Osteoporosis 2000 6(1):1-3 (in Chinese).
- [2] YU W. The diagnosis of osteoporosis and the determination of the bone mineral content. Journal of Chinese Physician 2006, 34(7):7-11 (in Chinese).
- [3] YE GJ. Child and Adolescent Health. 4th, ed, Beijing: People's Medical Publishing House 2001, 38-63 (in Chinese).
- [4] DU RF. Populations genetics of Chinese. Beijing: Science Press, 2004, 794-797 (in Chinese).
- [5] YE GJ. Modern Child and Adolescent Health. Beijing: People's Medical Publishing House, 1994, 23-27, 207-209 (in Chinese).
- [6] Research Section of the Constitution and Health of Chinese Students. Report on the physical fitness and health surveillance of Chinese school students. Beijing: Higher Education Press 2002, 251-252, 325-332, 341-348 (in Chinese).
- [7] LI NH, OU PZ, ZHU HM, et al. Comparison study on application of different diagnostic criteria to the prevalence rate of osteoporosis. Modern Rehabilitation 2001, 5(12):114-115, 117 (in Chinese).
- [8] OU PZ, DENG LP, LIU YS, et al. Prevalence rate of osteopenia and osteoporosis in middle-aged and elderly population in Guangzhou. Chinese Journal of Osteoporosis 2002 8(4):314-318 (in Chinese).
- [9] AN Z, WANG WZ, YANG DZ, et al. The epidemiological investigation of primary osteoporosis patients in urban and rural area in Chengdu. Chinese Journal of Osteoporosis 2002 8(3):233-236 (in Chinese).
- [10] LIU ZG, PAN ZA, WANG SL. A probe into diagnostic standard of primary osteoporosis. Chinese Journal of Osteoporosis 1997 3(1):1-15 (in Chinese).
- [11] LE J. Obstetrics and Gynecology. 5th, ed, Beijing: People's Medical Publishing House 2000, 382-386 (in Chinese).
- [12] QUAN XH, ZHU XY, WU XW, et al. Relationship of age, age at menopause, menopausal period with BMD of lumbar spines and hip in postmenopausal women. Chinese Journal of Osteoporosis 2006, 12(3):226-228 (in Chinese).

(收稿日期:2007-10-24)