

· 统计、测量与方法 ·

实证研究中的因果推理与分析*

温忠麟^{**1,2}(¹ 华南师范大学心理应用研究中心 / 心理学院, 广州, 510631) (² 香港考试及评核局, 香港)

摘要 心理学期刊中的实证研究论文, 很多时候都在检验变量之间的因果关系, 但学界对因果研究存在一些不同的看法。本文试图回答下面问题: (1) 实验中不能操纵的变量是否可以作为原因? (2) 非实验研究能不能检验因果关系? (3) 因果分析 (尤其是中介分析) 是否一定要使用追踪数据? 追踪设计的主要目的是什么? 通过因果要素和因果推理逻辑的辨析, 对前面两个问题都给出了肯定的回答, 并讨论了判明变量先后顺序、统计控制无关变量的方法; 为了回答问题 3, 厘清了追踪设计在因果分析中的作用: 区分变量的先后顺序、有效获得历时性的影响结果, 但即时性的因果影响采用追踪设计可能是不合适的。

关键词 因果关系 时间顺序 恒定关联 实验 问卷 变量控制

心理学基于实验和问卷的实证研究论文, 通常都试图探究因果关系 (causality)。通过实验去检验因果关系顺理成章, 但基于问卷的因果研究, 经常会碰到两方面的问题。一方面, 有的研究欠缺了因果分析中的某些环节, 得到的因果关系似是而非。常见的情况是, 变量之间谁是因谁是果没有理论根基或没有经验支持, 研究者根据自己的设想就画出模型路径图进行统计分析得出结果。不妨把这方面的问题说成“程序不足”, 这是造成问卷研究做因果分析屡遭质疑的主要原因。另一方面, 有人囿于方法成见, 对基于问卷数据得到因果关系的做法一律持否认态度, 认为变量之间只是相关关系而不是因果关系。不妨把这方面的问题说成“方法偏见”。例如, 有人认为像“性别”这样不能操纵的变量都不能作为原因; 有人认为基于问卷做回归、中介、调节分析的研究都只是考察变量之间的相关关系, 不是因果关系; 有人认为用横断 (cross-sectional) 数据验证中介作用在逻辑上都有缺陷。

本文试图回答下面的问题: 实验中不能操纵的变量是否可以作为原因? 非实验研究能不能检验因果关系? 因果分析 (尤其是中介分析) 是否一定要采用追踪 (longitudinal) 数据? 追踪设计的主要目的是什么? 通过回答这些问题, 可以一举拆解“程序不足”和“方法偏见”两方面的问题。本文首先介绍因果要素和因果推理; 接着回顾实验研究中的

因果逻辑; 然后讨论问卷研究中如何落实因果要素, 包括如何区分变量先后和进行统计控制; 最后就中介分析与追踪研究的作用做出说明, 并归纳本文结论。

1 因果要素与因果推理

因果关系, 是指事件的原因与结果的关系。原因和结果是一对共生的概念, 原因是指引起结果的现象 (或属性), 结果是指由于原因的作用而引起的现象 (或属性)。早期对因果关系说得比较清楚的是哲学家休谟, 他认为因果需要满足三个要素: 第一是原因和结果事件之间在时空上毗连 (contiguity), 第二是时间顺序 (succession), 第三是必然联系 (necessary connection) (Hume, 2005/1739; 彭玉生, 2011)。

1.1 “时空上毗连”可以忽略

在时空上毗连是因果要素之一, 直观地说就是, 两件八竿子打不着的事情之间是没有因果关系的。尽管有蝴蝶效应的说法存在, 但“一只蝴蝶在巴西轻拍翅膀, 可以导致一个月后德克萨斯州的一场龙卷风”就不算是因果了。不过, 怎样才算“毗连”其实是很模糊的, 未必能从时空上判定, 而需要对概念背景进行深切了解才能判定。例如, 宗族网络 (或宗族势力) 与计划生育 (彭玉生, 2011), 表面上看

* 本研究得到国家自然科学基金 (31271116) 的资助。

** 通讯作者: 温忠麟。E-mail: wenzl@snu.edu.cn

DOI:10.16719/j.cnki.1671-6981.20170130

来不毗连，但如果了解“宗族的核心价值是生儿育女、延续香火”，两者就是毗连的了。心理学研究中，所考虑的因果（变量）几乎都被试的各种心理特质以及人口学变量（如性别、年龄等），这些在同一个被试上的不同变量，说它们毗连相信不会有异议，所以时空上毗连与否可忽略不论。

1.2 先因后果是逻辑上的要求

时间顺序的意思很明确：先因后果。从时间角度说，可以将因果效应分为两种，一种是即时性（real-time）的因果作用，另一种是滞后的（lagged）、历时性的因果作用。虽然这种区分难以从时间上严格界定，但在社科研究中对于设计合适的实证研究的类型（横断研究还是纵向研究）具有参考意义。看看因果即时作用的两个例子，一是康德的“放置铅球在软垫上成窝状”，窝状的形成在放置的动作后即时发生，很难观察到时间先后；二是“高消费满足身份表达”，高消费与身份表达其实是同一行为的两个方面，几乎没有时间上的先后之分。这两个例子中的先后顺序更多是概念上的先后（王天夫，2006）。至于历时性的因果作用，则是心理学研究中的常见现象。例如，一项心理干预措施或者一项教育改革实验，需要一段时间后才能看出效果。

有人认为先因后果不必要，有教科书还举例说，过节前的购物高潮，结果就出在原因之前，这显然是误解了因果变量。有人为了苗条而节食，节食当然不是苗条的结果；同样道理，为了过节而购物，购物也不是过节的结果。先因后果应当是逻辑上的要求。

如何知道谁先谁后？这在实验研究中不是问题，实验因素的操纵在先，然后才有实验结果。在追踪设计的问卷研究中，谁先谁后通常也没有问题。剩下的问题是同一次的问卷中，如何确定变量的先后顺序？将在3.1节专门讨论。

1.3 “必然联系”的可操作化演绎

必然联系是超越感官经验的，实践中无从检验和判断现象之间是否有必然联系。取而代之的是所谓恒定关联（constant conjunction, Hume, 2005/1739）。恒定关联的意思是两个现象相伴而生：总是同时出现或者同时不出现。通过有限的观察，加上逻辑推理，就可以推论现象之间是否恒定关联了。通常使用的是密尔（也译为穆勒）逻辑，通过有限的观察归纳出两个现象恒定关联。后面将会看到，单就恒定关联这个要素，就衍生出两个要求：一是因果要共变（相

关是其特例），二是无关变量要控制。

1.4 因果推理的密尔逻辑

密尔逻辑是推论变量之间恒定关联的方法，包括求同法、求异法、剩余法、共变法（Mill, 2002/1848）。在知道了变量之间的先后顺序后，密尔逻辑也就成了推论变量之间因果关系的方法。方便起见，下面直接使用带箭头的因果推理的叙述。

求同法 如果所研究的两个或两个以上的场合（也称为案例）只有一对现象（如A和a）是共同的，则这一对现象恒定关联。简单点说，如果两个现象总是同时出现，则恒定关联。推理形式：

已知： $ABC \rightarrow abc$ （场合1）；
 $ADE \rightarrow ade$ （场合2）；
 $AFGH \rightarrow afgh$ （场合3）；
 ...

则有： $A \rightarrow a$ （ $\because A, \therefore a$ ）

求异法 比较某现象出现的场合和不出现的场合，如果这两个场合除一个不同外，其他条件都相同，那么这个不同的条件就是这个现象的原因。简单点说，如果两个现象同时出现或者同时不出现，则恒定关联。推理形式：

已知： $ABC \rightarrow abc$ ；
 $-BC \rightarrow -bc$ ；

则有： $A \rightarrow a$ （ $\because A, \therefore a$ ）

剩余法 如果某一复合现象已确定是由某种复合原因引起的，把其中已确认有因果联系的部分减去，那么，剩余部分也有因果联系。推理形式：

已知： $ABC \rightarrow abc$ ；
 $B \rightarrow b$ ；
 $C \rightarrow c$ ；

则有： $A \rightarrow a$ （ $\because A, \therefore a$ ）

剩余法比较少用，因为前提太强，除了一对现象外，其他因果关系都要预先知道才行。

共变法 在其他条件不变的情况下，如果某一现象发生变化另一现象也随之发生相应变化，那么，前一现象就是后一现象的原因。推理形式：

已知： $A_1BC \rightarrow a_1bc$
 $A_2BC \rightarrow a_2bc$
 $A_3BC \rightarrow a_3bc$
 ...

则有： $A \rightarrow a$ （ $\because A, \therefore a$ ）

共变法包含了两个要求，一是因果共变（A与a共同变化），二是变量控制（除了A以外，其他

条件保持不变)。不难看出,求异法是共变法的特例,因为出现与否是极端的变化。

1.5 小结

休谟所指的因果,要求两个现象总是同时出现或者同时不出现,是一种狭义的因果:一果一因。密尔逻辑的求同法和求异法并用,得到的因果满足一果一因。共变法则允许一果多因,因果的外延扩大了很多,这对心理学和其他社科研究很重要,因为其中考虑的原因,多数时候都只是部分原因,不是唯一原因。

2 实验研究的因果逻辑

2.1 实验研究的逻辑基础是共变法

求同法要求“其余条件都不同”;求异法和共变法则要求“其余条件都相同”。在自然条件下,很难出现“其余条件都不同”或者“其余条件都相同”这样的理想状态。在实验室里,通过人为操纵和干预,则有可能做到尽量接近这样的理想状态。而且,控制“其余条件都相同”比“其余条件都不同”容易办到,因而通常的实验都是通过控制使得“其余条件都相同”。这样,实验研究的因果推理的逻辑基础通常是共变法(包括其特例求异法)。

在理化实验中,要控制使得“其余条件都相同”比较容易。在生物实验中,控制起来比较难,因为被试会存在各种差异(如种群、地域等)。在心理学和其他社科研究领域,控制起来就更难,因为被试经常是人,除了生理差异外,还会有教育程度、宗教、文化背景等的差异。不过,很多条件就算不同,对结果的影响也无足轻重,因而实验中只要控制对结果可能会造成较大影响的条件就够了(即局部控制),其余次要条件造成的影响,都当做误差,并通过所谓随机化,降低误差的影响。在实验研究中,需要控制的变量也称为无关变量,所谓无关变量是指与因变量有关、而与实验目的无关的变量。

心理学和其他社科研究还有一个特点,原因对于结果的影响往往不具有必然性,而是具有或然性,就个案来说,原因对结果未必有确定性作用,作用反映在集体层面,呈现出集体规律。相应的,表述形式也是非决定性的。与“若A出现,则a必然出现”的论断相比,更为常见的是“若A出现,a出现的概率增加了”。很多时候这种概率差异,会从平均值上表现出差异。其实生物学和医学上的许多因果关系也有或然性。

Fisher (1935) 提出了实验设计应遵循的三个原则:随机化、局部控制和重复。从上面的讨论可知,随机化和局部控制都是为了使得“其余条件都相同”,重复实验不仅是考察集体性规律的需要,也是减少误差的需要。如何实现随机化、局部控制和重复,这是实验设计教科书上会有内容,这里不拟赘述,下面看看实验研究的因果逻辑。

2.2 差异分析检验共变与否

先看实验组控制组的对比实验,其因果推理的逻辑基础是求异法。实验组有实验处理A,结果有a;控制组无实验处理A,结果无a。因为因果的或然性,通常是比较并检验实验结果的均值差异。如果差异是统计显著的(就是通常的 $p < .05$)并且是实质性的(通常由相对差异或者效应量大小来判断),则推论A影响a。

再看单因素多水平(多组比较)的实验,此时一个水平就是一个处理,其因果推理的逻辑基础是共变法。怎么知道实验结果是否随着实验处理在变化呢?只要假设所有实验处理下的因变量均值相同,做方差分析进行检验。如果拒绝了原假设,说明共变了,如果这种共变导致实质性差异,就可以推出实验因素影响结果。

最后看看多因素实验,其中可能会影响因变量但研究者不感兴趣的因素会被作为区组看待,实现了局部控制。在统计的时候,将这些区组的作用分离出来,就可以知道感兴趣的因素对结果的影响了。可以说,实验设计的各种做法,其实都是试图做到或者接近密尔逻辑的基本要求。

2.3 小结

或然性因果,即集体规律上的因果,在一果多因的基础上再次扩大了因果概念,也是心理学和其他社科研究中通常所说的因果。在因果概念扩张的过程中,随着外延扩展的是对因果关联要求上的弱化。

3 问卷研究中的因果推理

这一节将讨论问卷研究如何尽量做到符合因果要素和密尔逻辑。

3.1 问卷研究中变量先后顺序的确定

前面说过,在基于一次性问卷数据的研究中,如何确定变量的先后顺序是一个问题。有一些简单的原则可以帮助建立社会现象之间的时间顺序(王天夫, 2006; Davis, 1985):

(1) X 已经停止变动之后, Y 才出现。如结婚时年龄与是否离婚。

(2) 一个常见的事件系列中, 与 Y 相比, X 是其中较为靠前的事件。如学历与工资收入。

(3) X 从来都不变化, 而 Y 则有变化。如成长地与成年后的食物口味。

(4) 比较而言, X 相对较为固定和不易变动。如工作职位与收入。

上面的 (1) 和 (2) 已经有明显的先后顺序, (3) 和 (4) 其实就是说 X 比 Y 稳定。温忠麟和叶宝娟 (2014) 总结得比较到位而简洁, 可以通过看两个变量的属性决定先后: 如果 X 是比 Y 更加本质的 (或者是长久的、稳定的、客观的) 属性, 则 X 在 Y 的前面。

心理学研究中的常见变量, 特质变量 (如智力、可遗传的性格特征) 很稳定、很难改变; 类特质变量 (如大五人格、美德) 次之; 类状态变量 (如心理资本、幸福感) 比较容易改变和开发; 状态变量 (如积极情绪) 则很不稳定、很易改变 (柯江林, 孙健敏, 李永瑞, 2009)。这样, 依稳定性高低可以将变量类别排序: 特质、类特质、类状态、状态。这种稳定性排序为区分变量顺序提供了一定的理据, 而性别和年龄与心理学研究中的其他变量比较, 总是最靠前的变量。

区分变量顺序的一个方法是颠倒两个变量的顺序, 看哪一个更容易解释。例如, 家庭收入影响支出, 而不是相反; 收入影响幸福感, 而不是相反。有时候, 两个变量相互促进, 互为因果, 例如健康状况与幸福感。如果不能用上面方法分清先后, 可以通过所谓的交叉滞后设计 (白学军, 2012; Kenny, 1975), 获取历时性的数据, 看看两个变量究竟谁影响谁, 或者谁影响更多一些。交叉滞后设计是一种追踪设计, 详见第 4 节。

3.2 相关显著强于差异显著

考虑完因果顺序, 现在考虑因果共变。基于问卷数据的研究, 相关分析往往多于差异分析。两个变量相关显著是比差异显著更强的结果, 例如小学生身高与年龄相关显著, 说明不同年龄的身高差异显著。但反之不一定成立, 因为通常所说的相关是线性相关, 有时候差异显著也不一定相关显著。但在有一个变量是二分变量的情形中, 相关 (点二列相关) 显著与差异显著是等价的 (温忠麟, 2016), 例如身高与性别相关显著等价于男女的身

高差异显著。许多文献和教科书都强调, 相关是因果的必要条件, 如果把相关理解为可以是非线性相关, 这个说法是对的。但比较清晰的说法是, 关联是因果的必要条件。无论是相关显著还是差异显著, 都是在推证 “关联” 而已。

3.3 变量的统计控制

除了因果共变, 共变法还要求变量控制。与实验相比, 问卷中的变量控制得较少, 弥补的办法是使用统计方法进行控制。这里以自变量、因变量以及无关变量都是连续变量的情形为例, 说明非实验数据如何控制无关变量, 为避免一些无关要旨的细节, 本节假设所有变量都是标准化变量 (因而变量之间的关系可以用皮尔逊相关系数来反映), 并且假设已经明确自变量在前、因变量在后。这里只考虑只有一个无关变量的情形 (多个无关变量情形类似考虑), 依据无关变量在 $X \rightarrow Y$ 关系中的位置进行讨论。无关变量 Z 的可能位置如下: (1) 在 X 前面, (2) 与 X 并列, (3) 在 X 之后、 Y 之前。而与 Y 并列或在 Y 之后的变量, 都不会影响 $X \rightarrow Y$ 关系, 不用考虑。

3.3.1 无关变量是自变量的前置变量

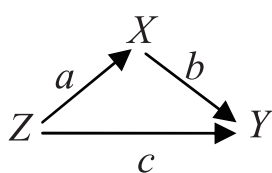
如图 1(1) 所示, 无关变量 Z 在 X 之前出现, 这时候 X 与 Y 因为有共同的起因 Z , 可能存在虚假效应 (侯杰泰, 温忠麟, 成子娟, 2004)。有文献称此种情形为第三变量问题 (third-variables problem, 例如, Goodwin, 1995)。 $X \rightarrow Y$ 的总效应有如下的效应分解公式:

$$r_{YX} = b + ac \quad (1)$$

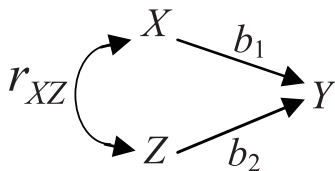
其中总效应 r_{YX} 是 X 与 Y 的相关系数, a 是 X 对 Z 的回归系数 (等于 r_{XZ}), b 和 c 分别是 Y 对 X 和 Z 的回归系数 (对应的回归方程为 $\hat{Y} = bX + cZ$)。总效应 r_{YX} 分解为两部分, 第一部分 b 是 $X \rightarrow Y$ 的直接效应, 第二部分 ac 就是虚假效应。

直接效应 b 的含义是, 在控制了前置变量 Z 后, X 变化一个单位, Y 将变化 b 个单位。如果 b 显著, 并且有实质性意义, 则可以推论 X 是引起 Y 变化的一个原因。可以用 “ R^2 的变化” 作为效应量, 就是在回归方程中已经包含 Z 的条件下, 增加变量 X 时 R^2 的变化。 R^2 的变化衡量了在控制 Z 的条件下, X 对 Y 的变异的解释的比例, 正好就是部分相关系数 $r_{Y(X.Z)}$ 的平方 (温忠麟, 2016)。因为通常会认为 R^2 的变化超过 2% 才有实质性意义, 所以 $r_{Y(X.Z)}$ 达到或者超过 0.15, 才有实质性意义。

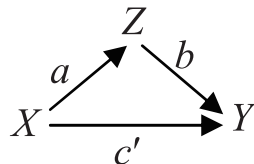
当 X 和 Y 有共同起因时, 需要排除共同起因引起的虚假效应, 才能确定 X 与 Y 是否还有关联。但不能说有共同起因的两个变量因果关系就不成立了, 还要看直接效应才能做出判断。如果在考虑了第三变量后直接效应还显著且有实质性意义, 则因果关系存在。



(1) 无关变量是前置变量



(2) 无关变量是并列变量



(3) 无关变量是中介变量

图 1 无关变量的位置与统计控制模型路径图

3.3.2 无关变量与自变量并列

与自变量并列的无关变量也应当在模型中作为控制变量, 参与统计分析。如图 1(2) 所示, 因变量 Y 除了受到 X 的影响外, 还受到无关变量 Z 的影响。通常的做法是做 Y 对 X 和 Z 的回归分析, 得到回归方程为 $\hat{Y} = b_1X + b_2Z$, 如果 b_1 显著 (不为零), 并且有实质性意义, 则可以推论 X 是引起 Y 变化的一个原因。参数 b_1 的统计意义和 3.2.1 节的 b 一样, 效应量也类似计算。 $X \rightarrow Y$ 的总效应有如下的效应分解公式:

$$r_{YX} = b_1 + b_2r_{XZ} \quad (2)$$

总效应 r_{YX} 分解为两部分, 第一部分 b_1 是 $X \rightarrow Y$ 的直接效应, 第二部分 b_2r_{XZ} 称为未分解效应 (侯杰泰等, 2004), 是因为存在相关的自变量引起的效应。效应分解式 (1) 和 (2) 在数量上是一样的, 但后半部分的含义不同。

3.3.3 无关变量作为中介变量

如图 1(3) 所示, Z 成了 X 与 Y 之间的中介变量 (mediator)。这时, 可以按中介模型那样进行分析, 结果可以知道 X 对 Y 的影响机制, 即 X 是如何影响 Y 的 (温忠麟, 叶宝娟, 2014)。按通常中介模型中的参数符号, $X \rightarrow Y$ 的总效应有如下的效应分解公式:

$$r_{YX} = c' + ab \quad (3)$$

其中 a 是 Z 对 X 的回归系数 (等于 r_{ZX}), b 和 c' 分别是 Y 对 Z 和 X 的回归系数 (对应的回归方程为 $\hat{Y} = bZ + c'X$)。总效应 r_{YX} 分解为两部分, 第一部分 c' 是 $X \rightarrow Y$ 的直接效应, 第二部分 ab 就是中介效应。

从数量上看, 效应分解式 (3) 和前面的 (1) 和 (2)

看两个例子, 被试都是 6~12 岁小学生。例 1 中, 被试智商测验成绩与身高相关, 在排除共同的起因年龄后, 两者不再相关 (同年级学生智商测验成绩与身高无关); 例 2 中, 被试体重与身高相关, 在排除共同的起因年龄后, 两者仍然相关 (同年级学生的体重仍然与身高相关)。

是一样的, 第一部分都是 $X \rightarrow Y$ 的直接效应, 但此处含义已经不同。即使在控制了中介变量 Z 后, c' 不再显著, 也不会改变 $X \rightarrow Y$ 的因果关系及其大小, 只是直接因果和间接因果大小变化而已。也就是说, 中介变量的存在, 不影响因果关系及其大小。而且, 对一般的因果, 都可以设法引入中介变量, 形成因果链。例如, 酒驾增加交通意外, 自变量是血液中的酒精浓度, 因变量是交通事故频率, 中介变量可以是反应时间、行车速度等, 醉酒导致反应时间变慢、车速加快, 进而导致车祸频发。但无论中介变量有几个、中介效应有多大, 都改变不了酒驾增加交通意外的因果事实。

3.4 小结

上述讨论可以帮助理解社科中比较经典的推论因果关系的三个准则 (Cook & Campbell, 1979): (i) 因在果之前发生; (ii) 因和果共同变化; (iii) 排除因果联系外的其他解释。

当无关变量是自变量的前置变量或者与自变量并列时, 需要作为控制变量, 习惯上将它们统称为协变量 (covariate)。例如, 要研究外貌吸引力对爱情的影响, 理论上说空间距离、相似性和熟悉度都是需要加以控制的协变量。问卷中除了感兴趣的因果变量外, 还应当包含可能影响因变量的协变量, 并进行统计控制。一般地说, 被试性别和年龄等人口学变量, 如果与自变量相关, 则需要控制; 如果与自变量不相关, 由公式 (1) 和 (2) 可知, 无需控制, 说明一下便可。

除了将控制变量作为协变量放入模型外, 对变量进行控制的简单方法是分别做自变量和因变量对控制变量的回归分析, 用所得的残差作为变量值进

行后续分析。例如要分析身高对体重的影响,控制变量是性别(0~1编码)和年龄。做身高对性别和年龄的回归,得到的残差值作为身高的观测值;做体重对性别和年龄的回归,得到的残差值作为体重的观测值。然后做体重对身高的回归,就是控制了性别和年龄后的结果。

自变量的前置变量和并列变量都有可能是因果关系的调节变量(moderator)。至于引进中介变量,可以知道因果如何发生作用,但因果关系不会受到中介分析的影响。有关中介变量与调节变量的区别,可参考有关论著(温忠麟,刘红云,侯杰泰,2012)。

控制变量的道理在实验情形也是一样的,实验中需要控制的无关变量,也是自变量的前置变量或者并列变量。无论是问卷还是实验,都不能控制所有的变量,例如性别差异的实验,除非使用龙凤胎的孪生子女进行实验,否则总有变量未能控制。但通常的研究并没有使用龙凤胎,正如彭玉生(2011)指出的那样,“常规统计分析中只能量力而为,能控制多少是多少。定量研究有一条不成文的行规:只需要控制那些已知的相关因素,次要并与核心变量无关的因素归入随机残项;至于可能重要而未知的因素,留给后人去发现。”学术共同体的这种共识很重要,否则对任何研究都可以挑刺或者否定。

4 中介模型与追踪研究

有研究者认为横断研究不能证明因果关系(Cole & Maxwell, 2003),横断研究中的中介效应是有偏的(Maxwell & Cole, 2007; Maxwell, Cole, & Mitchell, 2011)。国内外都有建议使用追踪数据进行中介效应分析的呼声(杜岸政等,2014;甘怡群,2014;Preacher, 2014; Taris & Kompier, 2006)。

简单的中介模型,由三个箭头组成,每个箭头表示一对因果的影响关系。中介效应分析是否应当使用追踪设计,与因果效应分析是否应当使用追踪设计的道理是一样的,所以下面讨论追踪设计对分析因果关系有哪些作用。

4.1 历时性因果需要追踪设计才能有效测量

对于一对因果,如果是即时性的作用,一次性收集数据进行分析是合适的;如果是历时性作用,则要间隔一定的时间收集数据。对于中介模型,如果三对箭头中的因果影响,至少有一对是历时性的,那么中介模型分析就需要追踪设计,获取历时性的

数据。

从因果影响的测量来说,是否需要追踪数据,完全取决于所研究的问题和变量。如果变量之间的影响是历时性的,像发展心理学中的多数变量那样,一个变量(如一种学习方法的训练)对另一个变量的影响(如学业成绩)需要经过一段时间才能起作用,这时追踪数据是必要的。

但是,任何事情都要看实际情形。例如,一年的工作业绩→年收入→年底幸福感,在适当的时候,对这样的因果链条中的所有变量一次性获得数据,足以建立中介模型了。如果每隔一年去追踪调查,就变成了这样子:2014年工作业绩→2015年收入→2016年底幸福感,这已经不是研究者想要的了,这个例子使用追踪数据反而是不合适的。所以,不能一刀切地要求中介研究必须使用追踪设计。

4.2 追踪设计有助确定变量顺序

虽然在第3.1节中讨论了如何在问卷研究中确定谁先谁后,给出了一些准则,但有时候还是难以确定先后。例如,是幼儿母亲抑郁导致教养压力,还是教养压力导致母亲抑郁?在横断数据中就难以确定,这时可以考虑两次测量的追踪设计,就是前面提到的交叉滞后设计。

交叉滞后分析的原理是,如果变量 A 是原因,变量 B 是结果,用下标表示两次测量的结果,那么 A_1 与 B_2 的相关,应该明显大于 B_1 与 A_2 的相关。同时,因为原因相对稳定, A_1 与 A_2 的相关也会大于 B_1 与 B_2 的相关。通过这样的数据分析,有助确定变量顺序。

4.3 追踪设计可以分析变量的发展变化

基于问卷的研究变量,基本上都是之前已经存在的、并非研究过程中引起的变量,即所谓的存在变量(莫雷,温忠麟,陈彩琦,2007)。多个时间点的测量可以分析一些研究变量的发展变化,还可以分析因果关系是否有跨时间的稳定性。不过,不论因果影响是否历时性的,都不合适控制结果变量在上一时间点的测量去分析因果,除非预先知道上一时间点不存在所论的因果影响。

4.4 小结

显然,追踪设计类似于实验中的被试内设计,在某种意义上减少了误差。但通常对因果研究要求追踪设计并不是从这个角度考虑的。在变量先后可以确定、因果影响可以在一个时间点合理地测量的时候,横断研究不仅省事,而且也是正确的做法。

对于即时性因果影响,就算获得了纵向数据,因果配对数据也应当使用同一个时间点数据,而将纵向数据当做被试内设计进行分析。

5 结论

根据前面讨论,回头检视一下文章开头提到的那些问题,已经可以得出清晰的答案。

5.1 实验中不能操纵的变量可以是原因

不能操纵的变量,最典型的是性别,与社科研究中的其他变量相比,性别无疑是比较靠前的。如果能证明性别与某变量恒定关联,就可以推论性别是原因了。例如,选择100对龙凤胎初中生进行普通话测试,结果发现女生成绩显著并且实质性地高于男生,则足以认定性别是影响初中生普通话成绩的一个原因。为了分析影响机制,可以从生理、心理、语言行为习惯和社会规约诸方面进行分析(侯涛,谢旭慧,2008),说明性别为何会影响普通话能力。

还有很多例子,一些不能操纵的变量可以作为原因,如红绿色盲人开车容易闯红灯、多动症影响学业成绩、抑郁症增加自杀风险等。这些不能操纵的变量,可以通过变量选择,成为实验因素。

5.2 非实验研究也可以验证因果关系

问卷与实验相比,差别在哪里?一是问卷中的变量先后顺序有时候不够明确,而在实验中总是明确的;二是问卷中的变量控制没有实验那么严格,因而统计控制更加重要。如果问卷中的变量顺序可以通过一定的方法确定,对无关变量进行统计控制,问卷中常见的存在变量之间的因果关系也就可以验证。

这里以简单的身高和体重(都是存在变量)为例。身高与体重相关高是公认的事实。即使固定种族、性别、年龄、营养状况等等可能同时影响身高和体重的主要条件,还是越高的人往往越重。理论根据是:圆柱体保持底面直径不变,高度越大其体积越大(因而也越重),所以身高与体重恒定关联没有疑问。谁先谁后呢?身高比体重稳定,成年人 体重哪怕增加很多,身高几乎不变。理论上也能说明身高先于体重,重量是与体积密切相关的概念,在一定的物质中两者可以相互转换,而长度在概念上比体积靠前。这都说明身高与体重不是对称关系,而是有方向的,并且是身高影响体重。通过上述分析,可以说身高和体重满足了因果关系的全部要素。再说白一点,如果甲比乙重,不论初始原因是什么,

直接的原因无非是甲比较高、或者比较胖、或者比较结实(比重大),所以身高是影响体重的一个原因(但不是唯一原因)。

还可以举出很多例子,非实验研究可以得到因果关系,例如收入影响幸福感,学业成绩影响自我概念等。如果变量之间的影响关系得到理论、文献或者生活经验的支持,应当可以视为因果关系。

5.3 根据研究需要采用追踪设计

许多期刊在投稿须知中强调横断研究和自我报告等方法的局限性,提倡使用追踪研究、实验研究、非自我报告等,但并非针对中介效应分析。期刊 *Journal of Occupational and Organizational Psychology* 的投稿须知提出,只在特殊情况下(如非常大的有代表性的样本)才可以考虑采用横断数据的研究;而期刊 *Journal of Research in Personality* 的投稿须知甚至指明,采用横断设计基于方便取样得到的问卷数据所写的论文常常会被编辑部直接拒稿而不送审。

虽然所有基于抽样的研究都与样本代表性和样本大小有关,样本容量越大、越具代表性,研究结论越精确和可信。然而,因果关系分析是否需要追踪数据的关键不是样本大小,也不是样本是否具有代表性,而是取决于所研究的问题和变量。如果因果影响是历时性的,则需要追踪设计才能有效测量因果影响;如果因果影响是即时性的,则追踪数据反而可能是不合适的(除非只是当做被试内设计看待)。特别是在管理心理和市场行为等领域,由于行为与心态的即时性关联,可能并不适合追踪设计。例如,品牌负面新闻对品牌信任度的影响,使用相隔半年的纵向数据就是不妥的。设计出因果作用能够得到充分反映的数据收集方法,是提高因果关系真实性的关键环节。在变量先后顺序难以区分的时候,采用两个时间点测量数据的交叉滞后分析是合适的。

5.4 结语

因果分析中,一般地说问卷法是不如实验法严密。但如果提出的因果顺序在常识上、逻辑上和学理上是合理的,而且主要的无关变量也得到控制,那么问卷法也可以验证因果关系。不能一概否定非实验的因果分析。与其强调相关不一定是因果、统计方法(包括结构方程模型)不能确立因果关系,不如强调因果分析需要做什么、怎样建模:模型路径(涉及变量先后)的设定需要有理据,周知的无关变量要控制,统计方法要得当。通过这样的因果

验证，我们就提高了对因果结论的信心。

不过，社科中的因果，通常是一果多因，而且是或然性的，普遍的做法是把原因 X 说成是影响结果 Y 的因素，影响有程度上的不同（与效应量有关），以区别决定性的原因。但“ X 可以预测 Y ”这样的说法就言过其实了，因为多数时候解释率都远低于 50%，预测误差太大，就像地震预测一样，根本谈不上可以预测。

因为密尔逻辑所做的因果推理不是必然推理，所以即使是实验得到的因果也不一定正确。不论哪种方法，都只能在一定程度上实证因果关系，并且需要有不同的研究进行重复验证，因果关系的确立最终还是要经过实践的检验。

参考文献

- 白学军. (2012). *实验心理学*. 北京：中国人民大学出版社.
- 杜岸政，古纯文，丁桂凤. (2014). 心理学研究中的中介效应分析意义及方法评述. *中国心理卫生杂志*, 28, 578–583.
- 甘怡群. (2014). 中介效应研究的新趋势——研究设计和数据统计方法. *中国心理卫生杂志*, 28, 584–585.
- 侯杰泰，温忠麟，成子娟. (2004). *结构方程模型及其应用*. 北京：教育科学出版社.
- 侯涛，谢旭慧. (2008). 师范生普通话水平的性别差异及其成因分析. *上饶师范学院学报*, (1), 68–73.
- 柯江林，孙健敏，李永瑞. (2009). 心理资本：本土量表的开发及中西比较. *心理学报*, 41, 875–888.
- 莫雷，温忠麟，陈彩琦. (2007). *心理学研究方法*. 广州：广东高等教育出版社.
- 彭玉生. (2011). 社会科学中的因果分析. *社会学研究*, (3), 1–32.
- 王天夫. (2006). 社会研究中的因果分析. *社会学研究*, (4), 132–157.
- 温忠麟. (2016). *心理与教育统计*. 广州：广东高等教育出版社.
- 温忠麟，侯杰泰，张雷. (2005). 调节效应和中介效应的比较和应用. *心理学报*, 37, 268–274.
- 温忠麟，刘红云，侯杰泰. (2012). *调节效应和中介效应分析*. 北京：教育科学出版社.
- 温忠麟，叶宝娟. (2014). 中介效应分析：方法和模型发展. *心理科学进展*, 22, 731–745.
- Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, 112, 558–577.
- Davis, J. (1985). *The logic of causal order*. Newbury Park, CA: Sage Publications.
- Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver and Boyd.
- Goodwin, C. J. (1995). *Research in psychology: Method and design*. New York: John Wiley & Sons, Inc.
- Hume, D. (1999/1748). *An enquiry concerning human Understanding*. Oxford, UK: Oxford University Press.
- Hume, D. (2005/1739). *A treatise of human nature*. New York: Barnes and Noble.
- Kenny, D. A. (1975). Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin*, 82(6), 887–903.
- Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, 12(1), 23–44.
- Maxwell, S. E., Cole, D. A., & Mitchell, M. A. (2011). Bias in cross-sectional analyses of longitudinal mediation: Partial and complete mediation under an autoregressive model. *Multivariate Behavioral Research*, 46(5), 816–841.
- Mill, J. S. (2002/1848). *System of logic: Ratiocinative and inductive*. Honolulu, Hawaii: University Press of the Pacific.
- Preacher, K. J. (2014). Advances in mediation analysis: A survey and synthesis of new developments. *Annual Review of Psychology*, 66, 825–856.
- Taris, T. W., & Kompier, M. A. J. (2006). Games researchers play—extreme-groups analysis and mediation analysis in longitudinal occupational health research. *Scandinavian Journal of Work, Environment and Health*, 32, 463–472.

Causal Inference and Analysis in Empirical Studies

Wen Zhonglin^{1,2}

(¹ Center for Studies of Psychological Application / School of Psychology, South China Normal University, Guangzhou, 510631)

(² Hong Kong Examinations and Assessment Authority, Hong Kong)

Abstract Causality is frequently examined in empirical studies in psychology and other social and behavioral sciences. In this paper the following three popular opinions are questioned: (1) A variable is not a cause if it cannot be manipulated in an experiment; (2) A cause and effect conclusion cannot be made from a non-experimental study; and (3) Longitudinal data is necessary for supporting causal inference, especially in a mediation model. The focus is on how researchers can identify causal relations among psychological and social phenomena.

According to David Hume, when we say that “*X* causes *Y*”, we mean that (i) Contiguity: *X* and *Y* must be contiguous in time and space; (ii) Succession: *X* must be prior to *Y*; and (iii) Constant conjunction: there must be a constant union between *X* and *Y*.

‘Contiguity’ is clear in most empirical studies because the research variables are observed from the same individuals of a sample. For experimental and longitudinal data, ‘succession’ is obvious. For a cross-sectional design of study, however, it may be a thorny issue how to find out one variable is prior to another. It is summarized that *X* is much more likely to act as a cause than *Y* if *X* is more essential (or more stable, or more objective, or more long-standing, etc.). In psychological studies, variable categories are ranked by stability as follows: traits, traits-like variables, states-like variables, and states, whereas gender and age precede all the other variables.

According to John Stuart Mill’s methods of investigating causes, ‘Constant conjunction’ leads to two requirements in statistical analysis as a component of causal inference. One of the requirements is concomitant variation between cause *X* and effect *Y* with correlation as the most frequently used numerical character for reflecting concomitant variation. The other is variable control for eliminating the spurious effect when there exists a variable which is a common cause of *X* and *Y*, or removing the unanalyzed effect when there exists a covariate with which *X* influences *Y*. Causal relationship is identified if the direct effect of *X* on *Y* is significant and substantial after eliminating the possible spurious effect and unanalyzed effect.

An existing variable (e.g., gender, age) cannot be manipulated, but by subject selection, it can become a factor in an experiment. If a research variable (e.g., achievement in language learning) is correlated with gender significantly and substantially, gender is a cause due to: (1) Gender is prior to the research variable; (2) Gender is the apposite of sex chromosomes such that no third variable is known. There are lots of examples showing that a non-manipulated variable could be a cause. For instance, children with Attention Deficit Hyperactivity Disorder would have lower academic achievement, and red-green color blind people would be much more dangerous than normal people when driving a car.

Participant’s height and weight are two existing variables that can be measured in a non-experiment study. It is well known that height and weight are highly correlated. Height is more stable than weight since the weights of adults are more likely to change than their height. Conceptually, height is a component of a volume, so height precedes volume which is equivalent to weight. After all known possible third variables are controlled, height and weight are still highly correlated. As a conclusion, height is a cause of weight. There are lots of examples indicating that causation can be analyzed in a non-experiment study. For instance, income may affect subject well-being, and academic achievement may influence academic self concept.

There are three causal arrows in a simple mediation model for analyzing a system of causality. If there is at least one causal arrow where the effect arises sometime after the cause, a longitudinal design will be necessary for effectively observing the causation. When all variables being considered have existing relationship or change promptly one after another within a short time, a cross-sectional design is valid enough for observing the causation if the sequence of the variables is predetermined, otherwise a cross-lagged design is proper for determining the sequence of the variables.

It is important for researchers in psychology and other social sciences to detect or infer causation from common sense, prior experience and academic knowledge. Our confidence about a causal explanation can be enhanced after the causation is examined. Any established causality, however, should be tested in practice.

Key words causality, succession, constant conjunction, experiment, questionnaire, variable control