



动物和人类药物毒性一致性的大数据

发布时间: 2019-10-09 09:09:17 分享到:

摘要: 尽管缺乏疗效是药物开发后期消耗的一个重要原因, 但临床前开发过程中观察到的毒性转化为临床试验或批准后观察到的毒性转化的缺点是一个正在进行的研究课题。临床前和临床安全性观察结果之间的一致性仅在相对较小的数据集上进行了分析, 主要是在药物批准的短时间内。因此, 我们探讨了对一套3290批准的药物和制剂进行大数据分析的可行性。其中, 在超过70年的监管提交中报告了1637449例人类和动物物种的不良事件。5个物种(大鼠、狗、小鼠、兔子和食蟹猴)报告的事件被视为人类事件的诊断测试, 并使用似然比计算每个事件/物种对的诊断能力。许多关键观察结果的动物-人翻译被证实是可预测的, 例如狗的QT延长和心律失常。我们的研究证实了动物安全观察对人类的一般预测性, 但也确定了这种自动化分析的问题, 一方面与数据管理和控制词汇有关, 另一方面与随着时间推移的方法学变化有关。

关键词: 动物研究的可预测性 一致性 不良事件 预测值 翻译 信号检测 大数据 风险评估

简介: 对晚期药物消耗原因的分析表明, 除了缺乏疗效外, 临床前研究中没有充分预测到的临床安全性失败仍然是主要因素。预测人类不良事件的关键性转化仍然主要基于动物实验。这些动物模型中的许多都是通过长期的经验发展而成的, 并已编入国际法规中。尽管经验表明许多动物模型是人类的有效替代物, 但对动物模型观察所隐含的人类风险的严格统计评估仍然有限, 主要原因是缺乏足够大且精心整理的临床前和临床数据集。最重要的是提高对临床前发现转化为人类不良事件的可译性的理解, 这将有助于早期药物开发的决策, 是否继续或停止基于临床前发现的特定化合物。它还将有助于对当前进行的动物研究和研究项目进行批判性评估, 并最终促进为那些显示出低一致性的终点寻找新的和更好的临床前安全标记。最后但并非最不重要的是, 出于伦理原因, 必须进行彻底的一致性分析, 以证明动物的使用是正当的, 特别是因为一些作者声称普遍缺乏可预测性。奥尔森和他的同事利用一组由150名候选药物组成的精选集, 计算了动物对人类预测的敏感性 $TP / (TP + FN)$, 在一致性分析领域进行了一项重要的研究。奥尔森的出版物引发



比，狗表现出更好的预测能力。Tamaki等人比较142种批准药物的药物不良反应的一致性，确认血液系统疾病的药物不良反应的一致性大于70%，而肌肉骨骼、呼吸和神经系统的药物不良反应的一致性小于30%。他们得出的结论是，动物没有毒性对人类没有不良事件具有很低的预测性。另一方面，毒性的存在为人类不良事件的发生增加了大量的证据，这些不良事件通常由高阳性似然比（IR+）证明，这表明在被调查物种中有毒的药物可能会导致人类不良事件。然而，不同种类的化合物之间的IR+浓度及其作用被发现是低的。更具体地关注单个终点，如QT延长，显示一小组化合物（n=23）通过使用相关的非啮齿类动物模型预测人类qt结果的概率为90%。Mead等人根据IRwin试验（FOB/IT）中的神经功能评估，分析了141种化合物与中枢神经系统（CNS）相关的安全问题的一致性。作者发现第一阶段临床试验中最常见的5个不良事件，即头痛、恶心、头晕、疲劳/嗜睡和疼痛，FOB/IT无法预测。大多数动物与人的一致性研究是基于对真阳性或真阳性率（TP/（TP+FN））的评估，对假阳性的分析有限。事实上，许多动物阳性不能区分真假，因为它们是通过不进行药物临床试验来避免的，这是一致性分析的一个不可避免的现实。作者将3815种药物的临床前安全性研究结果与所选临床MedDRA不良类别的不良事件相关联，结合所有种类并研究一些类别的事件。研究发现，一些非临床观察表明，在临床研究中观察相同不良事件的风险更高，特别是心律失常、QT延长和肝功能异常。另一方面，发现在非临床研究中缺乏这些事件并不能很好地预测人类的安全性，因此部分证实了贝利等人的发现。本研究是先前方法的延伸，现在分析了五种最常用的动物和人类报告的所有事件，以建立一个与统计指标一致的数据库。随着文本挖掘和大数据评估方法的出现，我们现在可以根据FDA和EMA监管文件评估所有已批准药物和制剂的所有报告事件，以评估当前的经验。这些文件提供了一个很好的数据集，因为每种药物都有一组经过严格审查的动物和临床试验。在我们的研究中，我们决定排除上市后事件，因为这些事件是通过不太严格的标准和条件报告的。然而，必须指出的是，即使一致性可以通过大数据分析来量化，从物种之间似乎不一致的观察中得出的结论最多的是，某些物种被优先用于调查在其他物种中未搜索到的特定终点（例如，光毒性是在啮齿动物身上测试的，而不是在狗身上测试的）。此外，许多动物阳性病例，无论是真是假，都会在动物观察阻止临床试验进展的情况下被排除，从而导致数据库倾斜。由于FDA/EMA数据集仅包括已获批准或已获批准的药物的信息，因此它代表了一个故意限制人类观察严重性的选择。我们的研究旨在解决以下问题：临床试验报告的不良事件与动物观察结果的正一致性是什么？正一致性，即动物对人类观察的可翻译性，是否依赖于所使用的动物物种？是否可以通过观察高于首选术语水平的更高级别观察分组来增加正一致性？这项研究还旨在回答这样一个问题：在多大程度上，这样一种大数据方法是可行的，而且比以前出版的一致性分析中繁琐的数据收集、整理和分析更有效。上述问题的答案对以下方面很有价值：通过评估该物种对预期观察的预测能力，指导选择研究物种，指根据经验采用强统计标准观察所隐含的风险，根据具体的动物观察结果，加强对潜在临床风险的评估。

方法：数据：Elsevier Pharma Pendium数据库被用作临床前安全相关观察以及临床试验记录和药物批准文件中报告的不良事件的来源。目前包含3920种药物和药物制剂的1637449次不良事件报告，并跨越超过70年的药物批准期。该数据库对适应症没有限制，涵盖了小分子和其他形式（“生物制品”）。文本挖掘确定研究作者是否陈述了丙氨酸转氨酶（ALT）的升高，并确定为不良事件。当然，许多报告测量并报告了谷丙转氨酶水平，但只有当研究作者得出结论，即可能与治疗相关，药物治疗中的不良事件记录才会产生。也就是说，该索引使用研究主管向FDA所作的结论和声明来判断丙氨酸氨基转移酶是否显著高于正常水平。仅使用对照临床研究的报告，即上市后不良事件报告不包括在本研究中，因为这些药物警戒数据包含许多难以



这些关系的实用性和统计显著性。这些总数是根据同一类别中的任何观察结果计算得出的。例如，对于心脏疾病，如果一种药物在该器官类别下发生任何事件（例如室性心动过速），并且该动物在该类别中发生任何事件（即使不是同一事件），则将该药物计算为真阳性。（例如心动过缓）。一致性度量：通过每对物种观察的2×2应急表评估动物与人类的一致性关系。我们将动物观察作为临床观察的诊断试验，并使用制定的统计方法评估诊断试验的疗效。真阴性（TN）值被计算为人类或动物均未观察到的值。这是一个估计，因为并非所有可能的观察都是测量或试图观察每种药物。在我们的研究中，我们使用正似然比“IR+”，根据图1中的值用公式 $TP^* (FP+TN) / FP^* (TP+FN)$ 计算，以表示动物和临床观察之间的统计关系。似然比表示在动物研究中观察到不良事件时临床风险的变化。似然比的优点是它独立于每项观察的患病率（贝叶斯先验概率），因此它在不同不良事件中比条件概率或阳性预测值更具可比性。在这项研究中，我们使用IR+10或更大值作为强一致性的截止值。还计算了负似然比“IR-”，它表示如果不进行动物观察，风险降低。一般来说，较高的正似然比表明了比较低值更可靠的相关性。用Yates校正卡方量表测量统计学意义。卡方检验测量动物和人类观察之间是否存在统计学上的显著关系。除非另有说明，否则只讨论99%置信水平或以上的关系。似然比测量动物观察的预测值。如果卡方很高，似然比很低，我们可以说动物观察不能预测人类观察，这是很有信心的。

结果：总体真阳性率：按MedDRA系统器官类别和种类分组的真实阳性不良事件的总总和如图2所示。一个物种内不同器官类别的事件分布如图3所示。统计结果表明，在物种或器官分类中，一致性事件并不均匀分布。真正阳性一致性总数最高的是狗的“胃肠道疾病”，紧随其后的是“一般性疾病和给药部位疾病”。图3显示了这些事件是如何分布在器官分类中的，归一化后每个物种的楔形物总和为100%。这个百分比分布一方面说明了一些物种专门用于测试特定终点的事实，例如，近25%的家兔TP观察值属于给药部位反应器官类别，反映兔子在皮肤、眼睛和注射部位的局部耐受性测试的使用。另一方面，它也证实了物种对特定终点的记录敏感性，例如狗23%的TP观察结果为胃肠道疾病，17%为心脏疾病。为了能够区分TP分布百分比和单个物种用于测试特定终点和器官类别的使用频率，图4显示了根据物种使用频率调整的百分比分布。在大鼠和狗身上观察到动物使用频率标准化的最高真阳性率，狗表现出较高的TPS，尤其是心脏和胃肠道疾病。而大鼠的肾、尿、呼吸、胸、纵隔疾病以及内分泌疾病的发病率更高。

分组为更高级别MedDRA类别的总体统计数据：在下一步中，我们更详细地研究了特定的端点以及根据MedDRA级别聚合数据的效果。表5显示了在MedDRA层级增加水平下观察qt延长的统计效应。在优选的术语水平上，有69种药物在动物和人身上观察到效果，可能性比为10.7，即表明如果在动物身上报告，观察到人类不良事件的概率会发生重大且经常是决定性的变化。“心电图调查”，有更多的一致性事件，但可能性比较低，因为增加了包含更多假阳性和阴性的子类别的其他成员。“心电图qt延长”效应与其他67个心电图参数（如PQ、PR、PRS、QRS和其他间隔）一起归入高级术语“心电图调查”。“心电图调查”是10个高级术语之一，包括心脏成像程序、血管成像和其他属于高级组术语“心脏和血管调查（不包括酶试验）”的术语。

物种间事件一致性的比较：图7显示了5种动物在人类一致性连接的首选术语水平上的事件，在99.99%置信水平下过滤，5种或更多的真实阳性药物，似



是心律失常，这表明狗和兔子有很高的一致性。虽然遥测犬是人类心血管事件的已知预测物种。小鼠和猴之间“存在药物特异性抗体”的观察结果很可能具有很强的一致性，反映了转基因小鼠（或使用小鼠替代物的野生型小鼠）和非人类灵长类动物在评估蛋白质药物（“生物制剂”）方面的优先选择。同样的道理也适用于局部作用于眼睛和皮肤，在兔上表现出高度一致性。如上所述，这种一致性很可能是由以下事实驱动的：这些特定的终点在过去曾在兔子身上被优先研究过（皮肤或眼睛腐蚀或刺激的“Draize测试”）。总之，这一图形分析一方面表明，动物物种和人类之间的一致性主要取决于特定安全终点的物种选择。另一方面，它表明所选物种对感兴趣的终点具有预测性。

按物种划分的最一致的动物模型：高似然比表明，在动物模型中观察每一个这些事件表明在人类中观察的风险很高。TP值对应于图1中动物和人类观察到的药物数量。如上文所述，某些高TP值和阳性似然比的一致性是由特定动物物种对这些终点的几乎独家使用引起的，这一点可以再次为兔子说明：除血肌酐增加外，可以推测，在观察结膜充血、粘液症、注射部位红斑和注射部位反应时，所有其他高一致性值实际上是由于该物种长期用于检测的特殊用途所致。皮肤和眼睛的耐受性。另一方面，高TP和IR+值证实了该物种对这些终点的预测值。在人类和动物（所有物种组合）中频繁发生的事件，具有统计学意义，但阳性似然比较低，即预测值较低。

结论：使用大数据方法是一种强大的分析方法，它产生了大量的分析结果，这些结果本身没有先入为主的概念。我们强调了与最高的TP数字的一致性以及正可能性。最一致的观察，即心脏效应，如QT延长和心律失常，与先前发表的报告一致。我们的结果有助于回答引言中提出的问题：在动物身上进行生物医学观察对人类风险有什么意义？涉及的人类风险对于每对动物事件都是独一无二的。一般来说，许多翻译被证实是可预测的，如QT延长和其他心律失常。其他两组具有统计学相关性，但预测值较低。这项研究为每一个事件-物种对建立了一个指标，可以用来预测人类的危险。翻译是否取决于使用的动物种类？在某些情况下，已经在几个物种中进行了观察。然而，由于每个试验所用的特定物种都有偏差，因此从生物学的角度来看，无法直接确定哪种物种是最好的。然而，可以从TP和IR+得出结论，这些特定终点的物种选择可产生预测能力，证实动物研究对人类风险评估的价值。是否可以通过观察高于首选术语水平的更高级别观察分组来增加一致性？只有在少数情况下，更高级别的分组才能产生有意义的结果。随着数据分组到更高的级别，统计显著性往往会降低。许多医学上显著的一致性被证实具有很高的意义和预测能力，特别是心脏事件。

来源：Regulatory Toxicology and Pharmacology Volume 96, July 2018, Pages 94-105



地址：北京市朝阳区潘家园南里5号（100021）

电话：010 - 67776816 传真：010 - 67781534 E-mail: calas@cast.org.cn

技术支持：山东瘦课网教育科技股份有限公司

| 站长统计

