

·综述·

骨质疏松症临床试验设计的统计学要点及相关问题

宇传华 方积乾

骨质疏松症是以低骨矿密度和骨纤维结构退化为特征的一种系统性骨骼疾病。由于骨脆性的增加,病人很容易在近端股骨(髋部)、椎体、桡骨远端、肱骨近端和踝部发生非创伤性骨折。骨质疏松症临床试验的目的就是寻求较好的措施降低和停止骨量丢失,建立新的骨组织(如果可能),最终达到减少骨折的发生与复发率。一项临床试验的成败与否,与其统计学设计有着密切的关系。

一、骨质疏松症临床试验的设计

骨质疏松症临床试验是一类前瞻性研究,分为治疗性和预防性两种,分别以骨质疏松症病人或健康人为研究对象,比较临床治疗或预防干预措施与相应对照措施的效果。

临床试验一般分为Ⅰ、Ⅱ、Ⅲ、Ⅳ期。Ⅰ期临床试验是初步的临床药理学及人体安全性评价试验,观察人体对于新药的耐受程度和药物代谢动力学情况,为制定给药方案提供依据。Ⅱ期临床试验是随机化、有对照的临床研究,对新药有效性及安全性作出初步评价,建立最小有效剂量和剂量反应曲线,推荐临床给药剂量。骨质疏松症Ⅱ期临床试验的研究时间一般应达到12个月,并应采用双盲法和安慰剂平行对照。Ⅲ期临床试验是扩大的多中心临床试验,遵循随机对照原则,进一步评价药物的有效性与安全性;这一时期一般应成立专门的委员会或小组,对各试验点执行情况实施监控;美国食物与药品管理部门FDA建议,骨质疏松症治疗性临床研究时间为3年,预防性临床研究时间至少2年。Ⅳ期临床试验是新药上市后的监测,在广泛使用的条件下全面考察研究药物的治疗效果与不良反应,尤其是罕见的不良反应^[1]。

下面讨论骨质疏松症的Ⅱ期与Ⅲ期临床试验设计的统计学要求。

1. 明确研究终点

美国FDA推荐的治疗性与预防性用药研究的疗效终点见表1^[1]。除了非雌激素类药物的测量终点为骨折外,其余均可采用骨矿密度(bone mineral density, BMD)作为研究终点。治疗前应注意记录和测量病人的基线数据,如BMD是多少,以前有无骨折等。

严格说来,某药物可维持或提高BMD仅仅说明它具有减少骨折危险的可能性,而骨折发生率的减少才能真正证明药物的疗效。此外,反映椎骨高度的身高改变、疾病痛苦与

活动受限情况的改善也只能作为疗效的辅助指标。进行骨质疏松症的治疗性临床试验需要对骨折进行评估,因此需要有明确的骨折定义。70年代和80年代早期,许多研究采用骨矿密度间接反映骨折情况,但这种方法明显存在着问题,例如Riggs及其同事(1990)^[2]采用氟化钠治疗已绝经妇女骨质疏松症骨折的研究发现,尽管BMD明显增加,但脊椎骨折发生率没有明显降低,非脊椎骨折反而有所增加。其原因在于BMD的增加影响了骨的结构,从而降低了骨的质量。

表1 用药研究的疗效终点

名称	治疗性	预防性
雌激素类药物的测量终点	BMD(因已证实治疗能明显降低椎骨及非椎骨的骨折危险)	BMD
非雌激素类药物的测量终点	骨折	BMD

脊椎骨折最常见,有人建议用它作为研究终点,但由于脊椎骨折的临床定义与解释多种多样,它的发生可以没有症状,诊断新的脊椎骨折有一定难度。髋部骨折的发生有症状表现,相对容易跟踪,但由于低发生率,将髋部骨折作为研究终点,需要相对大的研究样本量(4 000以上)和较长的观察时间(几年),才能获得有统计学意义的结果。因此,大多数骨质疏松症研究采用BMD和脊椎骨折的联合改变作为研究终点,最近也有采用髋部骨折作为终点的研究。

除研究终点外,还应注意年龄、性别、体重指数、骨折史(特别是近5年)、骨折家族史、吸烟否、使用糖皮质激素否、维生素D缺乏否、基因缺陷等情况。体重指数(body mass index, BMI)定义为体重(kg)/[身高(m)]²;一般认为BMI小于19,则有可能导致患骨质疏松症。

2. 确定研究对象入选与排除标准

为了保持研究对象的合理性,首先应根据具体的研究目的明确受试对象的人选标准与排除标准。

美国FDA推荐的骨质疏松症治疗性临床试验的人选标准是:绝经至少达5年以上的非住院病人;行走自如;具有1处或多处与骨质疏松有关的椎体骨折或腰椎骨矿密度比一般绝经前妇女低2个标准差。这类研究不应选用以下对象(排除标准):影响骨代谢的有关疾病患者;前6个月内曾接受骨质疏松症药物治疗(不包括补钙)者;长期持久使用可能影响骨钙代谢的药物者;经骨活组织检查确诊为骨软化者;维生素D缺乏者^[1]。

预防性临床试验的入选标准是:绝经后1~3年、45岁以上、行走自如的非住院骨质疏松症高危人群。这类研究不应选用以下对象(排除标准):确诊患各种骨质疏松症,或卵巢切除术后促卵泡激素明显升高大于 $50\text{U}\cdot\text{L}^{-1}$ 者^[1]。

3. 样本量的估计

对于Ⅱ期、Ⅲ期临床试验,如果试验组与对照组之间的药效有差异,为了在统计学上显示这种差异,研究前应根据药物的预期效果与个体差异,以及试验组与对照组间骨折发生率等,按统计学要求估计所需的样本量。通常情况下,如果不是明确知道试验组药物好于对照组,试验组与对照组的样本量应按1:1随机分配,且按双侧假设检验确定研究的样本量。

一般而言,临床试验需要较大的样本量,不单是为了显示统计学差异,更是为了考察可推广性。为了考察1个药物能否应用于广大人群,试验样本应能代表各式各样的人,通常总要以千计。否则,小样本的两组有差异,并不能保证结果的可推广性和在广大人群中的可应用性。

4. 常用的3类对照组

临床试验中常用的对照组有以下3种类型

(1)安慰剂对照:安慰剂是1种虚拟药物,其外观剂型、大小、颜色、重量、气味、口味和服药方式等都与试验药物尽可能保持一致,但不含有试验药物的有效成分。设置安慰剂对照的目的在于:消除研究者、受试者和评价人员等由心理因素等影响而形成的偏倚,从而分离出由试验药物所引起的真正疗效与不良反应。

(2)阳性药物对照:在临床试验中采用已知的、已批准上市的有效药物作为试验药的对照,称为阳性药物对照。阳性对照药物必须是合法的、公认有效的、最为安全的药物。

(3)剂量-反应对照:根据试验药物的性质设计几个剂量组,将受试者随机地分入其中的1个剂量组;这种对照也可包括安慰剂对照即零剂量,也可不包括。剂量-反应对照主要用于研究疗效或不良反应随剂量变化的情形。剂量-反应对照有助于回答给药方案中哪种剂量较合适。

临床试验中试验组与对照组唯一的差别应该是试验组受试者接受试验药物治疗,而对照组受试者接受对照药物治疗。1个临床试验不一定只有1个对照组,可根据实际情况设立多个对照组。如在一个阳性药物对照的临床试验中,增加1个安慰剂对照组,就形成同时使用安慰剂和阳性药物对照组的试验,常称为3组试验;又如在安慰剂对照试验中,根据医学伦理要求,有时需在给每个受试者同一种标准治疗措施的前提下,试验组给予试验药物、对照组给予安慰剂,这种试验称为标准治疗加安慰剂的试验。

一般说,对照可分平行对照和交叉对照。平行对照是将来自同一总体的受试者随机平行分成若干组,除施加的处理不同外,其他均相同或相似;交叉对照是指同一受试者在不同的时期先后给以处理和对照。骨质疏松症的预防性与治疗性临床试验应采用平行对照,而不宜采用交叉对照。应特别强调,对照组必须和试验组在同时、同地、同条件下进行,

不宜以回顾性资料、文献资料或异地的资料为对照。

5. 随机化分组

为了突出反映试验组与对照组药物间疗效和不良反应的差别,应采用随机化方法对受试对象进行分组,使得每个受试对象具有同等的机会被分配到不同的组,以排除其他因素的混杂作用。例如,通过随机化可使基线骨折次数、绝经后的年数、骨矿密度的基本水平等在试验组与对照组间基本均衡,以致两组间的差异主要反映试验药物与对照药物的差异。

随机化的方法很多,日常生活中的抽签或掷骰子就是最简单的随机化方法。在科学实验中,常通过随机数字表或计算机随机数发生器来实现随机化。

尽管简单随机化分组能使干扰因素在各组的分布基本均衡,但并不一定能完全保证某些重要的干扰因素在组间的均衡性。此时宜先对可能影响实验结束的重要干扰因素进行分层,然后在每一层内进行随机化,这种方法称为分层随机化(stratified randomization)。例如,为了保证试验组与对照组受试者的性别比例相同,可分别按男女性别施行随机化分组。又如,在多中心临床试验中,受试者分别在各中心内进行随机化分组。注意,分层随机化应使每一层具有足够的样本量,总样本量过少时不宜采用过多的分层。

6. 双盲法

在试验的全过程应采取特定措施,不让参与临床试验的人员知道病人所接受的是何种药物,这些人员包括受试者、研究者、疗效和安全性评价者、监视员、数据管理者和统计分析者,从而避免他们对试验结果的人为干扰。这类措施称为盲法(blinding)。通常分为单盲和双盲两种,仅仅受试者一方不知道接受的是何种药物时为单盲,受试者和直接研究人员双方都不知道是何种药物时为双盲。

从医学伦理学考虑,双盲试验应为每一个受试者编号并设置一个应急信件,信件内容为该编号的受试者所分入的组别(盲底)。应急信件是密封的,随相应编号的试验药物发往各临床试验中心,由该中心负责人保存,非必要不可拆阅。在发生紧急情况(如严重不良事件,或病人需要抢救)必须知道该病人接受的是何种处理时,由该中心负责人拆阅。一旦被拆阅,该编号病例就作为脱落处理,不计入疗效分析;但有不良反应时仍需计入安全性分析;应急信件的拆阅率超过20%,意味着双盲试验失败。所有应急信件在试验结束后随受试者报告表一起收回^[3]。

双盲临床试验常采用两次揭盲,第1次揭盲是将各病例号所对应的分组(如A组与B组)告诉生物统计学家,以便他们对全部数据进行统计分析;当分析结束、总结报告完成时,再作第2次揭盲,宣布A、B组哪一个是试验组。

7. 伦理学问题

与动物实验不同,临床试验是以人为研究对象,必须考虑病人的依从性问题,研究者不能完全支配病人的行为,只能要求病人避免采用某些干扰试验的治疗或行为。当新药已被证实对病人弊大于利,尽管试验尚未完成也要终止;当

有确切证据表明研究药物对疾病有效时,继续使用安慰剂对照是不适当的;当病人出现紧急情况需要采用额外治疗措施时,也必须给予。

二、骨质疏松症临床试验设计中存在的问题

回顾有关杂志的报道,我们发现骨质疏松症临床试验设计中存在一些共同的问题,现罗列于后,以引起注意。

1. 试验前缺乏周密的研究设计

大部分骨质疏松症的临床研究事先都没有1份详细的设计书,试验目的、研究对象入选和排除标准、样本含量、疗效及安全性评价指标等并不是事先设计好的,而是从事后病例总结中归纳出来的。这类研究只能视为病例总结,而不能算作临床试验。

2. 对照的设立不规范

骨质疏松症研究的对照应和试验组来自同一总体,除施加的处理不同外,其他均应相同。某研究以不足10例门诊绝经后骨质疏松症患者为试验组成员,却以数十例非老年健康女性作为对照组成员。尽管此研究设有对照,但试验组和对照组来自不同的总体,除地理因素以外两组人员的年龄、健康状况等都不相同,如果两组的疗效不同,究竟归功于药物还是归功于其他因素?

3. 未按随机化原则分组

大多数研究人员都知道随机化的重要性,不少文章声称将患者随机分成两组,但只字不提如何进行随机化。也许实际上根本就没有随机化,有的是根据工作的方便来分组,有的是根据病人的意愿来分组。这样不但不能保证其他因素在组间的均衡性,而且极容易带来病人意向等主观偏倚。

4. 试验没有采用双盲法

对骨质疏松症研究的受试者和各类直接研究人员采用盲法也是消除各种影响因素的重要方法之一。绝大多数骨质疏松症的临床试验没有采用盲法,即使有的采用了盲法,也只是对受试者采用盲法(单盲),而很少有采用双盲法的。这样,在测定、评价过程中由人为因素导致临床研究结果偏倚是不可避免的。

5. 样本量不足

近10年我国骨质疏松症临床试验中,绝大多数研究的样本量在100例以下,多中心的临床试验研究很少有报道。如鲑鱼降钙素治疗绝经后骨质疏松症仅9例患者(1997年);骨康口服液治疗绝经后骨质疏松症,患者为36例(1999年);阿法D₃软胶囊与葡萄糖酸钙治疗NIDDM伴骨质疏松症,患者为57例(1997年);阿法骨化醇(立庆片)治疗原发性骨质疏松症,患者为60例(1999年)等等。因样本量较少,反映药物疗效的BMD、骨折发生率的变化往往无统计学意义。

6. 观察时间短

我国骨质疏松症临床观察时间一般都只有半年左右,如某研究的临床试验时间仅为20周,达不到美国FDA推荐的Ⅱ期临床试验观察年限1年的水平。欲使治疗骨质疏松症的药物获得批准,临床研究一般需要3~5年^[1]。

7. 诊断标准含糊不清

如某研究治疗的患者标准为:“骨矿密度及骨矿含量低于非老年健康女性,同时有腰背酸痛、乏力及骨折等”,定义较含糊,执行者操作困难。

骨质疏松症诊断的指导思想是将个体的骨矿密度与年轻成人峰值骨量(peak bone mass, PBM)的均数和标准差(s)相比较。所谓年轻成人是指排除了对骨矿密度有影响的有关疾病和因素的20~40岁同民族、同性别、同地区成人。1994年Kanis等^[5]以25~50岁人群作为年轻成人提出的诊断标准得到了WHO的推荐。所谓年轻成人峰值骨量是指人在生命成熟期获得的最高骨量,一般可按5岁1个年龄组求每组的骨矿密度平均值,由最高者作为峰值骨量均数。

1994年WHO建议的女性骨质疏松症诊断标准为^[5]:与参照组年轻成人的BMD均数相比,减少量不足1 s 者为正常;减少量在1 s ~2.5 s 之间为骨量减少(osteopenia);减少量大于2.5 s 为骨质疏松症(osteoporosis);减少量大于2.5 s ,且伴有一处以上脆性骨折者为严重骨质疏松症(severe osteoporosis)。

个体BMD与峰值骨量均数之差和 s 的比值称为T值(T-score)。鉴于国内现有统计资料提供的 s 值偏大,有学者建议将我国骨质疏松症诊断标准定为:女性,T值大于2;男性,T值大于2.5^[6]。

8. 信息记录与利用不足

治疗性临床试验的主要目的是降低骨折发生率,但药物以外的许多因素,如健康状况、平时体育锻炼、钙与维生素D的补充、预防摔跤的健康教育等对其骨折发生率有很大影响^[4]。这些因素在试验组与对照组间的不均衡势必会影响研究的结果,现有的研究工作较少注意记录各种可能的影响因素,如试验组与对照组的基线骨矿密度、年龄、民族、身高、体重、吸烟史、(女性)绝经年限、以往骨折次数等,即便有所记录,也很少采用多因素统计分析方法充分利用这类信息。

三、讨论与建议

1. 大规模流行病学调查是当务之急

骨质疏松症的诊断与参照人群的选取有关,不同的参照人群有不同的骨峰值均数与标准差。国外人群的骨峰值均数与标准差显然不能为我所用。我国幅员广大、民族众多,用单一的参照人群和单一的峰值均数与标准差来进行诊断也容易出现较多的误诊与漏诊。

迄今为止,国内已有一些流行病学调查资料可供参考,但由于覆盖面、样本量以及测量的信度、效度等问题,现有统计资料的权威性尚嫌不足。因此,在我国有计划地开展较大规模的BMD流行病学调查以取得各类人群的基础资料是一项首要的工作。

2. 尽快开展多中心、长时间临床试验研究

现有的骨质疏松症临床研究大多基于单独一家医院的资料,属于少数人的个体行为,难免出现小样本问题。这种分头进行的小样本研究充其量只能起预试验的作用,重复报

(下转第176页)

去势大鼠股骨强度逐渐增大,其中,XW630组术后60 d 股骨强度已与 Sham 组接近,与 OVX 组之间存在差异($P < 0.05$),且高于 CFT 组,而术后 90 d 股骨强度则恢复至正常。由此可见,XW630 具有预防骨质疏松性骨折发生的作用。

骨组织的力学特性不仅决定于骨质密度,还与骨小梁的微观结构有着非常密切的关系^[12,13]。骨小梁体积分数(Vv)、骨小梁厚度(Tb、Th)、骨小梁数目(Tb、N)、骨小梁分隔距离(Tb、Sp)等参数指标可以反映骨小梁的基本结构特点及其与骨组织整体结构的关系。本研究对不同时期各组胫骨骨小梁结构参数测定结果显示,OVX 组的 3 个骨形成参数 Vv、Tb、Th、Tb、N 均逐渐下降,而骨吸收参数 Tb、Sp 则逐渐增加,表明骨吸收大于骨形成,呈骨量丢失状态。两个治疗组骨形成参数均逐渐增加,而骨吸收参数则逐渐减小,与 OVX 组比较存在不同的显著性差异($P < 0.05$ 和 $P < 0.01$),表明 CFT 和 XW630 均有促进去势大鼠骨组织成骨的作用。其中,XW630 组胫骨骨小梁平均厚度(Tb、Th)测定值在术后 90 d 高于正常对照组,这一结果与其相应的组织学和 SEM 观察结果相符,表明 XW630 促进去势大鼠骨组织的成骨作用较雌酚酮强,至术后 90 d,骨小梁数目基本恢复至正常,且骨小梁的形态更显粗壮^[3]。由此可见,XW630 可以有效地促进去势大鼠骨组织的成骨和预防骨质疏松性骨折的发生,预示着 XW630 有可能成为临床上广泛应用的新一代抗骨质疏松症和促进

骨修复的药物。

参 考 文 献

- 1 涂桂生 防治骨质疏松症药物研究进展. 化学医药工业信息, 1995, 11(7): 1-5.
- 2 Riggs BL. Tibolone as an alternative to estrogen for the prevention of postmenopausal osteoporosis in selected postmenopausal women. J Clin Endocrinol Metab, 1996, 81: 2417-2418.
- 3 费伟, 王大章, 郑虎, 等 XW630 促进去势大鼠骨组织成骨作用的实验研究. 华西口腔医学杂志, 2001, 19(6): 383-387.
- 4 Riggs BL, Melton LJ. Involutional osteoporosis. N Engl J Med, 1986, 314: 1676-1680.
- 5 Eriksen EF, Colvard DS, Berg NJ, et al. Evidence of estrogen receptors in normal human osteoblast-like cells. Science, 1988, 214: 84-86.
- 6 Komm BS, Terperung CM, Benz DJ, et al. Estrogen binding, receptor mRNA, and biologic response in osteoblast-like osteosarcoma cells. Science, 1988, 214: 81-84.
- 7 Grady D, Gebretsadik T, Kerlikowske K, et al. Hormone replacement therapy and endometrial cancer risk: a meta-analysis. Obstet Gynecol, 1995, 85: 304-313.
- 8 Rubin G, Peterson H, Lee N, et al. Estrogen replacement therapy and the risk of endometrial cancer. Am J Obstet Gynecol, 1990, 162: 148-154.
- 9 Steinberg K, Thacker S, Smith S, et al. A meta-analysis of the effect of estrogen replacement therapy on the risk of breast cancer. JAMA, 1991, 265: 1985-1990.
- 10 Thompson DD, Simon HA, Pirie CM, et al. FDA guidelines and animal models for osteoporosis. Bone, 1995, 17(4S): 125S-133S.
- 11 SØgaard CH, Mosekilde L, Schwartz W, et al. Effect of fluoride on rat vertebral body biomechanical competence and bone mass. Bone, 1995, 16: 163-168.
- 12 Dempster DW, Ferguson-Pell MW, Mellish RWE, et al. Relationships between bone structure in the iliac crest and bone structure and strength in the lumbar spine. Osteoporos Int, 1993, 3: 90-96.
- 13 Odgaard A. Three-dimensional methods for quantification of cancellous bone architecture. Bone, 1997, 20: 315-328.