arXiv:2302.01371v2 [stat.AP] 9 Jan 2024

个性化医疗中“伤害”的观点

Aaron L.Sarvet*，Mats J. Stensrud´洛桑联邦理工学院数学系，瑞士

摘要

避免伤害是个性化医疗及其他流行病学举措无可争议的目标。然而，“伤害”的精确数学表述仍存在争议。本文采用正式的因果语言，探讨几种常见但彼此不同的“伤害”定义。我们阐明，认同某种伤害定义会对决策产生重要的实践和哲学推论。我们还将实践和哲学考量与医学伦理及法律实践中的观点联系起来。

*通讯作者联系方式：Aaron L. Sarvet，数学系，´洛桑联邦理工学院，瑞士。
aaron.sarvet@epfl.ch

1. 引言

个性化医疗旨在利用观测数据，为未来的治疗分配提供个性化依据。当治疗因人而异时，治疗分配取决于患者特征。如果个性化治疗能提高总体的（期望）效用，那么这种做法就是合理的。在某些情况下，个性化治疗经过（生物学上的）设计，以适应已知的患者特征，例如免疫靶向癌症疗法。在另一些情况下，已有多种治疗方案，我们可以根据数据了解每位患者的最佳治疗方案。要找到最佳治疗方案，需要进行因果推断：我们需要确定并进一步估计患者在每种可用治疗方案下的期望效用（Robins，1986年；Richardson and Robins，2013；Murphy，2003；Robins，2004；Tsiatis 等，2019年；Kosorok 等，2021年）。

设定效用函数——将患者的结局及其他特征映射为某个数值——是提出这一任务的基本步骤，因为它构成了判定对治疗方案的偏好之基础。然而，流行病学家和统计学家往往会回避这一步骤，或许是因为它被认为带有哲学性和主观性。尽管如此，决策者和方法学家通常至少会表面上认同经典的希波克拉底格言：“首先，不伤害”。

本文聚焦于对“伤害”的考量。我们以一项新兴文献中的研究结果为基础，该文献提出了一种“伤害”的操作化定义，并认为这或许能够“为新一代个性化决策提供关键所在”（Mueller and Pearl，2022年）。他们的方法提出了若干策略，用存在混杂的非实验数据补充实验数据；在经典方法中，这类非实验数据往往被忽视（Tian and Pearl，2000）。我们将说明，他们的主张预先假定了某种特定的“伤害”操作化定义，并将其与由实验数据识别出的参数所界定的伤害概念加以对比。为了

阐明问题，我们提出两种替代方法，分别称为反事实方法和干预主义方法。这两种方法的区别在于，它们采用的效用函数在本质上截然不同。此外，两者分别对应不同的“伤害”概念，也对希波克拉底原则作出了不同的诠释。

2. 预备知识

假设我们从一个大型总体中观察到一组随机抽取的患者。我们将 Y 设为结局，将 A 设为二元治疗，将 L 设为基线协变量。

本文将始终使用潜在结局记法，表示在将治疗固定为某个特定值的干预下会出现的结局（Rubin，1974）（更多说明见网络附录1）。例如， $Y^{a=1}$ 表示患者在 A 被固定为 $a=1$ 时（即使这与事实相反）会出现的潜在结局。

我们将 $\{L, A, Y, Y^{a=1}, Y^{a=0}\}$ 视为完整数据，其中只有 $\{L, A, Y\}$ 被观察到。在讨论反事实方法时，我们用 S 表示联合主层变量 $\{Y^{a=1}, Y^{a=0}\}$（Robins，1986年；Frangakis 和 Rubin，2002年），其中事件（$S=1$、$S=2$、$S=3$、$S=4$）分别对应事件（$\{Y^{a=1}=1, Y^{a=0}=0\}$、$\{Y^{a=1}$ =0、$Y^{a=0}=1\}$、$\{Y^{a=1}=1, Y^{a=0}=1\}$、$\{Y^{a=1}=0, Y^{a=0}=0\}$）。

2.1. 定义效用函数

效用对于决策理论至关重要（冯·诺依曼和摩根斯坦，2007）。最简单的效用函数，就是患者在某种治疗下的结局a, Y^a。更复杂的效用函数还可能取决于实际接受的治疗。我们将考察一种这样的效用 U_1^a under a,

$$U_1^a = \mu(Y^a, a).$$

这种效用完全由在治疗 a下可直接观察到的变量定义。我们将这类效用称为干预主义效用。

然而，原则上，效用可以是任意一组定义明确的患者变量的函数，例如，与未分配给该患者的治疗相对应的潜在结局。与干预主义效用不同，这种效用部分由反事实变量定义，而这些变量在治疗a下无法直接观察。因此，我们将这类效用称为反事实效用。一种特殊情况是 U_2^a，它同时取决于两种治疗水平下可能产生的干预主义效用：

$$U_2^a = \gamma^*(\mu(Y^a, a), \mu(Y^{1-a}, 1-a), a) \tag{1}$$

$$= \gamma(S, a)。 \tag{2}$$

U_2^a 也可以理解为患者主分层 $S = \{Y^{a=1}, Y^{a=0}\}$ 及其分配治疗 a的函数。指定反事实效用的患者或研究者可以像式（1）那样指定 μ 和 γ^*，也可以像式（2）那样直接指定 γ。

我们将分别采用反事实效用函数和干预主义效用函数的方法称为反事实方法和干预主义方法。

3. 个性化医疗中伤害的双重定义

临床培训生常会举行背诵希波克拉底誓言“首先，不伤害”的仪式（Orr 等人，1997）。然而，广泛共识认为，该誓言并非足以充分阐明

现代医学伦理的丰富表述，其含义也并非不言自明（Beauchamp and Childress, 1994）。尽管如此，这并未阻止一些当代因果推断领域的作者声称存在一种唯一的数学表述：例如，Ben-Michael 等人（2022）将“首先，不伤害”原则等同于反事实效用函数的一种特定设定，γ，使得

$$\gamma(s=1,a=0)-\gamma(s=1,a=1)>\gamma(s=2,a=1)-\gamma(s=2,a=0). \tag{3}$$

当 Y 表示患者死亡时，$S=1$ 表示只有在接受治疗时才会死亡的患者（“受害者”），而 $S=2$ 表示只有在接受治疗时才能存活的患者（“获救者”）。因此，Ben-Michael 等人（2022）考虑了这样一种 γ：对于“受害者”患者，停止治疗所带来的效用增益 $\gamma(s=1,a=0)-\gamma(s=1,a=1)$，大于对“获救者”患者施予治疗所带来的效用增益 $\gamma(s=2,a=1)-\gamma(s=2,a=0)$。

这些似乎是对称的行动，因此应被赋予相同的收益：据推测，这两种决策都能拯救一条生命，因为它们选择了唯一能带来有利结果的决策。然而，根据 Ben-Michael 等人（2022年）采用的伤害译法，治疗 $a=1$被赋予了“作为”的能力。因此，第一种收益显然避免了错做错误，而第二种收益则避免了不作为错误。这种不对称性显然源于“首先，不要造成伤害”这一告诫，也就是说，“首先”不要犯错。Ben-Michael 等人（2022年，第8页）断言，这种特定类型的不对称性“由希波克拉底誓言中的‘不造成伤害’原则所编码”。他们的译法引出了以下反事实伤害定义：

定义1（反事实伤害）。假设某患者被分配接受治疗 $a = 1$ （相对于不接受治疗 $a = 0$），且其接受治疗时的干预主义效用低于不接受治疗时的效用，即 $\mu(Y^{a=1}, a = 1) < \mu(Y^{a=0}, a = 0)$。这就是反事实伤害。

换言之，当患者接受治疗 $a = 1$ ，但该治疗值并未使其个人干预主义效用最大化时，就会发生反事实伤害。之所以称其为反事实，是因为其定义至少依赖于患者的一个潜在结局，而原则上该结局不可能被观察到。我们将这一反事实伤害概念与 Ben-Michael 等人（2022）以及 Li 和 Pearl（2019）；Mueller and Pearl（2022）的提议联系起来，因为他们的提议同样要求考虑患者的主分层。

虽然定义1表面上是对希波克拉底意义上“伤害”的一种自然诠释，但对现代医学伦理的考察表明，实践中可能采用另一种表述。Beauchamp and Childress（1994）指出，不伤害原则与“首先，不伤害”原则被视为完全相同。然而，他们区分了两种情况：一种是负面结果几乎可以确定由某种行为引起，另一种则只是行为主体造成了伤害风险。例如，Beauchamp and Childress（1994，第154页）指出：“不伤害义务不仅包括不造成伤害的义务，也包括不施加伤害风险的义务。”在放弃确定性的情况下，恶行被界定为偏离“应有注意标准”。但在现代循证医学中，往往存在多种护理标准；也就是说，每当第二种治疗 $a = 1$ 被认定为相对于第一种治疗 $a = 0$非劣效时，便是如此。这类判定通常依据临床试验作出，在这些试验中， $\mathbb{E}[Y^{a=1}]$ 和 $\mathbb{E}[Y^{a=0}]$ 被确定。

那么，请看以下例子：假设 $U_1^a = (1 - Y^a)$ 定义了某种特定的干预性效用，因此 $\{Y^{a=1} > Y^{a=0}\} \equiv \{S = 1\}$ 构成反事实伤害。这

种情况是可能的： $P(S=1) > 0$ 和 $\mathbb{E}[Y^{a=1}] = \mathbb{E}[Y^{a=0}]$可以同时成立；按照定义1的意义，治疗会对一些患者造成反事实伤害，但医生在分配 $a = 1$时仍可能被认为遵循了“应有的医疗标准”。这立即带来一个重要含义：如果遵循“应有的医疗标准”等同于遵循不伤害原则，而不伤害原则又等同于“首先，不伤害”这一原则，那么我们就不能简单地将希波克拉底意义上的“伤害”等同于定义1中的反事实伤害。这一论证促使我们提出另一种定义；当结局不确定时（而这几乎总是如此），这种定义与现代生物伦理学中的不伤害原则更为一致。为了阐述这一伤害定义，我们首先要界定医疗服务提供者为遵循“应有的医疗标准”必须考虑的患者特征。

定义2（相关特征）。令 W 表示在治疗前具有明确定义取值的特征。令 X 表示 W 中在治疗前测量的子集，并假设某位患者的观测值为 $X = x$。令 X' 表示 X 中最大的子集，使得对于 $x' \subset x$,

$$sign\Big\{\mathbb{E}[Y^{a=1} \mid X' = x'] - \mathbb{E}[Y^{a=0} \mid X' = x']\Big\}$$

就决策任务而言，研究者将其视为固定且已知。因此，X' 是该患者的相关特征。

定义2指出了患者的哪些特征可用于实际比较两种治疗下的预期结局。就实际应用而言，考虑到领域知识、规模足够大的临床试验结果或其他实证证据（例如，基于现有数据对条件因果效应进行的一致估计），对于具有相关特征 x' 的患者，平均处理效应的正负号或许可以视为已确定且已知。现在，我们可以给出“伤害”的第二个定义，称之为

干预主义伤害。

定义3（干预主义伤害）。假设 L 表示患者的相关特征。假设患者被分配接受治疗 $a = 1$（相对于不接受治疗 $a = 0$），且在给定 $L = l$ 的情况下，接受治疗时的预期潜在效用低于不接受治疗时的预期效用，

$$\mathbb{E}[\mu(Y^{a=1}, a = 1) \mid L = l] < \mathbb{E}[\mu(Y^{a=0}, a = 0) \mid L = l].$$

这就是干预主义伤害。

因此，当患者接受治疗 $a = 1$，但根据其相关特征计算的期望干预主义效用并未达到最大值时，就会发生干预主义伤害。这是一种干预主义概念，因为确定其是否发生所需的参数可以通过实验观察。此外，它还以事后可观察到的特定患者的事实响应为依据来设定效用。

4. 决策理论

经典假设认为，在未来的决策场景中，理性主体会选择使其期望效用最大化的治疗 a（Von Neumann and Morgenstern，2007）$\mathbb{E}[U^a]$。患者会利用自己在决策时对自身特征的了解，以及对完整数据分布的了解；也就是说，他们会根据与其相关特征 P 相对应的参数来做出决策。例如，当患者观察到其基线相关特征 $L = l$的取值时，就会选择治疗 a，使其等于

$$\text{argmax}_a \mathbb{E}[U^a \mid L = l].$$

假设一位具有相关特征 $X' = x'$ 的患者接受伤害的反事实概念

（定义1）。因此，他们通过指定干预主义效用函数 μ 及其修正函数 γ^*，来指定反事实效用函数 γ。设 U_2^a 定义了这一效用。如果这位患者是理性主体，他们会选择使效用最大化的治疗 a

$$\mathbb{E}[U_2^a \mid X' = x'] = \sum_s \gamma(s, a) P(S = s \mid X' = x').$$

或者，假设这位患者更认同干预主义伤害概念（定义 3）。因此，他们只需指定一个干预主义效用函数 μ。令 U_1^a 表示在 μ下的干预主义效用。如果这位患者是理性主体，他们会选择使下式最大化的治疗 a

$$\mathbb{E}[U_1^a \mid X' = x'] = \sum_y \mu(y, a) P(Y^a = y \mid X' = x').$$

5. 实践考量

5.1. 利用实验数据实现效用最大化

考虑这样一项实践任务：利用观测数据的参数来选择治疗。对于效用最大化的干预主义患者而言，这一任务可简化为根据这些观测数据的参数，确定每个 a 对应的 $\mathbb{E}[Y^a \mid L = l]$。当数据来自理想实验时，我们通过以下方式确定 $\mathbb{E}[Y^a \mid L = l]$

$$\mathbb{E}[Y^a \mid L = l] = \mathbb{E}[Y \mid L = l, A = a]. \tag{4}$$

这一恒等式源于实验设计所确保的假设（明确论证见网络附录 2）。给定这些参数和指定的干预主义者

效用函数 μ，那么 L便构成每位患者的相关协变量。具有$L = l$ 的患者可以通过以下方式轻松选择使期望效用最大化的治疗 a

$$\text{argmax}_a \mathbb{E}[Y \mid L = l, A = a].$$

对于效用最大化的反事实患者而言，这项任务可归结为对每个 s识别 $P(S = s \mid L = l)$，但这比最大化干预主义效用更具挑战性；仅凭理想随机试验所要求的那些假设，通常无法实现识别。

或者，利用已识别量施加的约束，在这些假设下可以部分确定 $P(S = s \mid L = l)$ 的值。具体而言，这些约束可用于构建主分层概率的逻辑上限和下限，即界限，从而充分利用这些假设。当上限和下限重合时，概率即可识别；否则，在这种情形下，界限内概率的真实值最终取决于一个参数，而观测数据完全无法为该参数的取值提供信息。

我们回顾两种界限足以识别使 $\mathbb{E}[U_2^a \mid L = l]$ 最大化的治疗 a 的情形。第一种情形是上限和下限重合，因此 $\{Y^{a=1}, Y^{a=0}\}$ 的联合分布完全由实验数据确定；这只会在至少一个治疗水平 a下，结果能够被确定无疑地预测时发生。因此， S 的条件分布是点识别的，且可直接评估 $\mathbb{E}[U_2^a \mid L = l]$ 。这是一个可通过实验数据验证的条件，即当 $P(Y^a = y \mid L = l)$ 得到识别时。

其次， γ 的某些设定允许在仅有 $P(S =$

$s \mid L = l$)界限的情况下仍实现最大化。如果我们将 $\delta(s) = \gamma(s, a = 1) - \gamma(s, a = 0)$视为在主分层 s中的患者选择治疗 $a = 1$ 所获得的反事实效用增益，就可以将 $\mathbb{E}[U_2^{a=1} \mid L = l] - \mathbb{E}[U_2^{a=0} \mid L = l]$重新表示为

$$\sum_s \delta(s) P(S = s \mid L = l). \tag{5}$$

特别是，当 γ 满足

$$\delta(1) + \delta(2) = \delta(3) + \delta(4), \tag{6}$$

那么 $\mathbb{E}[U_2^{a=1} \mid L = l] - \mathbb{E}[U_2^{a=0} \mid L = l]$ 可重新表示为

$$\left\{ \Big(\delta(1) - \delta(4)\Big) P(Y^{a=1} = 1 \mid L = l) \right\} + \left\{ \Big(\delta(3) - \delta(1)\Big) P(Y^{a=0} = 1 \mid L = l) \right\} + \delta(4). \tag{7}$$

表达式 (7) 仅取决于研究者指定的 δ，以及诸如 $P(Y^a = y \mid L = l)$ 这样的项，这些项可通过实验识别。表达式 (6) 中对 γ 的约束被一些作者称为“增益相等”（Li and Pearl, 2019年、2022b）。Ben-Michael 等人（2022年）考察了一类特殊的 γ，称为“对称”效用函数，它们也满足“增益相等”。这类函数对“作为”错误和“遗漏”错误赋予相同权重，因此不符合他们对“首先，不伤害原则”的解释。

然而，一般而言，只想依靠可通过实验加以确保的假设，来选择使其反事实效用最大化的治疗方案 a 的患者，将

只能依据这样一项认知采取行动： $P(S = s \mid L = l)$ 必须落在某个区间内。基于 Cui（2021）的研究结果，Ben-Michael 等人（2022）指出，尽管一般情况下无法识别使 $\mathbb{E}[U_2^a \mid L = l]$ 最大化的治疗值 a^*，但可以识别使最坏情形遗憾最小化的治疗值 a'，其中遗憾是相对于 a^* 而言的（Savage，1951），并建议患者采用这种选择程序。Li 和 Pearl（2019）根据两个模拟示例的结果，以启发式方式提出了相同的程序。尽管遗憾最小化使人们能够在不确定性下做出决策，但我们指出，采用这一方法背离了经典决策理论。

5.2. 关于实验数据与非实验数据的融合

为便于说明，我们假设没有测量任何基线协变量，即 $L \equiv \emptyset$，但这些论证可以直接扩展到测量了基线协变量的情形。

5.2.1. 反事实方法

Tian 和 Pearl（2000）说明，利用存在混杂的非实验数据进行补充，可以缩小由实验数据识别出的 $P(S = s)$ 的界限。为帮助理解这一结果，我们考虑任意一种治疗方案 g。比较“始终治疗” $a = 1$ 与 g 的平均治疗效应（ATE），即

$$\mathbb{E}[Y^{a=1}] - \mathbb{E}[Y^g]$$

是 $P(S = 1)$ 的有效下界（见网络附录 2 中的命题 2）。例如，考虑 g 为“从不接受治疗”的特殊情形， $a = 0$。Greenland 和 Robins（1986年）

经典地说明，对于二元 $Y \in \{0,1\}$，

$$\mathbb{E}[Y^{a=1}] - \mathbb{E}[Y^{a=0}] = P(S=1) - P(S=2).$$

因此， $P(S=1)$ 不可能低于这一平均治疗效应，且它高于这一平均治疗效应的幅度恰好等于 $P(S=2)$。在实验数据所识别的所有平均治疗效应中，正是这一特例使下界最大（见网络附录 2 中的命题 3）。对于其他制度 g，它们可能取决于未测量特征；若非此类数据无法识别任何这类制度，本可得到更高的下界。因此，通常使用这一典型平均处理效应来计算精确非参数界限。然而，当还使用存在混杂的非实验数据时，其中一个此类制度下的期望结果确实可以识别：即用于生成该非实验数据的实际（但未知）制度，也就是 $\mathbb{E}[Y]$。随后可将这一参数作为另一次机会，用来提高 $P(S=1)$ 的下界，如 Tian 和 Pearl（2000）所示。我们在网络附录 3 中，依据本文的符号和常规假设，对这些改进后的界限作了阐述。

尽管这一结果早已为人所知，但 Li and Pearl（2022a）以及 Mueller and Pearl（2022）目前都强调，它是实施反事实方法的一项重要实用工具。

5.2.2. 干预主义方法

实践中对患者的治疗分配可分解为两个步骤：1）治疗意向；以及 2）治疗分配。在第一步中，患者或医疗服务提供者会决定期望接受的治疗。这一决定通常与健康状况有关

以及其他可能预测结局的信息，因此有治疗意向者与无治疗意向者并非可交换的群体。这就是混杂。我们引入一个二元“意向性治疗”（ITT）变量 A^*，表示期望接受的治疗。分配治疗 A 是实际接受的治疗。在非实验（存在混杂）数据中，意向治疗与分配治疗可能始终一致，因为患者和临床医生可以自由按照自己的意向行事。在实验数据中，通常不会假设 A^*与 A 一致，因为在实验中，患者的治疗意向会受到随机治疗分配机制的干预。在这两种情况下，观测数据通常都不包含对 A^*的测量。不过，我们可以通过辅助假设利用 A^*，从而为实验数据与非实验数据的融合提供依据；如果实验组和非实验组患者是从同一来源总体中随机招募的，这些假设可能成立。我们在网络附录 2 中对这类假设进行了定义。

重要的是，当在未来决策情境中测量 A^*，且已知或识别出给定 A^* 时平均治疗效应的符号，那么 A^* 将是定义2中所述的相关特征。事实上，只要恰当地融合实验数据与非实验数据，这些效应便可识别（见网络附录 2）。直观而言，难以识别的量是：在分配治疗水平a 下，患者自身的治疗意向 a' ——在实验中未测量——与其分配结果不一致的患者的期望结局，即 $a \neq a'$。但我们也知道，在分配治疗水平 a 下的边际预期结果——可在实验中识别——只是以下两者的加权平均：这一条件量，以及患者自身的治疗意向实际上与其分配结果一致的患者的相应期望结局 $a = a'$；权重为各治疗意向的概率。由于第二组量可直接从非实验数据中观察到，因此难以识别的量是唯一未知项，

其数值可以直接推导出来。

如 Stensrud 和 Sarvet（2022年）所示，当额外将意向治疗变量 A^* 用作相关协变量时，患者的期望干预主义效用会持平或更高。此外，网络附录 2 中的命题 4 表明，当且仅当非实验数据提高了“伤害”概率的下界，即 $P(S = 1 \mid L = l)$，非实验数据才会提高干预主义者的期望效用。因此，存在混杂的非实验数据的反事实用途与干预主义用途彼此相关。

6. 哲学考量

我们以一个简单的干预主义效用 $U_1^a = \mu(Y^a, a) = 1 - Y^a$ 为基础来展开哲学考量，并假设 $Y = 1$ 表示患者死亡。我们还考虑由 γ^* 和 μ所诱导的反事实效用 U_2^a。我们强调，这些考量并非面面俱到，而是受到当代倡导者关于反事实方法的主张所启发。与这些倡导者不同，我们在此并不倡导任何一种方法。

6.1. 反事实效用下的超额死亡

Ben-Michael 等人（2022）指出，在反事实方法下选择治疗值 a 的任务，可以改写为以下约束优化任务：

$$\text{argmin}_a \mathbb{E}[Y^a] \text{ subject to the constraints that}$$

$$P(S = 1) < \chi_1 \text{ and}$$

$$P(S = 2) < \chi_2,$$

其中， χ_1 和 χ_2 是完整数据某一特定分布下由 γ 所蕴含的 $[0,1]$ 中的一些数值。相比之下，干预主义者的任务只是求解优化任务 $\text{argmin}_a \mathbb{E}[Y^a]$，且不受任何约束。

这一表述有一个简单的推论：当政策制定者优化反事实效用时，通常会有更多人死亡。

反事实方法的支持者可能会这样辩称：但并非所有死亡都应一视同仁；他们可能会说，由 γ 给出的真实效用，是一种在主分层上采用可能不对称的反事实效用函数的效用，该函数经过恰当构建，能够反映反事实伤害的概念。然而，一个严重的问题是，我们没有直接证据证明这些主分层确实存在。即使有人在形而上学上承诺这些主分层存在，患者也永远无法知道自己真正属于哪个主分层，除非处于极端情形；因此，没有患者能够知道自己的事后效用，也无法依据直接观察来评估任何政策，或将其与替代政策进行比较。

换言之，反事实方法要求人们相信某些形而上学对象，而这些对象的存在既无法证实，也无法否定。患者或许可以自由持有这类信念，但政策制定者应当了解其推论：当反事实框架被用于制定社会政策和法规时，它会强迫人们遵循一种不可验证的形而上学，以及相应的、以这些术语进行推演的逻辑。相比之下，在此类情境中采用干预主义框架，则不会施加这种强迫，而且患者结局和群体结局都是可观察的，因而是透明的。

6.2. 美国法律中的医疗责任

反事实方法的支持者（例如 Tian 和 Pearl（2000））通过援引现代法律和医学领域的论述，为其应用提供了依据。事实上，在美国，现代医学法律体系在判定患者结局的责任归属时，似乎采用了反事实逻辑，

法律判例已在医疗过失认定中确立了“若非”原则。一种解释认为，除非在治疗当时能够判定，若非医疗服务提供者的行为，负面结局就不会发生，否则不能认定医疗服务提供者应对患者的负面结局承担责任。然而，尽管显然援引了这类反事实术语，同样的法律现象也可能通过干预主义视角来理解。如果责任认定要求判定 $P(S = 1 \mid L = l) = 1$，那么仅仅具有“若非因果关系”的高概率并不足够。有关判例法的综述，参见 Becker 等人（2018）。但是，$P(S = 1 \mid L = l) = 1$ 恰恰发生在 $P(Y^{a=0} = 0 \mid L = l) = 1$且 $P(Y^{a=1} = 1 \mid L = l) = 1$之时。因此，尽管使用了反事实语言，当前医学法律框架所产生的实际现象仍可等效地用干预主义逻辑来解释：只有当 $P(Y^{a=0} = 0 \mid L = l) = 1$ 且$P(Y^{a=1} = 1 \mid L = l) = 1$时，医疗服务提供者才会被追究责任。

6.3. 部分识别的问题

采用最小假设的研究者通常只能部分识别预期反事实效用（见第5节）。补充非实验数据并不能解决这一问题，因为这些数据通常只能使界限缩窄。因此，真正能够最大化反事实效用的治疗仍然无法确定。Ben-Michael 等人（2022）提出通过遗憾最小化方法（Savage，1951）来消除这种模糊性。然而，遗憾最小化只是解决此类模糊性的多种成熟方法之一。其他方法包括选择使最小可能效用最大化的规则，或对未识别的干扰参数剩余可能取值赋予一个分布（综述见 Stoye（2012）和 Cui（2021））。只要反事实效用无法点识别，这些不同方法通常会在样本量不断增加时得出不同的

决策，因此研究者在这些方法之间的选择，构成了另一项主观的研究者自由度，而这不会给干预主义者带来负担。

7. 结论

我们定义并比较了用于决策的反事实方法和干预主义方法。与一些作者的观点相反（Li and Pearl，2019年；Mueller and Pearl，2022年），反事实方法不应必然预示着个性化医疗领域的革命。明确效用函数及其定义域，以及建立决策框架，仍然具有很强的主观性。与反事实方法的支持者观点存在张力的是，我们考察了若干实际和哲学考量；这些考量似乎使其应用变得问题重重，并对其某些核心前提提出了挑战，例如认为它以某种方式自然地对应于现行的医学伦理和法律实践。也许从人口政策制定者的角度来看，最成问题的一点是：当结局为死亡且采用反事实方法时，与干预主义方法所确定的政策相比，通常会有更多人死于所确定的最优政策之下。对反事实方法的一项严厉批评称其“危险地误导”，并警告说，“如果将其应用于实践，必将造成真正的伤害”（Dawid and Senn）。我们的立场如下：随着因果推断日益融入个性化医疗的开发，研究人员、从业者和其他决策者清楚理解不同决策方法及其实际和哲学后果，至关重要。

致谢

我们感谢 Philip Dawid 和 Stephen Senn 富有洞见的多次讨论与评论。本研究得到了瑞士国家科学基金会的支持（资助编号200021_207436：资源有限时的因果推断）。

参考文献

Tom L Beauchamp 和 James F Childress。《生物医学伦理学原理》。Edicoes Loyola，1994年。

Charles L. Becker、Shanin Specter 和 Thomas R. Kline。《最高法院与医疗过失法》。载于《宾夕法尼亚州最高法院：联邦中的生活与法律，1684—2017》，第241–258页。宾夕法尼亚州立大学出版社，2018年。

Eli Ben-Michael、Kosuke Imai 和Zhichao Jiang。《具有非对称效用的策略学习》。arXiv预印本 arXiv:2206.10479v1，2022年。

Yifan Cui。《部分识别下的个体化决策：三种视角、两个最优性结果和一个悖论》。arXiv预印本 arXiv:2110.10961，2021年。

A. Philip Dawid。统计因果关系的决策理论基础。《因果推断杂志》，9：39–77，2021年。数字对象标识符:10.1515/jci-2020-0008。

A. Philip Dawid 和 Monica Musio。《群体层面数据能告诉我们什么关于个体因果关系？》载于 A. Carriquiry、J. Tanur 和 W. Eddy 编，《公共利益中的统计学：纪念 Stephen E. Fienberg》，第 235–256 页。施普林格国际出版，

2022年。

DOI: 10.1007/978-3-030-75460-0_13。

Philip A Dawid 和 Stephen Senn。《无反事实的个性化决策》。《因果推断杂志》。

Constantine E Frangakis 和 Donald B Rubin。《因果推断中的主分层》。《生物统计学》，58(1)：21–29，2002年。

Sara G. Geneletti 和 A. Philip Dawid。《定义和识别已治疗者的治疗效应》。载于 Phyllis M. Illari、Federica Russo 和 Jon Williamson 编，《科学中的因果关系》，第 728–749 页。牛津大学出版社，2011年。

Sander Greenland 和 James M Robins。《可识别性、可交换性与流行病学混杂》。《国际流行病学杂志》，15(3)：413–419，1986年。

Michael R Kosorok、Eric B Laber、Dylan S Small 和 Donglin Zeng。《精准医学与个体化策略发现理论与方法专刊导言》，2021年。

Ang Li 和 Judea Pearl。基于反事实逻辑的个体选择。载于《第二十八届国际人工智能联合会议论文集》，IJCAI-19，第 1793–1799 页。国际人工智能联合会议组织，2019年7月。数字对象标识符:10.24963/ijcai.2019/248。

Ang Li 和 Judea Pearl。因果概率：观察性数据的作用。arXiv 预印本 arXiv:2210.08874，2022a。

Ang Li 和 Judea Pearl。个体选择：案例研究及与 A/B 测试启发式方法的比较。arXiv 预印本 arXiv:2210.05030，2022b。

Scott Mueller 和 Judea Pearl。《个性化决策：概念导论》。arXiv 预印本 arXiv:2208.09558，2022年。

Susan A Murphy。《最优动态治疗方案》。《英国皇家统计学会杂志：B辑（统计方法学）》，65(2):331–355，2003年。

Robert D Orr、Norman Pang、Edmund D Pellegrino 和 Mark Siegler。《希波克拉底誓言的使用：对20世纪实践的回顾，以及对1993年美国和加拿大医学院所采用誓言的内容分析》。《临床伦理学杂志》，8(4):377–388，1997年。

Thomas S Richardson 和 James M Robins。《单世界干预图（SWIG）：统一反事实方法与因果关系图形方法》。《统计与社会科学中心》华盛顿大学系列。工作论文，128 (30):2013，2013年。

James M Robins。具有持续暴露期的死亡率研究中因果推断的新方法——应用于控制健康工人生存者效应。数学建模，7(9-12)：1393–1512，1986年。

James M Robins。用于最优序贯决策的最优结构嵌套模型。载于《第二届西雅图生物统计学研讨会论文集》，第189–326页。施普林格，2004年。

James M. Robins、Tyler J. Vanderweele 和 Thomas S. Richardson。《对 Antonio Forcina〈存在不依从时的因果效应：一种潜变量解释〉一文的评论》

Metron，LXIV：288–298，2007 年。

Donald B. Rubin。《随机化与非随机化研究中治疗因果效应的估计》。《教育心理学杂志》，66：688–701，1974。

Leonard J Savage。《统计决策理论》。《美国统计学会期刊》，46(253)：55–67，1951。

Mats J Stensrud 和 Aaron L Sarvet。《算法辅助的人类决策的最优方案》。即将出版，2022年。

J¨org Stoye。《模糊性下统计决策的新视角》。《经济学年度评论》，4 (1)：257–282，2012。

Jin Tian 和 Judea Pearl。《因果概率：界限与识别》。《数学与人工智能年刊》，28：287–313，2000。

Anastasios A Tsiatis、Marie Davidian、Shannon T Holloway 和 Eric B Laber。《动态治疗方案：精准医学的统计方法》。Chapman and Hall/CRC，2019年。

John Von Neumann 和 Oskar Morgenstern。《博弈论与经济行为》。载于《博弈论与经济行为》。普林斯顿大学出版社，2007 年。

附录 A：补充命题与证明

在本附录中，我们将为正文讨论的结果提供形式化论证，包括实验数据与非实验数据的融合，以及使用当前研究数据解决未来的决策问题。

附录 A.1：预备知识

为了区分实验数据和非实验数据，我们引入一个额外变量 R，它是对 A 中 L各个水平进行对照实验时参与情况的二元指标。我们用 $O_F \equiv \{R, L, A, Y, Y^{a=1}, Y^{a=0}\}$ 表示完整数据，用$O \equiv \{R, L, A, Y\}$ 表示观测数据。

此外，我们将从两个维度区分概率测度：1）该测度描述的是完整反事实数据的概率分布，还是观测数据的分布；2）该测度描述的是当前研究中的概率分布（用于数据分析），还是未来情境中的概率分布（用于决策）。第一个区别使我们能够区分可直接观测的分布与部分观测的因果分布；第二个区别则使我们能够考虑当前研究场景与未来决策场景之间可能存在的差异。

因此，我们令 P^F 表示完整数据的任意联合分布，并令 $P \equiv P(P^F)$ 表示观测数据中 P^F 的子分布。此外，我们令 P_0^F 和 P_0 表示研究进行时真实的此类分布，并进一步令 P_1^F 和 P_1 表示未来决策时相应的此类分布。注意，定义2和定义3在逻辑上都涉及在 P_1^F（未来决策时完整数据的联合分布）下计算的期望。同样，效用最大化中所使用的期望，

在第4节（决策理论）中讨论的期望，也同样是在 P_1^F 下计算的。直观地说， P_1^F 是未来决策者将受其支配的数据分布，因此也是与该个体相关的测度。

令 g 为 A 的任意制度（处理分配规则），并考虑取值为 $\{0,1\}$ 的二元结果 Y。假设事实数据由一组非参数结构方程生成，且所有反事实均通过递归替代生成（例如 Richardson and Robins（2013）中的做法）。

令 A^{g+} 表示制度 g 下分配的治疗值，其中 A 是在生成事实数据的制度下所接受的治疗。令 $\tau^g_{P^F}$ 表示在分布 P^F 下比较 $a=1$ 与 g 的平均治疗效应（ATE）：

$$\mathbb{E}_{P^F}[Y^{a=1}] - \mathbb{E}_{P^F}[Y^g].$$

令 $\tau^0_{P^F}$ d 表示特殊的 ATE，其中 g 是无人接受治疗的规则， $a=0$,

$$\mathbb{E}_{P^F}[Y^{a=1}] - \mathbb{E}_{P^F}[Y^{a=0}]$$

并令 τ_{P^F} 表示特殊情形： g 是事实数据中实施的规则

$$\mathbb{E}_{P^F}[Y^{a=1}] - \mathbb{E}_{P^F}[Y].$$

同样，令 $\beta^g_{P^F}$ 表示比较 g 与 $a=0$ 的平均治疗效应（ATE），并令 β_{P^F} 表示特殊情形： g 是事实数据中实施的规则。

令 $\mathcal{L}_{P^F}$为仅知道参数 $\mathbb{E}_{P^F}[Y^{a=1}]$和 $\mathbb{E}_{P^F}[Y^{a=1}]$时， $P^F(S=1)$ 的精确下界。令 $\mathcal{L}^*_{P^F}$ 为参数 $P(Y=$

$y, A = a$) 对所有 a, y 也已知时的精确下界。

附录 A.2. 在模型 $\mathcal{M}$下识别条件期望潜在结局我们基于以下假设定义模型 $\mathcal{M}$ 。

A1. 对所有 r、l、a，$\mathbb{E}_{P_0^F}[Y^a \mid A = a, L = l, R = r] = \mathbb{E}_{P_0^F}[Y \mid A = a, L = l, R = r]$。

A2. 对所有 a，$Y^a \perp\!\!\!\perp_{P_0^F} A \mid L$、
$R = 1$。A3. ($Y^{a=1}$, $Y^{a=0}$)
$\perp\!\!\!\perp_{P_0^F} R \mid L$。A4. $P_0^F = P_1^F$。

假设 A1 指出，接受治疗 a、协变量为 l 且试验参与指标为 r 的患者，其期望结果等于试验中接受治疗方案固定为 a的类似患者的期望结果，这也称为“分布一致性”（Dawid，2021年）。假设 A2 指出，在试验中，患者在 L 各层内的不同治疗水平之间可交换（$R = 1$）。假设 A3 指出，在 L各层内，试验中的患者与试验外的患者可交换。假设 A4 指出，完整数据的真实概率分布在数据收集时（P_0^F）与未来决策情境（P_1^F）之间保持稳定。

我们认为， $\mathcal{M}$ 是最简的，因为它仅包含用于以下目的的基本模型模式：1）关联观测参数与因果参数（A1）；2）区分实验数据与非实验数据（A2）；3）论证实验数据与非实验数据融合的合理性（A3）；以及 4）论证在未来决策情境中使用观测数据的合理性（A4）。

我们可以按如下方式，在 $\mathcal{M}$ 条件下从 P_0 的参数中获得 $\mathbb{E}_{P_1^F}[Y^a \mid L = l]$ ：

$$
\begin{aligned}
\mathbb{E}_{P_1^F}[Y^a \mid L=l] =& \mathbb{E}_{P_0^F}[Y^a \mid L=l] \\
=& \mathbb{E}_{P_0^F}[Y^a \mid L=l, R=1] \\
=& \mathbb{E}_{P_0^F}[Y^a \mid L=l, A=a, R=1] \\
=& \mathbb{E}_{P_0}[Y \mid L=l, A=a, R=1],
\end{aligned} \tag{A.1}
$$

假设所有项都有明确定义。第一个等式基于数据收集时点 P_0^F的完整分布与未来决策时点 P_1^F的完整分布等价（A4.）；第二个等式基于给定L 时试验参与者与非试验参与者之间的可交换性（A3.）；第三个等式基于给定 L 时试验中处理组与未处理组参与者之间的可交换性（A2.）；第四个等式则基于一致性（A1.）。

附录 A.3 关于干预主义方法将治疗意向用作相关协变量

这里我们定义了一个扩展模型 $\mathcal{M}^*$，其假设允许融合实验数据与非实验数据，以便在干预主义方法下改善决策。我们令 $\mathcal{M}^*$ 表示由 $\mathcal{M}$ 定义并补充了假设 B1. 至 B3. 的模型。我们令 P^{F*} 表示完整数据的任意分布，其中包括意向治疗变量 A^*。

B1. $P_1^*(A = A^* \mid R = 0) = 1$。

B2. $(Y^{a=1}, Y^{a=0}, A^*) \perp\!\!\!\perp_{P_0^F} R \mid L$。
B3. $P_0^{F*} = P_1^{F*}$。

假设 B1. 断言，在非实验环境（$R = 0$）中，患者的意向治疗变量（A^*）几乎必然与实际接受的治疗（A）一致，并且

因此，意向治疗变量是间接测量的。假设 B2. 与 A3. 的不同之处在于，它还断言，在给定协变量的条件下，试验参与独立于患者接受治疗的意向 A^*。假设 B3. 与 A4. 的不同之处在于，它断言完整数据的扩展定律（其中包括 A^*）在数据收集时（P_0^{F*}）与未来决策情境（P_1^{F*}）之间也保持稳定。

以下命题断言，在 $\mathcal{M}^*$ 下，这些参数确实是可识别的。

命题 1. 在 $\mathcal{M}^*$ 下，对于所有 $(a, a^* = 0, 1)$，$P_1^{F*}(Y^a = y \mid A^* = a^*, L = l)$ 都是可识别的。具体而言，

$$P_1^{F*}(Y^a = y \mid A^* = a^*, L = l) = \begin{cases} P(Y = y \mid A = a, L = l, R = 1) & for\ a = a^* \\ \frac{P(Y=y|A=a,L=l,R=1) - P(Y=y,A=a|L=l,R=0)}{P(A=a^*|L=l,R=0)} & for\ a \neq a^*. \end{cases}$$

相关命题见 Stensrud 和 Sarvet（2022年）；Robins 等人（2007）；Geneletti 和 Dawid（2011年）；Dawid 和 Musio（2022年）。

证明. 当 $a = a^*$ 时，论证过程与 (A.1) 中完全相同。当 $a \neq a^*$ 时，根据全概率公式，我们有 $P_1^{F*}(Y^a = y \mid A^* = a^*, L = l)$ 等于

$$\frac{P_1^{F*}(Y^a = y \mid L = l) - P_1^{F*}(Y^a = y, A^* = a \mid L = l)}{P_1^{F*}(A^* = a^* \mid L = l)}.$$

然后，分子中的左侧项仅根据 $\mathcal{M}$ 中的假设由实验数据识别，其余项则根据 $\mathcal{M}^*$特有的假设由非实验数据识别。例如，我们有：

$$
\begin{aligned}
P_1^{F*}(Y^a = y, A^* = a \mid L = l) =& P_0^{F*}(Y^a = y, A^* = a \mid L = l), \\
=& P_0^{F*}(Y^a = y, A^* = a \mid L = l, R = 0), \\
=& P_0^{F*}(Y^a = y, A = a \mid L = l, R = 0), \\
=& P_0(Y = y, A = a \mid L = l, R = 0).
\end{aligned}
$$

在上式中，第一个等号由 B3. 得出，第二个等号由 B2. 得出，第三个等号由 B1. 得出，第四个等号则由概率法则和 A3. 得出。 □

上述证明依赖于 A （因此也依赖于 A^*）为二元变量，但 Y 不必是二元变量。

附录 A.4. 命题 2 和 3

命题 2. 比较 $a = 1$ 和 g 的平均治疗效应是伤害概率的一个有效下界。也就是说，对于所有定律 P^F:

$$
P^F(S = 1) \geq \tau^g_{P^F}.
$$

证明。

$$
\begin{aligned}
P^F(Y^g = 1) =& P^F(Y^g = 1, A^{g+} = 1) + P(Y^g = 1, A^{g+} = 0) \\
=& P^F(Y^{a=1} = 1, A^{g+} = 1) + P^F(Y^{a=0}, A^{g+} = 0) \\
=& \Big\{ P^F(S = 1, A^{g+} = 1) + P^F(S = 2, A^{g+} = 0) \Big\} + P^F(S = 3).
\end{aligned}
$$

第一个等式依据概率法则得出，第二个等式依据递归代入得出

（一致性），第三个等式依据概率法则得出。令 ϵ 表示项 $\{P^F(S=1, A^{g+}=1) + P^F(S=2, A^{g+}=0)\}$。那么

$$\begin{aligned}\tau^g_{P^F} =& P^F(S=1) - \{P^F(S=1, A^{g+}=1) + P^F(S=2, A^{g+}=0)\} \\ =& P^F(S=1) - \epsilon.\end{aligned}$$

第一个等式根据定义（并约去 $P^F(S=3)$）得到，最后一个等式也根据定义得到。由于对于所有规律 P^F，ϵ必须为正，因此 $\tau^g_{P^F}$是 $P^F(S=1)$ 的有效下界。

□

命题 3. 当相对于制度 g，在 P^F 下满足强可交换性时，

$$\{Y^{a=1}, Y^{a=0}\} \perp\!\!\!\perp_{P^F} A^{g+}$$

则 $\tau^0_{P^F} \geq \tau^g_{P^F}$。

证明。

$$\begin{aligned}\tau^g_{P^F} =& P^F(S=1) - \{P^F(S=1, A^{g+}=1) + P^F(S=2, A^{g+}=0)\} \\ =& P^F(S=1) - \{P^F(S=1)P^F(A^{g+}=1) + P^F(S=2)P^F(A^{g+}=0)\} \\ =& P^F(A^{g+}=0)\{P^F(S=1) - P^F(S=2)\} \\ =& \tau^0_{P^F} P^F(A^{g+}=0).\end{aligned}$$

第一个等式由命题 2 得出，第二个等式由强可交换性得出，最后两个等式则由概率法则得出。由于对于

所有法则 P^F， $P^F()$ 都被限制在单位区间内，因此命题 3 的结论立即成立。 □

论证 当所有项都进一步以事件 $L = l$ 为条件时，两者相同。

附录A.5. 命题 4

Tian 和 Pearl（2000）表明，加入非实验数据可以改善 $P^F(S = 1)$ 的下界，即

$$\mathcal{L}^*_{P^F} > \mathcal{L}_{P^F}$$

当满足以下条件时：

$$\max \left\{ \begin{array}{c} \tau_{P^F}, \\ \beta_{P^F} \end{array} \right\} > \max \left\{ \begin{array}{c} 0, \\ \tau^0_{P^F} \end{array} \right\}. \tag{A.2}$$

令 $\tau_{P^F,\ A=1}$ 表示处理组中的平均治疗效应，

$$\tau_{P^F,A=1} \coloneqq \mathbb{E}_{P^F}(Y^{a=1} \mid A = 1) - \mathbb{E}_{P^F}(Y^{a=0} \mid A = 1),$$

并令 $\tau_{P^F,A=0}$ 表示未处理组中的平均治疗效应， $A = 0$。

命题 4. 当且仅当以下条件成立时，条件（A.2）成立：

$$sign\Big\{\tau_{P^F,A=1}\Big\} \not\equiv sign\Big\{\tau_{P^F,A=0}\Big\}. \tag{A.3}$$

证明。注意，如果满足 $\tau_{P^F} > \{0, \tau^0_{P^F}\}$ 或 $\beta_{P^F} > \{0, \tau^0_{P^F}\}$，则 (A.3) 成立。考虑以下推论：

$$
\begin{aligned}
&\tau_{P^F} > \tau^0_{P^F} \\
\Longleftrightarrow\ & P^F(S=1) - \{P^F(S=1,A=1) + P^F(S=2,A=0)\} > P^F(S=1) - P^F(S=2) \\
\Longleftrightarrow\ & P^F(S=2) > P^F(S=1,A=1) + P^F(S=2,A=0) \\
\Longleftrightarrow\ & P^F(S=2,A=1) > P^F(S=1,A=1) \\
\Longleftrightarrow\ & P^F(S=2 \mid A=1) > P^F(S=1 \mid A=1) \\
\Longleftrightarrow\ & \tau_{P^F,A=1} < 0.
\end{aligned}
$$

另假设 $\tau_{P^F} > 0$。

$$
\begin{aligned}
&\tau_{P^F} > 0 \\
\Longleftrightarrow\ & P^F(S=1) > P^F(S=1,A=1) + P^F(S=2,A=0) \\
\Longleftrightarrow\ & P^F(S=1,A=0) > P^F(S=2,A=0) \\
\Longleftrightarrow\ & P^F(S=1 \mid A=0) > P^F(S=2 \mid A=0) \\
\Longleftrightarrow\ & \tau_{P^F,A=0} > 0.
\end{aligned}
$$

因此，当且仅当 $\tau_{P^F,\ A=1} < 0$ 且 $\tau_{P^F,\ A=0} > 0$时，$\tau_{P^F} > \{0, \tau^0_{P^F}\}$成立。通过同构论证，我们可以证明，当且仅当 $\tau_{P^F,\ A=1} > 0$ 且 $\tau_{P^F,\ A=0} < 0$时，$\beta_{P^F} > \{0, \tau^0_{P^F}\}$成立。由此立即可得命题 4 的结论。

□

当所有术语都进一步以事件 $L = l$ 为条件时，论证完全相同。因此，

可以得出，在 $\mathcal{M}^*$ 下， $A = A^*$ 以概率 1 发生且 (A.2) 成立时，仅以 L 为函数的最优制度的期望结局，将劣于同时以 A^* 为函数的最优制度的期望结局。因此，正文中的论证仍然成立：在 $\mathcal{M}^*$ 下，当且仅当干预主义者通过使用这类数据能够获得更高的期望效用时，非实验数据才会提高反事实伤害概率的下界。

附录B. 使用实验数据和非实验数据回顾主分层界限

Tian 和 Pearl（2000）指出，对于任意条件事件 $X = x$，这类约束

$$P_1^F(S = 1, A = 1 \mid X = x) + P_1^F(S = 3, A = 1 \mid X = x) = P_1^F(Y = 1, A = 1 \mid X = x)$$

这将对主分层概率施加约束，例如， $P_1^F(S = 1 \mid X = x)$ 必须大于 $P_1(Y = 1 \mid X = x) - P_1^F(Y^{a=0} = 1 \mid X = x)$。利用模型 $\mathcal{M}$ 下的实验数据，我们同样可得，$P_1^F(S = 1 \mid L = l) = P_1^F(S = 1 \mid L = l, R = 1)$ 必须大于 $P_1(Y = 1 \mid L = l, R = 1) - P_1^F(Y^{a=0} = 1 \mid L = l)$。我们还可得， $P_1^F(S = 1 \mid L = l) = P_1^F(S = 1 \mid L = l, R = 0)$ 必须大于$P_1(Y = 1 \mid L = l, R = 0) - P_1^F(Y^{a=0} = 1 \mid L = l)$。Tian 和 Pearl（2000）的结果启发了以下关于 $P_1^F(S = s \mid L = l)$ 的下界：

$$P_1^F(S=1 \mid L=l) \geq \max \left\{ \begin{array}{c} 0 \\ P_1^F(Y^{a=1}=1 \mid L=l) - P_1^F(Y^{a=0}=1 \mid L=l), \\ P_1(Y=1 \mid L=l, R=1) - P_1^F(Y^{a=0}=1 \mid L=l), \\ P_1^F(Y^{a=1}=1 \mid L=l) - P_1(Y=1 \mid L=l, R=1), \\ P_1(Y=1 \mid L=l, R=0) - P_1^F(Y^{a=0}=1 \mid L=l), \\ P_1^F(Y^{a=1}=1 \mid L=l) - P_1(Y=1 \mid L=l, R=0). \end{array} \right\} \quad \text{(B.1)}$$

利用实验数据的识别结果，我们将这些约束完全表示为观测数据：

$$P_1^F(S=1 \mid L=l) \geq \max \left\{ \begin{array}{c} 0, \\ P_0(Y=1 \mid A=1, L=l, R=1) - P_0(Y=1 \mid A=0, L=l, R=1), \\ P_0(Y=1 \mid L=l, R=1) - P_0(Y=1 \mid A=0, L=l, R=1), \\ P_0(Y=1 \mid A=1, L=l, R=1) - P_0(Y=1 \mid L=l, R=1), \\ P_0(Y=1 \mid L=l, R=0) - P_0(Y=1 \mid A=0, L=l, R=1), \\ P_0(Y=1 \mid A=1, L=l, R=1) - P_0(Y=1 \mid L=l, R=0). \end{array} \right\} \quad \text{(B.2)}$$

然而，由于存在这样的约束： $P_0(Y=1 \mid L=l, R=1)$是 $P_0(Y=1 \mid A=0, L=l, R=1)$和 $P_0(Y=1 \mid A=1, L=l, R=1)$的凸组合，因此集合中的第三项和第四项必定小于或等于第二项，因而实际上无法进一步收紧界限。然而，当对于某个 a，$Y^a \not\perp\!\!\!\perp_{P_0^F} A \mid L, R=0$时（这在非实验数据中是预期情况），则对 $P_0(Y=$

$1 \mid L = l, R = 0$)不施加此类约束，而这些额外约束（第五项和第六项）确实可能收紧界限。当然，只有在研究者还利用（存在混杂的）非实验数据时，才能利用这些约束。我们还要强调，这些界限只在类似 $\mathcal{M}$这样的模型下成立；在这类模型中，实验数据和非实验数据之间的条件平均处理效应保持不变，但治疗 A 与潜在结局的条件联合分布有所不同。

附录 C. 关于潜在结局记法的说明

尽管我们使用的记法往往与关于反事实变量存在的假设相对应，但我们强烈强调，对于干预主义方法而言，这类断言并非必要。我们之所以首先采用这种记法，是因为反事实方法与反事实的存在密不可分。我们在全文中使用这些记法来减轻符号负担，但鼓励读者将任何由单一干预定义的参数，理解为假想对照实验中与该干预对应的实验组内可观测到的类似参数。例如，在治疗 a Y^a的条件下，协变量为$L = l$ 的患者的期望潜在结局，应理解为在一项假想对照实验中， A 固定为 a时， $L = l$ 患者的期望结局。再举一例，关于单一干预反事实与治疗之间（条件）独立性的主张，例如 $Y^{a=1} \perp\!\!\!\perp A$，应仅理解为：观测数据中接受不同治疗的患者，就结局Y而言是（条件）可交换的。