
有向无环图的概念与基本原则

作者：倪凯文，赵一鸣 来源：临床流行病学和循证医学

本文原地址：<https://www.iikx.com/news/statistics/1673.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

随机对照试验

被视为临床研究中的金标准，是验证某种治疗方案有效性和安全性最可靠的依据，但在试验设计阶段存在许多苛刻条件。另外，如果实际的临床研究问题存在违背伦理道德的风险，研究者将无法开展随机对照试验。因此，观察性试验研究方法经常被采用来探讨暴露(临床干预)与结局之间的因果关系。但是，由于一个或多个潜在的混杂因素的影响，暴露与结局间的真正联系往往被错误的估计。

今天，小编和大家分享一个实用的工具——有向无环图(directed acyclic graph, DAG)，将各变量间的因果关系用直观的示意图表示，帮助大家在试验设计阶段和统计分析阶段更清晰的理解混杂与偏倚以及考虑相关变量的控制。

Judea Pearl教授在1995年首次提出DAG，其概念是用箭头连接节点，一个节点代表一个变量，一个箭头代表一个变量是另一个变量的直接因果效应(暴露 结局)，一个节点出发经过若干条边后无法回到该节点的图形。

接下来，小编用几个示意图让大家更好的理解DAG的基本原则。

箭头从暴露(吸烟)指向结局(肺癌)，表示暴露因素存在一条通路指向结局，即吸烟与肺癌间存在因果关联。

箭头从吸烟同时指向暴露(携带火柴)和结局(肺癌)，表示吸烟与携带火柴间存在因果关联，同时与肺癌情况也存在因果关联，暴露与结局间有一条“携带火柴 吸烟 肺癌”的后门通路，此DAG中的吸烟变量就是我们所熟知的混杂因素。

方框表示该变量被分层或者调整，因为我们已经知道吸烟是该研究中的混杂因素，其被方框包围，表示吸烟因素已被调整控制，因而原来的“携带火柴 吸烟 肺癌”后门路径被封锁，所以我们在此DAG背景下讨论携带火柴与肺癌关系时将不会受到吸烟(混杂因素)的影响。

两个箭头同时指向一个变量将形成碰撞节点，表示糖尿病与住院情况存在因果关联，肝癌与住院情况存在因果关联。由于“糖尿病 住院情况 肝癌”的路径中存在碰撞节点，所以无法形成通路。换句话说，就是我们在大环境下研究糖尿病与肝癌的关系，不受住院因素的影响。

碰撞节点被方框包围，表示该变量被分层，进而导致原本封锁的道路被打开，暴露与结局间形成了一条新的通路。如图，住院情况被方框包围，表示在医院环境内讨论糖尿病与肝癌的关系，由于受到住院情况(入院偏倚)的影响，从而得到错误的因果关联。

以上就是DAG的一些基本原则，大家是不是觉得对于混杂和偏倚有了一个直观可视的认识呢?在今后，小编会用DAG让大家更进一步理解混杂和偏倚以及判断是否需要控制相关变量，让大家体会到DAG对于我们临床研究设计与分析的利用价值。

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发