

论药品有效期估计的量化模型

纪建勋

厦门大学附属第一医院, 福建 厦门 361003

摘要: 运用量化模型估计药品有效期是长期稳定实验的核心及新药申报中的重要节点。通过对比最新《中国药典(2015版)》中的对应规定(第IV部, 9001章)与国际法规ICH Q1E中的估计方法, 阐明ICH Q1E是符合国内法规的。其次, 通过理论分析解释了国内与国际法规中量化模型所隐含的假设及局限性。最后, 通过蒙特卡洛模拟得出有效期估计值的经验分布, 表明取均值为最终有效期估计值的方法, 可能比ICH Q1E中的方法更倾向于低估真实有效期, 从而置制造方于不必要的劣势。

关键词: 有效期; 稳定性试验; 质量控制; 统计模型; ICH

中图分类号: R965.2 **文献标志码:** A **文章编号:** 1674-6376(2017)05-0642-06

DOI: 10.7501/j.issn.1674-6376.2017.05.011

Quantitative models for drug shelf-life estimation

Ji Jian-xun

The First Affiliated Hospital Of Xiamen University, Xiamen 361003, China

Abstract: Shelf-life estimation is the core of stability tests and critical to the final label claim on the package of new drug. Firstly, this paper depicts the methods of ICH Q1E for shelf-life estimation, compares it to the counterpart in the newest version of *Chinese Pharmacopeia* (Ch. P 2015, IV-9001), and proves its compliance. Secondly, this paper also unfolds the latent assumptions and hinged limitations of the quantitative methods in the guidelines. Finally, the empirical distributions of the shelf-life estimators are obtained via Monte-Carlo simulation to facilitate better understanding of properties of the estimators. It is shown that the approach that takes the average as final shelf-life estimate is likely to further underestimate the true shelf-life comparing to the ICH Q1E approach, and hence puts the manufacturers in unnecessary disadvantage.

Keywords: shelf-life; stability test; quality control; statistical models; ICH

如何确定药物产品(化学药品或生物制剂)的有效期是长期稳定性试验的核心,也是申请新药批准和上市中的一个重要节点。长期稳定性试验和其他稳定性试验需依照国内法规《中国药典(2015版)》^[1](第IV部, 9001章)执行,其中未确切规定的细项可以参照常用的国际法规,如美日欧国际协调委员会(ICH)指导原则Q1^[2-3]。

药物产品的有效期指的是,在给定的贮存条件下,该产品在市售包装下各关键质量(Critical Quality Attribute)指标预期能持续保持在检验标准(shelf life specification)内的时期。其中,(1)给定的贮存条件(一般指温度与相对湿度,无直接光照)应与实际保存环境匹配。根据我国南北地域条件,

药典中指定贮存条件为温度(25±2)℃、相对湿度(60±10)%,或温度(30±2)℃、相对湿度(65±5)%。(2)关键质量指标指的是那些非临床的可测量指标,用于从多个(物理或生化)角度描述药品。药品的质量最直观反映在临床的安全性和有效性上,但显然临床试验无法作为质量控制的直接手段。针对各药品类型(化学药品或生物制剂)和具体药品,专家们(药学、药物动力学和临床)与监管机构会研究提出一系列关键质量指标。通过控制这些指标以期保证生产的药品质量。(3)各关键指标所对应的检验标准通常是,先由制造方提议,再经监管机构审核批准后确定的。制造方提议时一般是根据部分长期试验数据的趋势适度外扩最终成为标准。

收稿日期: 2016-10-12

作者简介: 纪建勋, 厦门大学附属第一医院门诊西药房药师。Tel: 022-84845240 E-mail: 184849023@qq.com

据药典规定并结合常见的实际运作,长期试验需包含3批次药品的数据,每批次药品需以今后销售批次的规模(或稍小规模)生产,对包装后成品进行科学抽样,保留足够数量并贮存在规定条件下,之后在第0、3、6、9、12、18、24、36个月,随机取样以测量各关键质量指标并汇报数据。由于所有测量方法都存在一定的系统误差,汇报数据一般是取多个独立样本值的平均值。每个测量法的评估检验和汇报方法应在长期试验前就已经确定并提交监管机构。通常,每批次药品的每个关键质量指标在某试验时间点上仅有一个汇报数据。基于这些数据并给定检验标准,该药品关于每个关键质量指标的有效期就可以通过统计模型进行估计。最终该药品标注在包装上的有效期是,由制造方整合所有有效期估计值进行提议,并经由药监局批准后确定的。

下文将首先详细介绍有效期估计的量化模型,阐述国内药典相关要求与国际法规 ICH Q1E 是如何接轨的。然后,通过蒙特卡洛模拟比较国内与国际法规中估计方法的差异。最后,总结比较结果并讨论量化模型的统计学性质与隐含假设。

1 国际现行的量化方法(ICH Q1E^[3])与国内法规

现仅考虑关键质量指标会随时间线性渐增或渐减。基于前文所描述的长期试验的汇报数据,简单的线性回归模型(以关键质量指标为响应量,时间为自变量)可以用于拟合某批次的数据,然后预测在哪一个时间点其响应量均值的95%置信限与检验标准相交,该时间点就是该批次的有效期估计值。如图1所示,如果检验标准是单侧(如上侧),那么计算的95%置信限也应是相同的单侧(如上侧)。

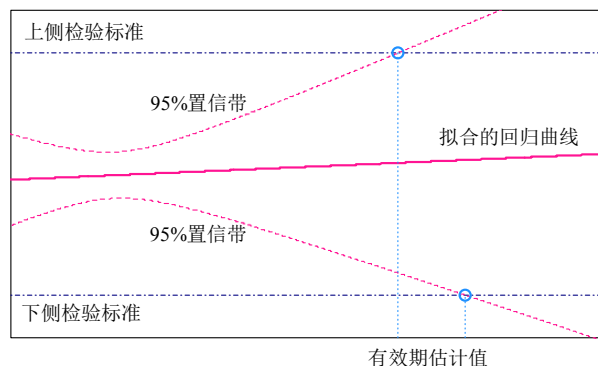


图1 当检验标准为双侧时,如何估计有效期

Fig. 1 Graphical illustration of shelf-life estimation with two-sided specification limits

如果检验标准是双侧,那么置信限也应是双侧,有效期估计值为任意一侧置信限最先与检验标准相交的时间点。

以上估计某批次有效期的框架同时满足《中国药典》和 ICH Q1A (R2)^[2]的要求。两者之间未完全相同的部分在于如何综合3批次数据(或3批次分别的有效期估计值)计算关键质量指标的有效期。相较于 ICH Q1E 指导原则中的具体估算模型,《中国药典》有更大的弹性。本小节将先阐述 ICH Q1E 中的量化方法(见模型1),解释它为何是符合国内药典规定的,然后再提出另一种更简单且也符合规定的实用方法,用作对比。通过之后小节3中的理论讨论和小节4中的模拟分析,读者可以更清楚的了解国内和国际法规建议方法的性质和局限。

阐述量化方法前,为了清楚的表述统计模型,本文作如下符号约定。 t 表示时间(单位:月), Y 表示关键质量指标的随机观测值, e 表示随机干扰项[通常假设服从正态分布 $N(0, \sigma^2)$], ID 是代表批次的虚拟变量组合, a 表示截距, b 表示时间回归系数。

模型1(ICH Q1E 分布模型):

1) 综合3批次的数据,拟合饱和模型 $Y=a+aID+bt+bIDt+e$,并用方差分析检验饱和模型中 aID 和 $bIDt$ 项是否显著, P 值检验标准为0.25。该饱和模型旨在使每个批次有不同的截距和时间系数,但用全部的数据拟合随机方差。

2) 如果 aID 项显著($P \leq 0.25$)但 $bIDt$ 项不显著($P > 0.25$),最终模型应该拟合 $Y=a+aID+bt+bIDt+e$,使3批次有不同的截距但相同的时间系数;如果 aID 和 $bIDt$ 项都不显著,最终模型应该拟合 $Y=a+bt+e$,称此为同系数模型,它使3批次有相同的截距和时间系数;其他情形下,最终模型即是步骤(1)中的饱和模型。

3) 依照最终模型计算其95%置信带(例如如图1中的红色虚线),然后取置信带与检验标准的交点所对应的时间为有效期估计。该置信带上的每一点都是对应时间点上预测均值的(单侧或双侧)95%置信限。如果同系数模型是最终模型,则3批次仅有1个有效期估计值,该值便是该关键质量指标的有效期估计值。如果同系数模型不是最终模型,则3批次各有一个有效期估计值,取其中的最小值为该关键质量指标的有效期估计值。

应注意的是,如果步骤(2)中 aID 项不显著

而 $bIDt$ 项显著, ICH Q1E 没有建议使 3 批次有不同截距但相同的时间系数, 而是使用饱和模型做最终模型。

区别于 ICH Q1E 的具体, 国内药典给予更多弹性, 仅规定某关键质量指标的最终有效期估计应如下计算: “如 3 批统计分析结果差别较小, 则取其平均值为有效期限。若差别较大, 则取其最短的为有效期”。然而如何定义差别和判别“差别较小”的阈值, 药典中没有明确给出。模型 1 是通过方差分析检验批次间的回归系数是否显著差异, 以此来判别“3 批统计分析结果差别”的大小。如果没有, 则使 3 批次有相同的回归系数得到唯一的有效期估计值。如果有, 则分别求取有效期再取最小值。因此 ICH Q1E 是符合国内法规的。

现在考虑更简单直接的建模方法, 即对每批次的数据分别建模以估计有效期, 并定义最大与最小估计值间的比率为批次间统计分析结果的差别, 标记为 $R_{\text{Max/Min}}$ 。当 $R_{\text{Max/Min}}$ 越靠近 1, 表示 3 个估计值间差别越小, 反之差别越大。如果 $R_{\text{Max/Min}}$ 小于或等于阈值, 则取平均值为最终有效期估计, 反之取最小值。关于如何确定阈值, 小节 4 运用蒙特卡洛模拟进行讨论。最后, 如何分别求取每批次的有效期估计值也需要明确说明, 可行的方法包括模型 1 中非同系数的两个模型。为了更好的对比, 现在仅考虑最简单直接的模型 (模型 2), 即对每批次数据分别独立建模。

模型 2 (例子模型): 对每批次数据分别独立拟合回归模型 $Y=a+bt+e$, 计算 95% 置信带, 并确定其与检验标准的交点为有效期估计值。取这 3 个有效期估计值的均值为该关键质量指标的有效期估计值。

这个模型 2 与模型 1 中饱和模型的区别在于批次内随机误差的估计。两个模型都使每批次有独立的截距和时间系数, 但模型 1 中的饱和模型假设 3 批次有相同的随机干扰, 所以综合 3 批次的残差用以估计随机干扰方差 (σ^2)。而模型 2 假设每批次有独立的随机干扰。

2 关于量化方法的理论分析

本小节进一步从理论上分析国际和国内法规中量化模型的隐含假设、性质和局限性。如果某批次的的数据是根据模型 $Y=a+bt+e$ 产生, 且模型系数 (a 和 b) 的真实值可知, 该批次的真实有效期即是

$\frac{\text{Spec } a}{b}$, 其中 Spec 表示检验标准。现实中, 真实的

模型不可知, 且观测的数据包含随机干扰, 因此基于数据推断的回归系数 $\hat{a}$ 和 $\hat{b}$ 也存在不可避免的误差。同批次内的观测数据至少包含两种随机干扰, 即样本的随机性和测量时的系统误差。为了安全性着想, 国内药典和 ICH 都规定有效期估计必须取 (预测均值的) 95% 置信限与检验标准的交点, 而不是取 $\frac{\text{Spec } \hat{a}}{\hat{b}}$, 以期该估计值有很大可能比真实有效期

小。如图 1 所示, 简单线性回归的置信带总在拟合的回归线的两侧, 且中部内凹向回归线, 内凹的最值在观测数据的平均样本时间。批次内的随机干扰越小, 置信带越靠近回归线。给定任何有限方差的随机干扰, 随着用于拟合回归线的样本数增加, 置信带会逐渐收敛于回归线。由于 ICH Q1E 的广泛使用, 已有国外文献对模型假设和性质进行了解析, 可参见参考文献^[4]。本小节对模型 1 和 2 的两个重要方面进行讨论。

(1) 线性和正态性假设

模型 1 和 2 都假设关键质量指标会随时间线性变化。如果已有研究表示某关键质量指标与时间的关系是非线性的, 且有具体的表达式, 非线性模型可以用于拟合模型。有效期的估计框架依旧是, 在给定模型下寻找关于预测均值的 95% 置信限与检验标准的交点。比较常见的情况是, 先将测量值进行对数转换后, 再假设其与时间的线性关系, 例如生物制剂的效价。另外, 如果对随即干扰的分布假设存在怀疑, 拟合模型后可以提取残差作 Shapiro-Wilk 正规性检验。当检验 P 值足够小 (如 ≤ 0.01), 分析员可为随机干扰选择其他更合理的分布假设。最后, 由于现实操作中每个时间点一般仅有一个汇报数据, 本文不讨论是否应该进行方差齐性检验。

(2) 固定影响模型的局限性

模型 1 和 2 都没有在建模时考虑批次间的随机性, 而是选用固定影响模型 (Fixed-effect model)^[5-6], 即假设每批次的数据是基于不可知的某线性模型生成, 而该模型的系数是不可知但固定的。统计理论上, 运用固定影响模型所确定的有效期估计无法将结论延展到其他批次, 即估计结果仅反映了长期试验 3 批次的稳定时期, 这显然与长期试验的根本目的相悖。确定有效期是为了说明在生产线运作良好的情况下, 现在和未来所有批次的稳定时期。尽管存在这样的理论缺陷, 固定影响模型还是被广泛使用, 这主要是由于它的易操作性。国内和国际法规

显然也考虑到了固定影响模型的局限性,所以采用最小值,即3批次中最坏的情况,来保守估计有效期。但是这样的策略依然假设了长期试验中3批次完全代表了未来所有销售批次。为了考虑批次间的随机性,可行的方法是运用随机系数模型(Random-coefficients model)^[5-6],该模型将综合3批次数据,假设每批次的数据是基于某线性模型产生,而模型的系数服从某分布(以此反映批次间的随机性),随机系数分布具有不可知的真实均值,这些均值确定了真实的有效期。拟合此种模型后,仍可以取预测均值的95%置信带与检验标准的交点为有效期估计值,该置信带将包含了批次间、批次内的随即干扰和样本数的影响。这种方法综合3批次数据仅有一个有效期估计值。关于这种方法更详细的讨论可参见^[4, 7-8]。

3 蒙特卡洛模拟

如前文论述,国内药典比ICH Q1E(模型1)更有弹性,允许运用其他模型(如模型2),并可以在3批次分析结果差异较小时取各自有效期估计的均值为最终有效期估计值。本小节将通过蒙特卡洛模拟^[9-10]说明确定 $R_{\text{Max/Min}}$ 阈值在实际操作上的困难和取均值的方法可能导致过于保守的有效期估计值。

借鉴于ICH Q1E,模型1通过拟合回归系数均无显著差异来判定3批次有效期估计差别较小,并采用同系数模型得到唯一有效期估计值。本小节用同系数模型随机产生3批次数据,即 $Y_{it}=1+0.5t+e_{it}$ 。其中 $i=1, 2, 3$ 表示批次, t 取值于 $\{0, 3, 6, 9, 12\}$, Y_{it} 表示第 i 批次在时间点 t 的关键质量指标观测数据, e_{it} 相互独立且取同分布 $N(0, s^2)$, s 取值于 $\{1, 0.5\}$ 表示随即干扰的强度。上单侧检验标准定为13,因此3批次的真实有效期均为2年。在时间回归系数不为零的情况下,单侧检验标准总能根据给定回归系数取值,而使真实有效期为2年。因此,回归系数1和0.5的设定不影响本小节结论的广泛性。另外,当真实截距为1时,3批次数据在时间点0(即长期试验开始时)各自服从 $N(0, s^2)$ 。所以, s 的取值也表示时间点0上每批次数据的真实相对标准差(或变异系数)。这样的设定使模拟的结果可以不限于关键质量指标的测量单位和量级。本文模拟假设有1年数据可用于预测,下文讨论均在此设定下,不另外强调。其他样本数设定可以用同样模拟方法分析,但预期对本文讨论结果没有影响,因此不做延展。

根据前文模型随机生成3批次数据值后,每批

次有效期可以按不同模型进行估计。重复这样的模拟5 000次,并记录每次不同模型拟合生成数据而计算的估计值。因而每个模型有5 000个模拟估计值所形成的经验分布来描述其样本随机性和平均精确度。对每个随机生成的3批次数据,三个模型将用于拟合:(1)完整的ICH Q1E方法(模型1),(2)仅使用模型1中的同系数模型(不进行子模型选择),(3)模型2。注意,在每次模拟中,前两个模型均仅有1个最终有效期估计值,最后一个模型(模型2)在未确定 $R_{\text{Max/Min}}$ 阈值前应保留3批次各自的有效期估计值。

现先仅观察完整的ICH Q1E方法和仅用同系数模型的模拟估计值经验分布(如图2所示)。同系数模型的估计值呈(向左)偏态分布(见图2蓝色直方图),而模型1的估计值经验分布(见图2绿色直方图)应是三个分布的混合体,混合系数是(基于样本数据分布)分别选择3个子模型为最终模型的概率。两种估计方法的经验分布均大部分在真实有效期(2年/24个月)的左侧,因此,如国内和国际法规期望的,取95%置信限与检验标准的交点有很大几率低估有效期。随着样本数据的随即干扰(s^2)的减弱,两个估计值的经验分布均向真实有效期收敛,因而平均精确度提高。另外,模拟发现不论真实随机干扰项取值大小,当真实模型为同系数模型,ICH Q1E方法仅有60%几率选择同系数模型为最终模型。

其次,观察ICH Q1E方法和模型2(且取均值)的估计值经验分布(如图3所示)。不考虑阈值,取3批次各自估计值的均值为最终有效期估计,这样的估计方法其经验分布(见图3红色直方图)比ICH Q1E(见图3绿色直方图)有更大面积偏左,因而对真实有效期估计的平均精确度更低。这种情况下,采用模型2(且取均值)的方法将使制造方处于没有必要的劣势。当然随着样本随机干扰的减弱,这种方法也会收敛向真实有效期。

运用模型2且取均值的方法如上所说可能过于保守。除此之外,确定 $R_{\text{Max/Min}}$ 阈值的也存在很大困难。模拟中,运用模型2估计的3批次的各自的有效期估计值均被保留,因而每次模拟都可以计算一个 $R_{\text{Max/Min}}$ 。在给定的随机干扰强度下,5000个模拟值形成的经验分布(见图4)表示其分布偏态向其理论最小值1。真实的随机干扰强度越小, $R_{\text{Max/Min}}$ 的分布越收敛向1。

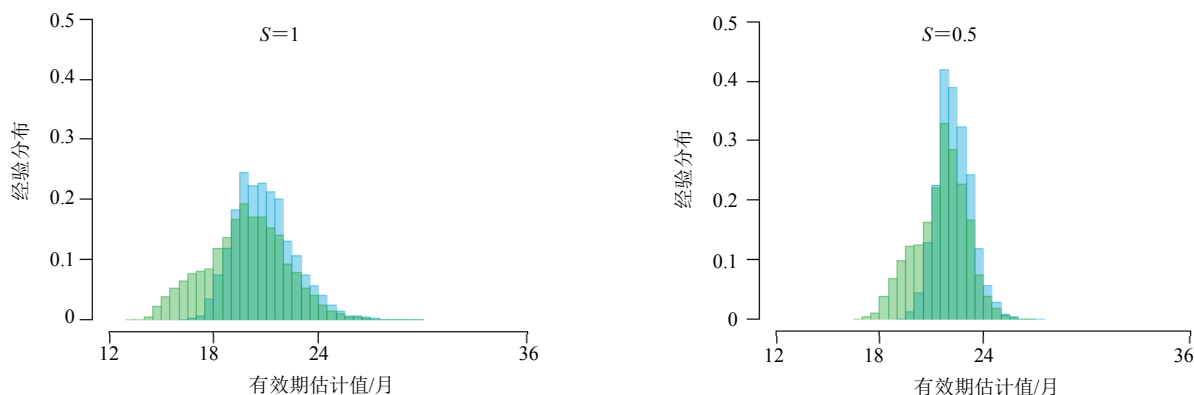


图2 在不同强度的随即干扰下 ($S=1$ 或 0.5), 完整的 ICH Q1E 方法 (绿色直方图) 和仅用同系数模型 (蓝色直方图) 的估计值经验分布 (基于 5000 次重复模拟)

Fig. 2 Empirical distributions of shelf-life estimator according to ICH Q1E (green histogram) and estimator with same coefficients across batches (blue histogram) under different settings of population standard deviation ($S = 1$ or 0.5), based on 5000 repeats of simulation for each scenario

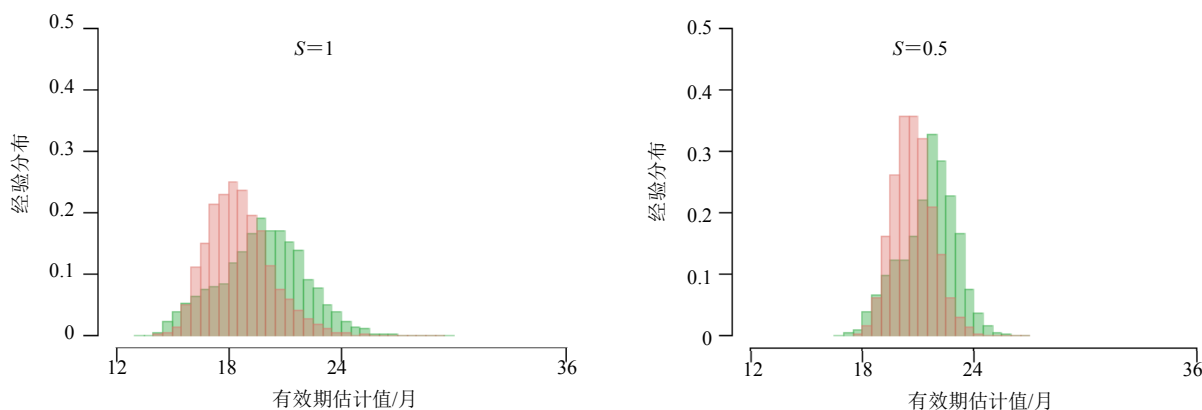


图3 在不同强度的随即干扰下 ($S=1$ 或 0.5), 完整的 ICH Q1E 方法 (粉色直方图) 和模型 2 且取均值 (蓝色直方图) 的估计值经验分布 (基于 5000 次重复模拟)

Fig. 3 Empirical distributions of shelf-life estimator according to ICH Q1E (Pink histogram) and estimator using Model 2 for each batch and then averaging (blue histogram) under different settings of population standard deviation ($S = 1$ or 0.5), based on 5000 repeats of simulation for each scenario

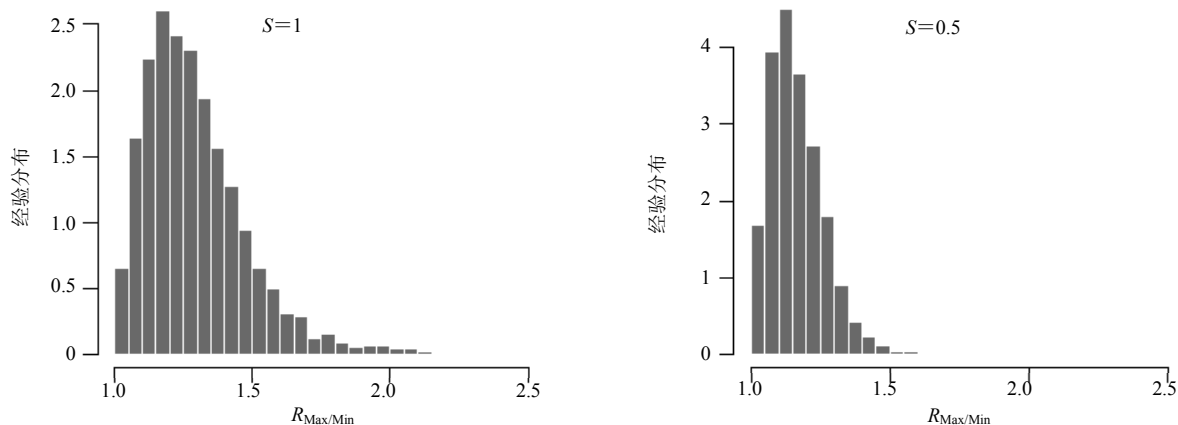


图4 在不同强度的随即干扰下 ($S=1$ 或 0.5), $R_{\text{Max/Min}}$ 的模拟经验分布

Fig. 4 Empirical distribution of $R_{\text{Max/Min}}$ under difference setting of population standard deviation ($S = 1$ or 0.5) based on 5000 simulations for each scenario

这不利于确定阈值。借鉴于 ICH Q1E 方法和同系数模型的对比,当真实模型是同系数模型时,假设阈值的设定应使有 60% 的几率取平均值。因此,阈值可以设定在模拟经验分布的 60% 分位数。当 $s=1$ 时,阈值应为 ≤ 1.30 ; 当 $s=0.5$ 时,阈值应为 ≤ 1.17 。真实随机干扰的强度越大,应设定的阈值越高。然而,真实的随机干扰强度是不可知的。即便运用本文中的模拟方法,通过设置不同的真实随机干扰强度,并找到对应的阈值,制成表格。在真实随机干扰强度不可知的情况下,只能用样本估计的干扰强度寻找对应阈值,这显然将引入不可预测的误差,阈值控制的通过几率可能高于或低于预期。因而,这样的操作不严谨。如果不用比率,而选用最大最小估计值间的绝对差异来描述 3 批次统计分析结果的差异,这个难题仍然存在。

4 结语

ICH Q1E 给出了具体的有效期估计的方法,它满足国内药典的要求。其建议的估计方法是基于固定影响模型,存在一定的局限性。但通过模拟,该估计方法确实很可能低估真实有效期,因此有效的保护病人利益。制造方可以通过优化关键质量指标的测量方法,减小数据中的随机误差干扰来得到更靠近真实有效期的估计值。虽然国内药典的规定比起 ICH Q1E 更有弹性,允许当 3 批次统计结果分析差异较小时采用均值为最终估计值。但采取最小值是减小固定影响模型局限性的关键。再者,通过模拟,本文发现这种方法可能比国际法规更保守(更倾向于低估真实有效期),从而置制造方于不必要的劣势。另外,如何量化“统计结果分析差异”和确

定阈值也存在实际中难以操作的难题。因此,本文建议采用 ICH Q1E 方法,避免使用均值。

参考文献

- [1] Ch. P (2015) Vol IV (中国药典 2015 年版. 四部) [S]. 2015.
- [2] ICH. Quality guidelines. Stability Q1A (R2), Stability testing of new drug substances and products [EB/OL]. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q1A_R2/Step4/Q1A_R2_Guideline.pdf. 06 /02 /2003.
- [3] ICH. Quality guidelines. Stability Q1E, Evaluation of stability data [EB/OL]. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q1E/Step4/Q1E_Guideline.pdf. 06 /02 /2003.
- [4] Capen R, Christopher D, Forenzo P, et al. On the shelf life of pharmaceutical products [J]. AAPS PharmSciTech, 2012, 13 (3): 911-918.
- [5] Demidenko E. Mixed models: Theory and applications with R [M]. Wiley-Blackwell, 2013.
- [6] Fitzmaurice G M, Laird N M, Ware J H. Applied longitudinal anlysis [M]. Wiley-Blackwell, 2011.
- [7] Chow S C, Shao J. Estimating drug shelf-life with random batches [J]. Biometrics, 1991, 47 (3): 1071-1079.
- [8] Quinlan M, Stroup W, Schwenke J, Christopher D. Evaluating the performance of the ICH guidelines for shelf life estimation [J]. J Biopharmaceut Stat, 2014, 23 (4): 881-896.
- [9] Robert C, Casella G. Monte carlo statistical methods [M]. Springer-Verlag New York, 2004.
- [10] Liu J S. Monte carlo strategies in scientific computing [M]. Springer-Verlag New York, 2001.
- [11] (上接第 641 页)
induce binucleation in L5178Y and TK6 cells [J]. Mutat Res, 2014, 770(8):29-34.
- [15] Vlastos D, Drosopoulou E, Efthimiou I, et al. Genotoxic and antigenotoxic assessment of chios mastic oil by the *in vitro* micronucleus test on human lymphocytes and the *in vivo* wing somatic test on drosophila [J]. Plos One, 2015, 10(6): e0130498. doi: 10.1371/journal.pone.0130498. eCollection 2015.
- [16] 胡燕平, 宋 捷, 王 欣, 等. 麦冬水煎剂的遗传毒性研究 [J]. 中国中医药信息杂志, 2009, 16(7): 38-40.
- [17] 胡燕平, 王 欣, 宋 捷, 等. 五味子水煎剂的遗传毒性研究 [J]. 癌变. 畸变. 突变, 2009, 21(4): 309-312.
- [18] Nai G A, de Oliveira M C, de Oliveira Tavares G, et al. Evaluation of genotoxicity induced by repetitive administration of local anaesthetics: an experimental study in rats [J]. Rev Bras Anesthesiol, 2015, 65(1): 21-26.
- [19] 肖 凯, 李宏霞. 环磷酸胺剂量, 取样时间对小鼠骨髓微核率的影响 [J]. 癌变. 畸变. 突变, 2005, 17(6): 367-369.
- [20] 欧红梅, 周长慧, 涂宏刚, 等. 流式细胞术检测体外微核方法的验证 [J]. 中国新药杂志, 2015, 24(14): 1590-1598.