调整性别相关的偏倚

数据分析的另一种方法

在发育过程中，激素发挥了调节各组织形成和功能的作用。男孩和女孩的不同激素背景意味着他们对内分泌干扰物（EDCs）的易感性可能不同。正如《环境与健康展望》（*EHP*）上发表的一项新研究所强调的那样，混杂因素和结局之间的关系本身就可能会因性别而异，如果不能说明这一点，可能会导致错误的效应估计。该研究的作者提出了一种解决这个问题的新方法。

“如果你认为暴露可能会因为性别不同而产生不同的结局，那么你应该考虑那些可能同样因性别而异的混杂因素，并对此作出解释，”约翰霍普金斯大学布隆伯格公共卫生学院的助理教授 Jessie Buckley 说。她是这项新研究的合著者。

研究人员通常使用两种方法的其中一种来确定暴露与结局之间的关联是否因性别而异。第一种方法是分层，即针对男性和女性使用不同的模型来估计性别特异性关联。第二种方法是使用带有性别相关暴露“乘积项”的单一模型，该模型允许估计的关联在男性和女性之间有所不同。

这两种方法产生的关联估计通常被假定为是基本相等的。然而，由于乘积项模型不适用于性别相关的混杂，情况并非总是如此。

在这项新研究中，作者提出了一个强化的乘积项方法。这个替代模型允许女孩和男孩的暴露 - 结局关联有所不同。不寻常的是，它也允许混杂因素（既是与感兴趣的暴露相关的因素，也是结局的独立原因或预测因素）因性别而异。

为了检验替代模型得到的估计结果是否比传统方法更准确，研究人员模拟了一个简单的 EDC 假设研究的数据，并创建了各种情景，其中混杂因素和 EDC 暴露的效应在女孩与男孩之间存在差异。然后，他们分别使用了这三种模型来估计假设的暴露与健康结局之间的性别特异性关联。

因为作者使用了模拟数据进行分析，所以他们可以将每种模型的性别特异性关联的估计值与“真实的”暴露 - 结局关联进行比较，这个“真实值”在他们为不同情景生成数据时就已经确定了。比较的结果表明，分层模型和强化的乘积项模型几乎在所有情况下都能准确估计出性别特异性效应，而传统的乘积项模型在 8 种情景中有 5 种表现较差。

接下来，研究人员将这些模型应用于他们早期在男孩和女孩中进行的关于产前邻苯二甲酸酯暴露研究中的数据。他们发现，传统的乘积项模型（不支持性别特异性混杂）得到的估计值与使用分层和扩充乘积项模型得到的估计值不同。

布朗大学流行病学助理教授 Joseph Braun 说：“不同的方法在分析模拟数据和实际数据时的确会产生不同的结果，但仅存在细微的差异。”然而，他也指出这种差异是值得关注的。“当你将其他的无法解释的事情（比如选择偏倚）一起考虑进去，你可能会漏掉一些东西或者得到错误的答案，”他说。Braun 没有参与这项研究。

男孩和女孩对某些环境暴露作出的反应存在性别特异性。一种分析研究数据的新方法可以帮助研究人员更准确地识别这些性别特异性反应。
Image: © monkeybusinessimages/iStockphoto.

该研究的作者建议，未来的 EDC 研究需要考虑混杂因素是否可能对所研究的结局产生性别特异性影响，以及将扩充乘积项方法运用于统计分析以减少得出不准确结论的风险。Buckley 说：“我们只是想确保人们知道他们正在运行的模型的假设条件。当混杂因素对性别相关结局产生不同的影响时，人们使用的两种方法并不等同。”

在 Braun 看来，注意这种等效性的缺乏是该研究的一个特别优势。“这是我见过的第一项已经采取行动去实际比较人们用来检验效应测量改进的两种标准方法的研究，”他说道，“我认为这是科学领域中一个真正的进步，它为我们提供了某些方法是否有偏倚的信息。”

Julia R. Barrett，硕士，生命科学编辑，居住在威斯康星州麦迪逊市，科学作家和编辑。她是美国科学作家协会（National Association of Science Writers）会员和生命科学编辑委员会（Board of Editors in the Life Sciences）的成员。

译自 *EHP* 125(10): 104001 (2017)

翻译：张蕴晖

*** 本文参考文献请浏览英文原文**

原文链接

https://doi.org/10.1289/EHP2473