

随机实验和观察性研究中因果推理的基本概念： 一个非凡的历史

唐纳德B. 鲁宾

清华大学杨数学科学中心教授，清华大学高级研究员，福克斯/寺庙大学名誉教授

因果推理的序言

- 我的介绍
 - Physics – John Wheeler 1961
 - Psychology & consciousness – Julian Jaynes 1964
 - Experimental design – William Cochran 1968
- 之间的清晰分离
 - 科学=推理的对象
 - 如何学习来了解科学
 - 采取干预措施，在某个时间点测量各个方面
- 同样的科学符号/表示，无论我们如何尝试学习或衡量它
- 缺少数据总是存在的
 - 无法及时地回到过去
 - 海森堡不确定性原理与观察者效应

Fisher (1925) -实际上是随机化的

- 从而在预期中的所有预处理变量上创造平衡
- 重新建议=如果不平衡重新随机化
 - 理论和应用最终被追求-计算成为可能
- 假设再随机化以进行评估-基准分布
 - 由尖锐零假设相矛盾得到的随机证明
 - 后验预测p值的特例(Rubin, AoS1984)
- Fisher, 1918年对非零因果效应的隐含定义):

如果我们说, “这个男孩长得很高是因为他吃得很好,” 我们不仅是在追踪个体的因果关系; 我们暗示他可能吃得更糟, 在这种情况下, 他会更矮。
- 但是, 尽管在基于对称参数的尖锐零假设下, 在基于对称论证的尖锐零假设下, 并没有巨大的几何见解

Neyman (1923) – 潜在结果

- 在随机实验中定义估计为N个单位的潜在结果的函数
 - $Y(0)$ = 治疗下的潜在结果阵列0
 - $Y(1)$ = 治疗下的潜在结果阵列1
 - 等等，隐含地假设SUTVA: $Y_i(w)$ (i 、 w) 的功能
 - 不能在任何一个单元上观察到“海森堡”
- 评估随机化分布上程序的操作特性（如估计的）
- 非加性单位级因果效应的作用

- 但后来在1923年否认了深入的理解

对这些见解的评论

- 20th世纪的洞察力，就像量子力学中一样
 - 估计是用可测量的量来定义的，即使在理论上，它们也不是同时可测量的
- 费雪真的是第一个吗？ 皮尔斯，19年末th世纪
- 随机对照试验很快主导了整个英联邦和美国的农业和动物繁殖(FisherDOE1935)
 - 更多的应用工作(例如，肯普索恩，科克伦&G考克斯，Box，D考克斯)
 - 支持数学工作（例如，ISI数学学，Bose、Nair、Rao）
- 随后，随机对照试验进入了西方工业
 - 二战后：（例如，1951年日本戴明奖章，为质量控制）

- 但这些见解仅限于具有无意识单位的随机对照试验

将洞察力运输到具有有意识单位的rct

- 医学
 - 1946年，英国MRC&希尔-斯特雷普
 - 1954年在美国进行的索尔克疫苗随机对照试验
 - 美国FDA和制药公司-Paul Meier 20世纪50年代
 - 过分热心地坚持ITT来评估分配的效果，而不是分配和接受药物的效果，而不是使用
 - 但在非rct中没有使用Fisher/Neyman的见解或符号
 - E. g. , 1964年美国卫生局局长关于吸烟和肺癌
 - Epi、econ、社会科学都使用回归和潜在结果代替观察结果和指标 w_i 对于每个单元
 - $$\frac{1}{W} \sum_i Y_i^{obs} = \sum_i w_i Y_i(1) + (1 - \sum_i w_i) Y_i(0)$$
 违反了分离的原则
- 科学从我们所做的事情来了解科学

RCM (1974、1975、1976、1977、1978)

- 潜在的结果定义了所有情况下的因果估计，而不仅仅是在随机实验中
 - Neyman不同意（“过于推测性”，没有随机化）
- 因果推理所需的分配机制：治疗指标 W 的概率 $\Pr(W|X, Y(0), Y(1))$ ，一般依赖于 $Y(0)$ 和 $Y(1)$
 - (1) 未混淆的= $\Pr(W|X)$ ，如rct
 - Ignorable= $\Pr(W|X, Y_{\text{obs}})$ ，如顺序rct
 - $Y_{\text{obs}}Y(0)$ 和 $Y(1)$ 的=观测值
- Bayes=除了建立分配机制之外，还建立了科学($X, Y(0), Y(1)$)的模型，从而产生缺失和观察到的潜在结

果；艺术触摸是需要的

因果推理面临的基本问题 (Rubin, 1975) 这是一个缺失的数据问题！

- 对于每个 i ，只有 $Y(1)$ 或 $Y(0)$ 可以观察到

		$Y(1)$	$Y(0)$	W
单位	1	✓	?	1
	.	✓	?	1
	.	✓	?	1
	.	?	✓	0
	.	?	✓	0
	.	?	✓	0
N				

- 随机分配主动和对照 \Rightarrow 代表样本的 $Y(1)$ 将与 Y 的代表性样本进行比较 (0) - 显而易见吗？
- 必须建模分配机制，以推断缺失的潜在结果

回归Y的“错误” obs, i 在 W_i 和 X_i

- 在使用观察值符号时，失去潜在的结果和关键的Fisher/内曼概念

$$Y_{obs,i} = W_i Y_i(1) + (1 - W_i) Y_i(0)$$

- 混合分配机制和科学！
- 抑制关键见解，不丢失数据！
 - 该估计什么？ 必须是参数吗？？！
- 生物统计学、经济学、流行病学的标准，无处不在！！
- 即使是伟大的统计学家和流行病学家（如费舍尔、科克伦、康菲尔德、……）也把自己与观

察值符号混淆了

9

对观察性研究的不良影响

—没有设计阶段，分析混乱

- 关联与因果关系之间的关系，混乱的
- E. g.， 病例-非病例（对照）研究
 - 这些方法使用的抽样机制是混淆的，因为根据定义，它们依赖于 观察到的潜在结果
- 试着用回归来“修复”，但没有任何原则

*“我们现在考虑观察性研究可以支持的推论类型，无论是前瞻性的还是回顾性的，以及那些可以由实验研究支持的推论。这种区别似乎是不可否认的，但其确切的性质却难以捉摸。” 康菲尔德，1959年，
《研究原理》*

启动结论

- 保留来自过去的见解
- 避开了过去的混乱
- 意识到RCT的想法是非常新的
- 更新针对设计和分析的统计方法，以利用现代计算环境
- 鼓励数学精度，特别是在符号和逻辑流程中
- 精度可能会对具有挑战性的应用产生关键后果，例如，安慰剂效应

一些新的一般想法

- 设计研究的倾向得分，包括实验性和观察性R&R（1983…）
- 在实验中重新随机化，以避免“不幸”的分配，并作为观察性研究M&R的扩展模板（2010……）。
- 处理并发症的主要分层，如不遵守指定治疗的R&T（2002）和安慰剂效应R（2020）