arXiv:2301.02904v1 [stat.ME] 7 Jan 2023

多研究、多结局情境下可迁移性的敏感性分析

Ngoc Q. Duong，Amy J. Pitts，Soohyun Kim，Caleb H. Miles

哥伦比亚大学梅尔曼公共卫生学院生物统计学系

摘要

现有的数据融合研究探讨了如何整合异质来源的数据以识别因果估计目标。这些识别结果通常需要额外的假设，才能进行有效的估计和推断。然而，对于多结局情境下因果效应的迁移和推广，相关文献仍然很少。在这种情境中，主要结局在研究层面系统性缺失，但其他结局变量可能充当代理变量。我们回顾一项正在进行的研究中提出的识别结果，该结果利用这些代理变量的信息来获得更有效的估计量，并介绍与之对应的关键识别假设。随后，我们提出评估这一方法对该识别假设敏感性的方法。

关键词：因果推断、数据融合、外部效度、可推广性、缺失数据、代理变量

1 引言

临床医学和公共卫生研究通常关注如何估计某种处理在特定目标人群中的效应。然而，即使在被视为金标准研究设计的随机临床试验中，确保外部效度仍然是一项挑战。这可能由多种原因造成，包括非随机抽样、过于严格的排除标准，或对感兴趣的目标人群界定不明确（Tan 等，2022年；Kennedy-Martin 等，2015）。基于汇总统计量的荟萃分析是整合并推广已发表研究的研究层面汇总统计量所得发现的常用工具，但往往依赖强假设，而这些假设通常并不合理。另一种方法则是在条件允许时，整合多项研究的个体层面数据，从而更好地控制各研究间的细微差异和异质性。每项研究单独来看可能都缺乏足够的信息来回答某个科学问题，但汇总起来便有能力回答该问题。关于可推广性和可迁移性方法的研究日益增多，这些方法有助于解决整合多项研究的个体层面数据所得效应估计值的外部效度问题。

可推广性关注的是研究人群属于目标人群子集的情形，而可迁移性则针对研究人群部分或完全处于目标人群之外的情形（Degtiar and Rose，2023年）。具体而言，只要研究人群与目标人群中的协变量具有共同支持，通常就可以将基于研究得出的因果效应估计值推广至目标人群（Gechter，2015；Tipton，2014）。另一方面，可迁移性是指将某项研究中估计的效应外推至一个由样本代表的人群，而该研究样本中并未测量感兴趣的某项主要结局。

现有方法会将某些估计得到的因果效应（例如平均处理效应（ATE））直接从观察到结局的研究迁移到结局缺失的其他研究中，或迁移到研究设计和情境存在异质性的研究中（Barein-

boim 和 Pearl，2016；Dong 等人，2020；Pearl 和 Bareinboim，2014；H¨unermund 和 Bareinboim，2019），或迁移到更广泛的目标人群（Dahabreh 等人，2020a,b；Lesko 等人，2017；Westreich 等人，2017）。在考虑多项研究时，通常会在随访中观察到不同的结局。然而，现有方法并未利用这些在随访中测量的、可能相关且具有信息价值的其他结局变量；利用它们有望大幅提升效率。现有的结局代理盲法通常依赖于条件潜在结局均值同质这一假设，以便有效地将估计结果从一个人群迁移到另一个人群。已有研究提出敏感性分析策略，用于考察违反这些假设在多大程度上会影响估计结果和所得推断（Nguyen 等人，2017；Dahabreh 和 Hern´an，2019；Dahabreh 等人，2022年）。

在进行中的研究中，我们开发了一种新策略，以便在多结局研究中主要关注结局在研究层面能否获得的情况不一致时，更高效地利用整合数据估计平均处理效应。所提出的方法学利用了主要结局潜在相关因素的随访测量，从而更精确地估计因果效应。本文在利用这些结局代理变量的同时，考察了可迁移性所依赖的关键共同结局回归假设（或针对研究选择的条件可交换性）；这一假设与传统上用于可迁移性的共同结局回归假设略有不同。我们讨论了该假设不成立时产生的偏倚，并开发了用于评估违反该假设情况的敏感性分析方法学。

本文其余部分安排如下。在第2节中，我们讨论多研究、多结局情境下平均处理效应的识别。在第3节中，我们讨论违反关键条件可交换性假设所造成的偏倚。在第4节中，我们将本研究情境中的条件可交换性假设与未利用结局代理变量的情境中所采用的假设进行比较。在第5节中，我们针对假设不成立的情况提出敏感性分析方法。我们将展示其经验

表现，并在第6节通过模拟研究展示所提出方法的性能，最后在第7节进行讨论。

2 主要结局系统性缺失时的研究数据整合

2.1 研究与数据设定

在此设定中，我们令 A 表示治疗指示变量， W 表示各项研究中共同观测到的一组协变量， Y 表示主要结局变量，集合 $\{T_1, \ldots, T_k\}$ 表示任一研究中在随访时测量的所有潜在结局代理变量，而 J_s 表示在研究 s 中测量的 $\{T_1, \ldots, T_k\}$ 的研究特定子集。假设共有 $\mathcal{S}$ 项研究，并按如下方式排序：对于前 s^* 项研究中的每个 s ，我们观测到变量集合 (Y, A, J_s, W)；对于其余 $\mathcal{S} - s^*$ 项研究中的每个 s，则仅观测到子集 (A, J_s, W)。换言之，在后一组研究中， Y 系统性缺失。不同于其他有关效应可迁移性的研究中仅涉及 (Y, A, W) 的标准设定，我们引入了 $\mathcal{T}_s$ 的使用，其中，对于每项研究 s，$\mathcal{T}_s \subset J_s$ 是用户指定的 J_s 子集。 $\mathcal{T}_s$ 可根据可用性和领域知识来选择，且必须确保它们至少在研究 $\{1, 2, \ldots, s^*\}$ 中的某一项里被观测到。

研究可以是随机试验，也可以是观察性研究；不过，本文不考虑部分研究为随机试验、而其他研究为观察性研究的情形。因此，针对特定研究的平均治疗效应和条件平均治疗效应可写为：

$$ATE(s) = E(Y_1 - Y_0 \mid S = s)$$

$$CATE(w, s) = E(Y_1 - Y_0 \mid W = w, S = s).$$

相应地，我们可以将总体平均治疗效应和条件平均治疗效应定义为:

$$ATE = \sum_{s=1}^{\mathcal{S}} ATE(s)$$

$$CATE(w) = E(Y_1 - Y_0 \mid W = w)$$

其中，权重可由用户指定，使得 $\sum_s \pi_s = 1$。例如，可以选择$\pi_s = P(S = s)$，或选择每项研究中个体所占的边际概率。或者，对于用户指定的已知分布 $Q_{W,S}$，我们可以定义 $ATE = E_{Q_{W,S}} CATE(W, S)$，该分布涉及 W和S。

由于在 $s \in \{s^* + 1, \ldots, \mathcal{S}\}$中未测量 Y，我们无法仅使用这些研究的数据直接估计平均处理效应和条件平均处理效应。我们的目的是将前 s^* 项研究中（其中观测到了 Y）的平均处理效应迁移到其余 $\mathcal{S} - s^*$ 项研究，同时利用结果代理变量集 $\mathcal{T}_s$ 中的信息来提高效率。为便于记号表达，令 σ_s 为前 s^* 项研究中同时观测到 $\mathbf{Y}$ 和 $\mathcal{T}_s$ 的研究子集。这样，我们就可以利用构成 σ_s 的研究中的信息，估计结果回归，从而将因果效应迁移到研究 s。在这一设定下，我们在一项正在进行、尚未发表的工作中证明，平均处理效应可以非参数识别为:

$$\begin{aligned}\Psi^{ATE} = &\sum_{s=1}^{s^*} \pi_s E\{E(Y \mid W, A = 1, S = s) - E(Y \mid W, A = 0, S = s) \mid S = s\} \\ &+ \sum_{s=s^*+1}^{\mathcal{S}} \pi_s E[E\{E(Y \mid \mathcal{T}_s, W, A = 1, S \in \sigma_s) \mid W, A = 1, S = s\} \\ &\qquad - E\{E(Y \mid \mathcal{T}_s, W, A = 0, S \in \sigma_s) \mid W, A = 0, S = s\} \mid S = s].\end{aligned} \tag{1}$$

第一项求和中的各项只是当观测到 Y 时，（特定研究中的）平均处理效应的标准识别公式。第二项求和可以识别，因为它仅取决于 σ_s 中各项研究里 Y 的分布，也就是说，在这些研究中， Y 确实

被观测到。

这里，我们对传统的可迁移性方法进行了改进，将随访时测量的一组与主要关注结局相关的结局信息纳入其中。

2.2 平均处理效应识别的假设

在所有结局均被观测到时，只要满足平均处理效应识别的标准假设，该推导所得的平均处理效应即可通过非参数方式识别：

假设 1（正值性）。 $P(A = 1 \mid W = w) > 0$ 对所有 w 均成立，且其概率为正。

假设 2（一致性）。 $Y = AY_1 + (1 - A)Y_0$。

假设 3（研究内条件可交换性）。

$$E[Y^a \mid W, A, S = s] = E[Y^a \mid W, S = s] \text{ for all } s.$$

我们的估计量的有效性依赖于第四个假设，该假设允许在不同研究之间迁移效应：

假设4（共同结局回归（考虑代理变量的版本））。

$$E(Y \mid \mathcal{T}_s, W, A = a, S = s) = E(Y \mid \mathcal{T}_s, W, A = a, S \in \sigma_s) \text{ for all } s.$$

这是一个随机缺失（MAR）类型的假设，其中 S 在某种意义上可以视为缺失指示变量，因为缺失情况因研究而异。

我们还可以引入第五个假设。识别并不需要该假设，但它允许在不同研究之间更多地借用信息，从而有助于提高效率：

假设 项 5（结局代理变量的共同分布）。 $\mathcal{T}_s \perp S \mid W$， A 对所有 s .

这意味着，在给定治疗分配和基线协变量的条件下，$\mathcal{T}_s$ 的分布在各项研究中相同。在这一附加假设下，识别结果可简化为：

$$ATE = \sum_{s=1}^{\mathcal{S}} \pi_s E[E\{E(Y \mid \mathcal{T}_s, W, A = 1, S \in \sigma_s) \mid W, A = 1\} - E\{E(Y \mid \mathcal{T}_s, W, A = 0, S \in \sigma_s) \mid W, A = 0\} \mid S = s].$$

在正在进行的研究中，我们开发了一种简单代入估计量：将每个期望替换为基于回归的估计值，并将外层期望替换为经验均值。

对于结局代理盲法方法，除前三个标准内部效度假设外，假设4被一个略有不同的均值结局可交换性假设所取代：跨研究假设（关于 S的可交换性）（Dahabreh 和 Hernán，2019年；Lesko 等人，2017）：

假设6 （共同结局回归（代理盲法版本））。

$$E(Y \mid W, A = a, S = s) = E(Y \mid W, A = a, S \in \sigma_s) \text{ for all } s.$$

假设4与假设6 的区别在于，它针对每项研究还对 $\mathcal{T}_s$ 进行条件化s。假设4 和5 共同蕴含假设6。本文仅讨论违反假设4时的敏感性分析。当假设5被违反时，基于假设4 的平均处理效应估计量（即基于识别公式（1）的代入估计量）仍然是一致的。

3 刻画违反识别假设所导致的偏倚

的有效性 Ψ^{ATE} 取决于关键假设4。该假设要求，在存在结局缺失数据与不存在结局缺失数据的研究之间，给定处理、协变量和结局代理变量后，条件结局均值不存在异质性（Y）数据。这使得条件结局均值，以及相应的平均处理效应和条件平均处理效应，可以从一项研究推广到其他研究并进行估计。

在实践中，这可能是一项强假设，而且无法通过观测数据进行检验。例如，在此前未发表的工作中，我们利用NIMH认知训练与矫正研究数据库（DoCTRS）中的三个试验的协调数据，估计了认知矫正（CR）疗法对社会行为量表（SBS）得分的平均治疗效应；该得分用于衡量社会功能。然而，CR疗法的有效程度，尤其是在功能和职业结局方面，尚不明确，且有研究指出，其效果可能因处理实施的情境而异（Barlati 等人，2013；Combs 等人，2008；McGurk 等人，2007年；Wykes 等人，2007年，2011年）。当这一假设遭到违背时，上一节所述的代入估计量将产生偏倚。因此，我们考察两种敏感性分析策略，以评估在不同程度的假设违背下估计结果的稳健性。

为量化违反程度，将偏差函数定义为：

$$
\begin{aligned}
u(A=1,\mathcal{T}_s,W) &= E(Y \mid \mathcal{T}_s,W,A=1,S=s) - E(Y \mid \mathcal{T}_s,W,A=1,S\in\sigma_s),\\
u(A=0,\mathcal{T}_s,W) &= E(Y \mid \mathcal{T}_s,W,A=0,S=s) - E(Y \mid \mathcal{T}_s,W,A=0,S\in\sigma_s) \qquad (2)
\end{aligned}
$$

那么，当假设 14 被违反时，方程（4 ）变为：

$$\begin{aligned}\text{ATE} = \sum_{s=1}^{s^*} \pi_s E\{E(Y \mid W, A=1, S=s) - E(Y \mid W, A=0, S=s) \mid S=s) \\ + \sum_{s=s^*+1}^{\mathcal{S}} \pi_s E\left[E\left\{E\left(Y \mid \mathcal{T}_s, W, A=1, S \in \sigma_s\right) \mid W, A=1, S=s\right)\right\} \\ -E\left\{E\left(Y \mid \mathcal{T}_s, W, A=0, S \in \sigma_s\right) \mid W, A=0, S=s\right)\right\} \mid S=s] \\ + \sum_{s=s^*+1}^{\mathcal{S}} \pi_s E[E\left\{u\left(A=1, \mathcal{T}_s, W\right) \mid W, A=1, S=s\right\} \\ - E\left\{u\left(A=0, \mathcal{T}_S, W\right) \mid W, A=0, S=s\right\} \mid S=s],\end{aligned}$$

其中最后一项求和无法识别。因此，研究 s 的研究特异性偏倚为：

$$\begin{aligned}&E\left[E\left\{u\left(A=1, \mathcal{T}_s, W\right) \mid W, A=1, S=s\right\} - E\left\{u\left(A=0, \mathcal{T}_s, W\right) \mid W, A=0, S=s\right\} \mid S=s\right] \\ &= E[\delta^*(W) | S=s]. \qquad (3)\end{aligned}$$

通过重新排列各项， $\delta^*(W)$ 也可以写为：

$$\begin{aligned}&E\left[E\left(Y \mid \mathcal{T}_s, W, A=1, S=s\right) - E\left(Y \mid \mathcal{T}_S, W, A=1, s \in \sigma_s\right) \mid W, A=1, S=s\right] \\ &\quad - E\left[E\left(Y \mid \mathcal{T}_s, W, A=0, S=s\right) - E\left(Y \mid \mathcal{T}_s, W, A=0, s \in \sigma_s\right) \mid W, A=0, S=s\right] \\ &= E(Y \mid W, A=1, S=s) - E(Y \mid W, A=0, S=s) \\ &\quad - \{E\left[E\left(Y \mid \mathcal{T}_s, W, A=1, s \in \sigma_s\right) \mid W, A=1, S=s\right] \\ &\quad\quad - E\left[E\left(Y \mid \mathcal{T}_s, W, A=0, s \in \sigma_s\right) \mid W, A=0, S=s\right]\}. \qquad (4)\end{aligned}$$

除非假设5 成立，否则后一项无法简化。

4 与未纳入随访替代结局的情形下偏差函数的比较

在近期的研究中，Dahabreh 和 Hern´an (2019年) 针对一种类似情形开展了可迁移性的敏感性分析，该情形涉及两类研究，其中一类存在结局缺失。基本情形中考虑了两项研究（结局变量的缺失情况以二元指标 S表示）。用我们的符号来描述这一情形，只需写作 $\sigma_0 = \sigma_1 = \{1\}$ （即，研究 $S = 1$ 中观测到的感兴趣结局用于插补研究 $S = 0$的条件结局均值）。等价地，为便于解释基本情形，令 $S = 1$ 和 $S = 0$ 分别表示观测到和未观测到感兴趣主要结局的研究。

在用于插补条件潜在结局的模型未利用来自 $\mathcal{T}_s$的信息时，Dahabreh 和 Hern´an (2019年) 定义：

$$u(A = a, W) = E[Y \mid A = a, W, S = 1] - E[Y \mid A = a, W, S = 0].$$

这些偏差函数之间的差异可由此得出：

$$\begin{aligned}\delta(W) &= u(A = 1, W) - u(A = 0, W) \\ &= E[Y^1 - Y^0 \mid W, S = 1] - E[Y^1 - Y^0 \mid W, S = 0]\end{aligned}$$

这一表达式可以定性地理解为两项研究的条件平均处理效应之差。这种定性解释有助于我们在考察结果的稳健性时，构思并思考更合适的敏感性参数取值和范围。更具体地说，假设较高的结局水平更受青睐，如果我们认为结局缺失研究中的参与者从处理中获益较少，那么可以假设真实的 δ 为正，反之亦然（Dahabreh 和 Hern´an，2019年）。由于我们的偏差函数以代理结局集为条件，因此

项 $\delta^*(W)$ 在（4）中遗憾的是无法进一步简化为更易于解释的统计实体。当我们将 $\mathcal{T}_s$ 设为空集时，偏差函数 $\delta^*(W)$ 会简化为相同的表达式。

5 通过敏感性分析处理共同结局回归假设的违反

我们考虑两种情形，假设偏差项 $u(A=1,\mathcal{T}_s,W)$ 和 $u(A=0,\mathcal{T}_s,W)$ 分别为：1）常数；2）结果代理变量和/或基线协变量的有界函数。第一种情形对偏差项作出了更强的假设。另一方面，第二种情形所需的假设较弱，但允许偏差项为非常数。

5.1 假设偏差函数为某个固定值 s

虽然假设偏差函数依赖于某些基线协变量可能更为合理，但为便于实施敏感性分析，也可以假设它们为常数。当 $u(A=1,\mathcal{T}_s,W)$ 和 $u(A=0,\mathcal{T}_s,W)$ 与基线协变量 W 和结果代理变量集 $\mathcal{T}_s$相互独立时，偏差函数的条件期望以及由此得到的项 $\delta^*(W)$ 在 (3) 中可化简为：

$$\delta = u_1 - u_0, \text{ where } \delta, u_1, \text{ and } u_0 \in \mathbb{R} \tag{5}$$

敏感性分析通过将上述偏差项加回识别公式 Ψ^{ATE} 来进行校正；该公式依赖于共同结局回归假设。

$$
\begin{aligned}
ATE =& \sum_{s=1}^{s^*} \pi_s E\left\{E\left(Y \mid W, A=1, S \in \sigma_s\right)-E\left(Y \mid W, A=0, S \in \sigma_s\right) \mid S=s\right\} \\
&+\sum_{s=s^*+1}^{\mathcal{S}} \pi_s E\left[E\left\{E\left(Y \mid \mathcal{T}_s, W, A=1, S \in \sigma_s\right) \mid W, A=1, S=s\right\}\right. \\
&\left.-E\left\{E\left(Y \mid \mathcal{T}_s, W, A=0, S \in \sigma_s\right) \mid W, A=0, S=s\right\} \mid S=s\right]+\sum_{s=s^*+1}^{\mathcal{S}} \pi_s\left(u_1-u_0\right) \\
=& \Psi^{ATE}+\sum_{s=s^*+1}^{\mathcal{S}} \pi_s\left(u_1-u_0\right) \qquad (6)
\end{aligned}
$$

其中， u_1 和 u_0 是标量。

在实践中，真实偏差项通常是未知的。因此，一种策略是提出一组敏感性参数网格，覆盖真实偏差项可能落入的数值范围。这组敏感性参数网格可以利用领域知识加以设定。随后，在估计步骤中，我们可以使用所提出的方法，将不同的敏感性参数加回估计的平均处理效应，以校正偏差项。这样还可以观察随着敏感性参数变化，估计的平均处理效应所呈现的变化。

5.2 有界协变量依赖型偏倚函数

人们也可能认为，偏差项并非在基线协变量和/或结局代理变量的所有取值水平上都保持恒定。当固定值偏差项这一假设被认为过强，而又无法根据对数据生成机制的现有认识有把握地确定偏差项的函数形式（通常情况如此）时，我们仍可以在无需正确设定偏差项的情况下，获取有关真实平均处理效应的一些信息。若改为假设偏差项是某些有界函数，我们可以通过改变这些函数的界限，计算出一个围绕（naïve）平均处理效应估计的界限范围，其中包含真实平均处理效应。这有助于了解真实平均处理效应可能与该估计值相差多远

而该估计值受偏差项界限的约束。

确定偏倚项的界限可表述为最大化和最小化目标函数：

$$E[E[u(A=1,\mathcal{T}_s,W)\mid W,A=1,S=s]-E[u(A=0,\mathcal{T}_s,W)\mid W,A=0,S=s]\mid S=s]$$

满足以下约束条件：

$$|u(A=1,\mathcal{T}_s=t_s,W=w)|\leq\gamma_1$$
$$|u(A=0,\mathcal{T}_s=t_s,W=w)|\leq\gamma_0$$

对于所有 t_s和 w，这意味着 $|E[u(A=1,\mathcal{T}_s,W)\mid W,A=1,S=s]|\leq\gamma_1$以及 $|E[u(A=0,\mathcal{T}_s,W)\mid W,A=0,S=s]|\leq\gamma_0$其中 $\gamma_1,\gamma_1\in\mathbb{R}^+$。

于是我们有 $-(\gamma_1+\gamma_0)\leq u(A=1,\mathcal{T}_s,W)-u(A=0,\mathcal{T}_s,W)\leq\gamma_1+\gamma_0$。如果没有理由认为我们对一个偏差函数界限的了解多于另一个（通常情况如此），我们可以简单地指定一个标量敏感性参数 γ，使其取 γ_1和 γ_2中的较大值，此时有 $-2\gamma\leq u(A=1,\mathcal{T}_s,W)-u(A=0,\mathcal{T}_s,W)\leq 2\gamma$。

根据方程（6），尽管我们不知道偏差函数的形式 $u(A=1,\mathcal{T}_s,W)$和 $u(A=0,\mathcal{T}_s,W)$，但我们可以利用朴素估计值周围的界限，部分恢复真实平均处理效应：¨

$$\Psi^{ATE}-2\max(\gamma_1,\gamma_0)\leq ATE\leq\Psi^{ATE}+2\max(\gamma_1,\gamma_0)$$
$$\Psi^{ATE}-2\gamma\leq ATE\leq\Psi^{ATE}+2\gamma \tag{7}$$

如果偏差函数实际上被限制在某个不大于我们为敏感性界限指定值的数值范围内，那么真实平均处理效应将落在 $[\Psi^{ATE}-2\gamma,\Psi^{ATE}+2\gamma]$之间。由此，即使不对函数形式作出假设，也能部分识别真实平均处理效应

的 $u(A=1,\ \mathcal{T}_s,\ W)$ 和 $u(A=0,\ \mathcal{T}_s,\ W)$。随后，可以使用识别公式（1）的代入估计量的自助法标准误，确定分别应从上界和下界加上及减去的数值，以便为敏感性参数每个取值对应的部分识别集构造置信区间。由于敏感性界限是敏感性参数的确定性函数，因此只需进行一次自助抽样。

6 模拟

6.1 数据生成机制

我们考虑两项研究的情形，其中 $S=1$ 表示可获得主要结局的研究。我们从两项研究中分别随机抽取样本，样本量均为 $\mathrm{n}=100$ 。数据生成机制如下。 $W,\ T_0$ 服从相互独立的标准正态分布，而 T_1 服从均值和方差均为 1 的正态分布。然后

$$
\begin{aligned}
T &= I(A=1)\times T_1 + I(A=0)\times T_0 \\
Y^0 &= -4T_0 + W + \epsilon_0 \\
Y^1 &= 4T_1 + W + \epsilon_1 \\
Y &= I(A=1)\times Y_1 + I(A=0)\times Y_0
\end{aligned}
$$

其中 $\epsilon_1,\ \epsilon_0 \sim N(0,\ 1)$。

根据这些设定， T 完全中介了 A 与 Y 之间的关系（将从 A 到 Y 的直接效应限制为 0）。因此，真实平均处理效应 $=4$。这也是一个更基础的情形，其中向量 T 在所有研究中均可观测。

由于 DoCTRS数据库 由随机临床试验组成，在我们的基础设置中，我们将边际概率 $\mathrm{P(A}=1)=0.5$ 设定为代表——

随机治疗分配。这种治疗分配满足正值性和可交换性假设。

具体而言，为了纳入缺失结局的研究中两类研究条件结局均值之间的差异，我们在反事实结局 Y_0 和 Y_1中加入了常数偏差项。与数据生成步骤类似，我们保留了与相应治疗分配相对应的观测反事实结局，这满足一致性假设。根据（5），我们有：

$$\begin{aligned} Y^0_{S=1} &= Y^0_{S=0} + u_0 \\ Y^1_{S=1} &= Y^1_{S=0} + u_0 + \delta \end{aligned} \tag{8}$$

当 $u_0 \in \{-3, 0, 3\}$时， $\delta \in \{-2, 0, 2\}$。

因此，计算平均处理效应时，偏倚简化为单个参数 δ，因为它不再是 u_0 的函数：

$$E(Y^1 - Y^0 \mid S = 1) = E(Y^1 - Y^0 \mid S = 0) + \delta \tag{9}$$

当偏差项是基线协变量和替代结局的函数时，我们对真实偏倚作如下设定：

$$\begin{aligned} u_0 &= b_0 \times \sin(T_s + W) \\ u1 &= b_1 \times \tfrac{\exp(T_s+W)}{1+\exp(T_s+W)} \end{aligned}$$

当 $b_0 \in \{2, 3, 4\}$ 和 $b_1 \in \{1, 2, 3\}$ 时。

6.2 在估计步骤中调整敏感性参数 p

存在非零偏倚时，若指定敏感性参数 δ 的值使其等于真实的 δ，则偏差调整后的平均处理效应估计往往比调整前更接近

真实平均处理效应。此外，相应的95%置信区间预计有95%的情况下会覆盖真实平均处理效应。尽管可以通过在所有模拟中使用自助法置信区间，以更稳健的方式考察覆盖概率，但在图1、2和A.1-A.4中，除一个实例外，95%置信区间均在反映假设违背程度的敏感性参数取值处覆盖了真实平均处理效应，这与我们的预期一致。

情景1。 假设偏差项为常数时，一个自然的做法是为两个标量分别指定一个由敏感性参数构成的二维网格 u_0 和u_1。然而，根据（8），这等同于指定 u_0 （或 u_1）以及 δ。事实上，由于 u_0 （或u_1）作为常数项会在调整过程中抵消，因此只需指定一个敏感性参数 δ （9）。我们还注意到， δ 为 0 并不一定意味着满足假设 4 ，因为偏差项 u_0 和 u_1 可能恰好相互抵消。

实施敏感性分析时，我们按以下步骤进行：

1. 指定敏感性参数网格 δ。网格范围应足够宽，以包含真实的 δ。

2. 使用（¨1）中的识别结果，估计朴素迁移的平均处理效应

3. 依次将敏感性参数网格中的值加到朴素¨估计的平均处理效应上，并使用（6）中的结果，得到偏差校正后的平均处理效应估计。

然后，我们将不同敏感性参数下的偏差校正后的估计与真实平均处理效应进行比较并绘图。此外，我们对偏差校正后的估计进行自助法重抽样，以获得95%的置信区间，并探究不同 u_0 和 δ取值下的覆盖率。

情景2。 当我们希望对偏倚的函数形式作出最少的假设时，仍然可以通过以下步骤对真实平均处理效应进行敏感性分析：

1. 指定一个敏感性参数网格，称为 γ ，其范围可能包含真实偏差函数的上下界

2. 使用（¨1）中的识别结果计算“朴素”平均处理效应估计

3. 使用（7）构造估计的平均处理效应的上下界，其中 γ 由敏感性参数替代。

我们还绘制了朴素平均处理效应估计，以及在每个敏感性参数取值下这些估计的上下界。在实践中，偏差函数当然是未知的，无法根据观测数据进行估计。因此，在设定敏感性参数网格时，分析者需要运用关于数据生成机制的领域知识来选择 δ 和 γ的取值。

随后，我们通过模拟考察调整估计量的表现。在第一种情形中，我们关注正确调整后的点估计在总体平均处理效应以及结局缺失研究中的平均处理效应方面是否总体无偏，并考察在不同程度的假设违背下（即在真实 u_0 和 δ的不同取值下）95%置信区间的覆盖率。在第二种情形中，我们考察是否能正确界定真实平均处理效应。

6.3 模拟结果

6.3.1 偏差项为常数

我们考察了在假设4不同程度违背的情况下，本方法产生的估计值，以及在纳入指定敏感性参数之前和之后的估计值。图1展示了在一次模拟中，偏差项的幅度和方向不同时，使用本方法对真实总体平均处理效应所得到的估计值（95%置信区间）。

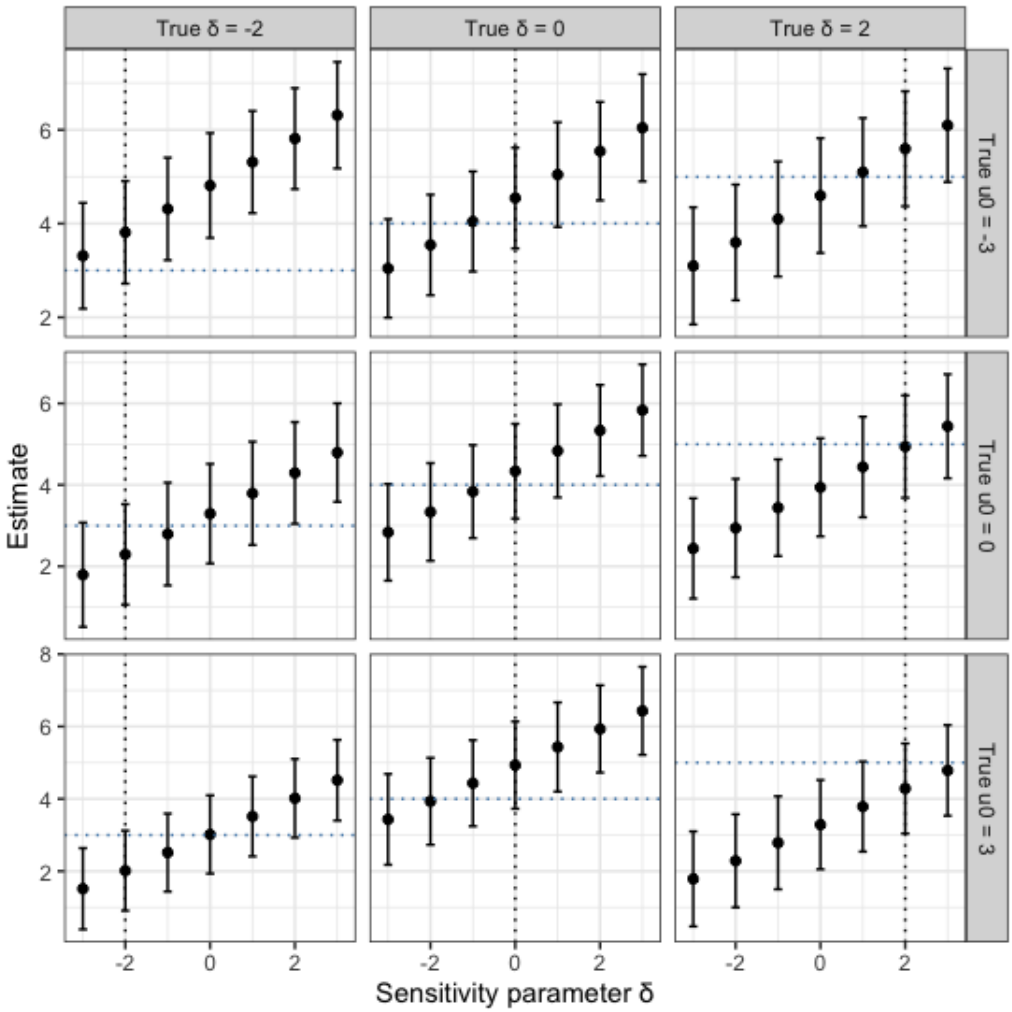

图1：在真实偏倚和敏感性参数取不同值时，经敏感性参数调整的平均处理效应估计与真实总体平均处理效应的对比；每项研究的 n=100，95%置信区间根据 1000 个自助抽样样本构建。当敏感性参数为 $\delta = 0$ 时，调整后估计对应于未调整估计值。水平点线表示给定真实 δ 时的真实平均处理效应；垂直点线表示敏感性参数 δ 等于真实 δ

在存在非零偏倚时，当敏感性参数 δ 被指定为等于真实 δ时，进行偏差调整后的平均处理效应估计相比调整前往往更接近真实平均处理效应。此外，相应的95%置信区间预计有95%的概率覆盖真实平均处理效应。尽管可以通过在所有模拟中使用自助法置信区间，以更加稳健的方式进一步考察覆盖概率，但在图1、2和A.1-A.4中，当敏感性参数的取值反映假设违背程度时，95%置信区间除一次外均覆盖了真实平均处理效应，该例位于

与我们的预期一致。

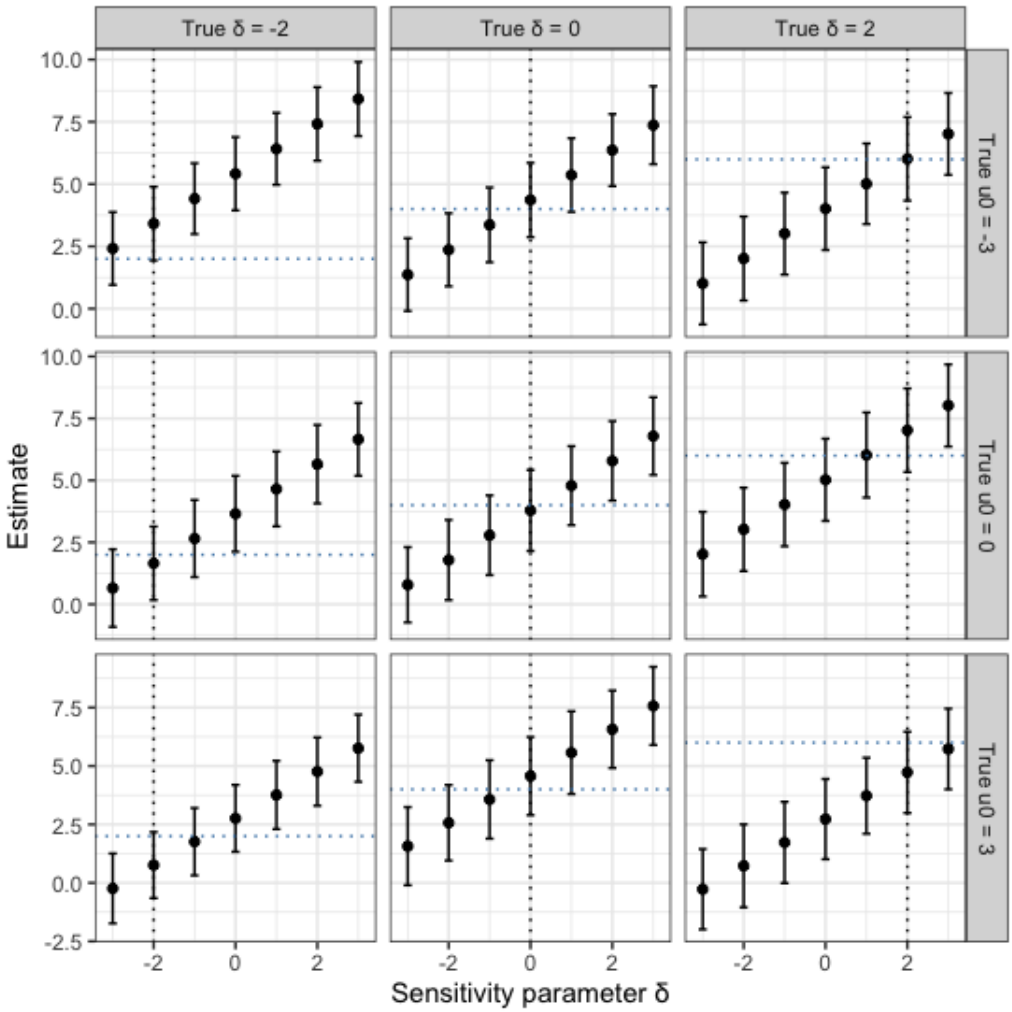

图2：在结局未观测到的研究中，经敏感性参数调整的平均处理效应估计与真实研究特异性平均处理效应的对照图，展示了不同真实偏倚值和敏感性参数值下的结果；每项研究的 n=100 ，95% 置信区间根据 1000 个自助抽样样本构建。当敏感性参数为 $\delta = 0$ 时，调整后估计与未调整估计值相对应。水平点线表示给定真实 δ 时的真实研究特异性平均处理效应；垂直点线表示敏感性参数 δ 等于真实 δ

图2显示了同一模拟数据中结局缺失研究的研究特异性平均处理效应估计（偏差调整前后）的相似结果。与图1中的结果相比，使用正确的敏感性参数进行调整后，结局缺失研究中主要结局未调整估计的置信区间更少包含真实平均处理效应，而95%置信区间更频繁地包含真实平均处理效应。图2还展示了一个例子，其中推断对违反假设的程度很敏感：当其幅度介于-1和-2之间（，左下图）时，95%置信区间会从不包含零变为包含零。

推断对我们的假设被违反的程度较为敏感：当其处于 -1 到 -2 之间时（$u_0 = -3$，左下图），95%置信区间会从不包含零变为包含零。

当我们增大样本量（n=200 和 n=500）时，发现这些单个估计的误差总体有所降低（图A.1、A.3）。在大多数情况下，即使调整后估计存在误差，95%置信区间自助法置信区间仍能提供良好的覆盖率（图1、A.1、A.3）。与合并两项研究的总体平均处理效应相比，在结局缺失研究中估计研究特异性效应时，误差的降低和覆盖率的改善更为显著（图A.2、A.4）。

我们还在相同的数据生成机制下进行了 1000 次模拟，并为每次模拟计算了未调整估计和经敏感性参数调整后的估计。随后，我们展示了真实偏差值的每种组合下这些估计的均值、第 2.5 和第 97.5 分位数。可以看到，在 1000 次模拟中取平均后，当敏感性参数使用真实的 δ 值时，调整后估计非常接近真实平均处理效应（图 3、4）。

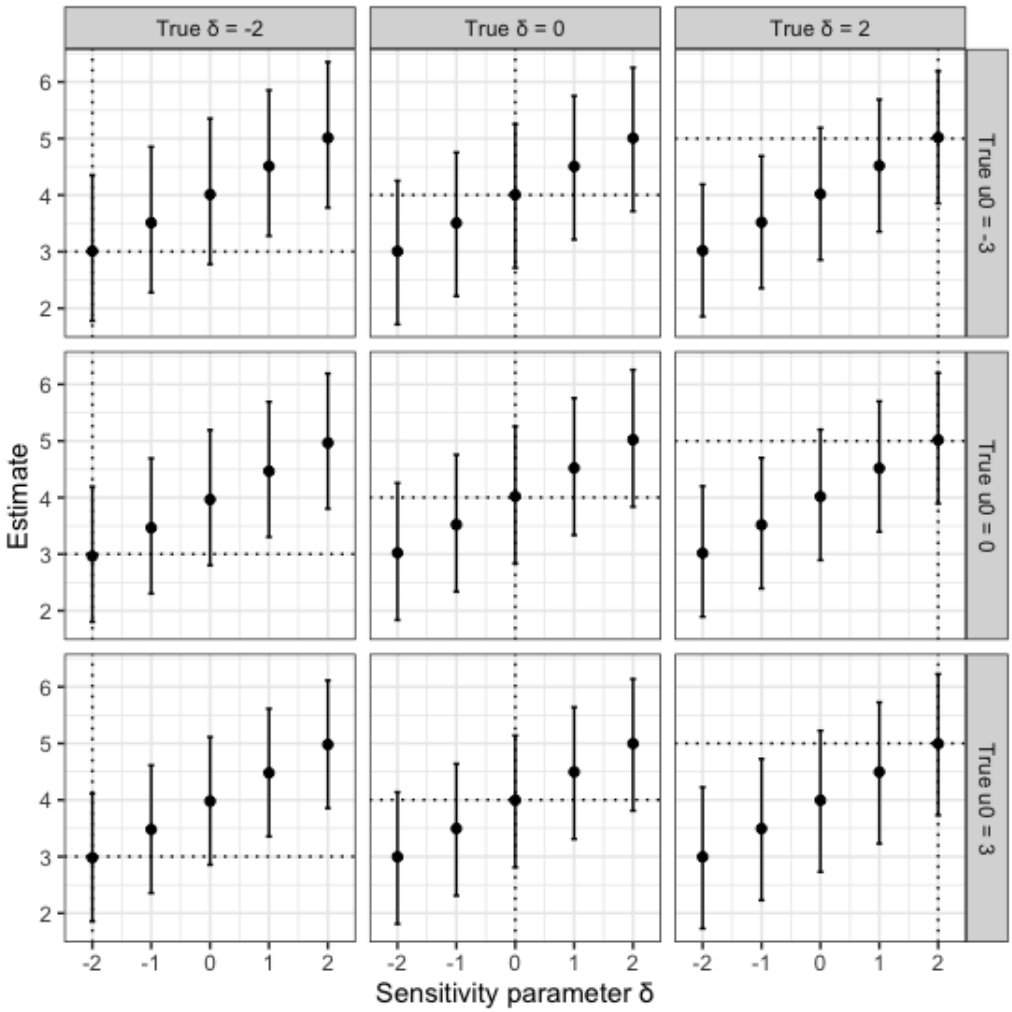

图3：在真实偏倚敏感性参数的不同取值下，经敏感性参数调整的平均处理效应估计与真实总体平均处理效应的对比；均值、第2.5和第97.5分位数基于1000次模拟得出。当敏感性参数为 $\delta = 0$时，调整后估计对应于未调整估计值。水平点线表示给定真实 δ时的真实总体平均处理效应；垂直点线表示敏感性参数 δ 等于真实 δ

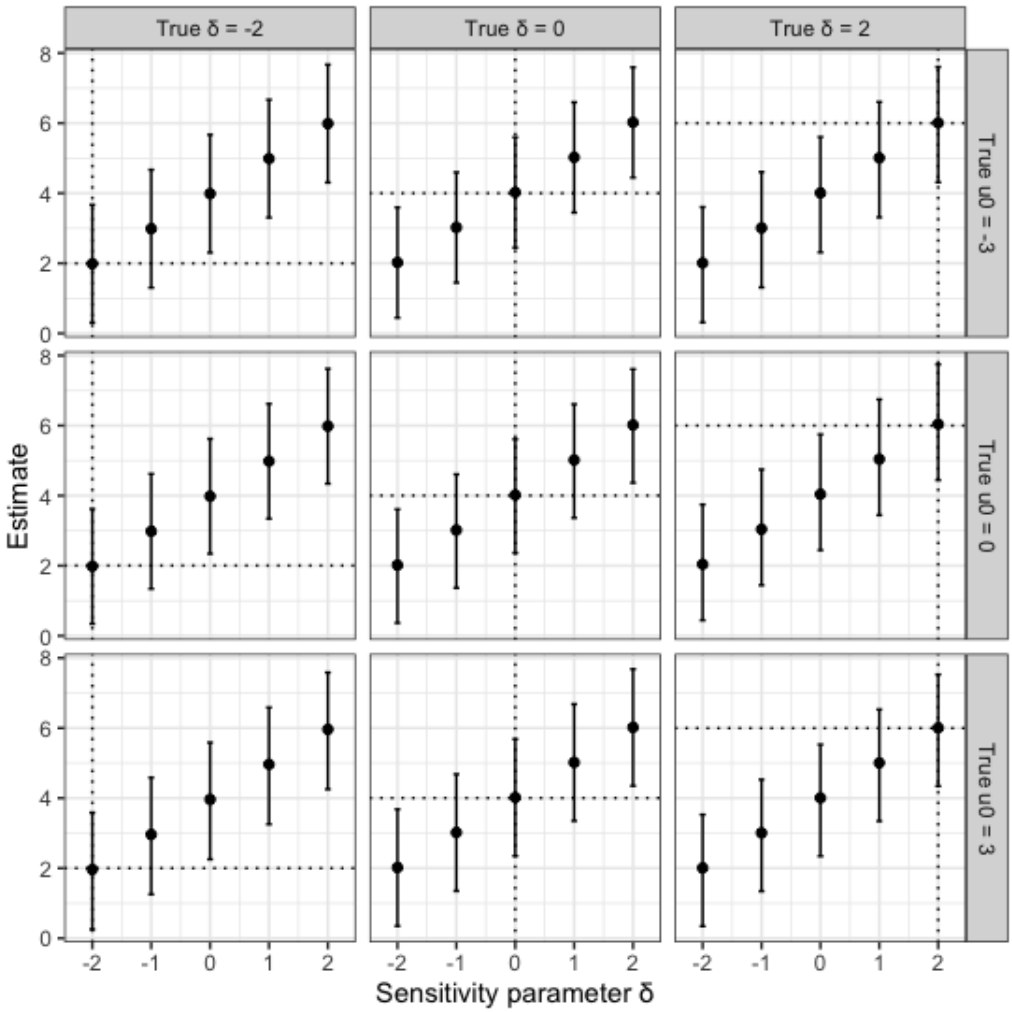

图4：在不同真实偏倚和敏感性参数取值下，结局缺失研究中经敏感性参数调整的平均处理效应估计与真实平均处理效应的对比；均值、第2.5和第97.5分位数由1000次模拟得出。当敏感性参数 $\delta = 0$时，调整后估计对应未调整估计值。水平点线表示给定真实 δ时的真实研究特异性平均处理效应；垂直点线表示敏感性参数 δ 等于真实 δ

使用近似敏感性参数 δ时（$\delta \in \{-1, 1\}$ 真实 $\delta \in \{-2, 2\}$时），调整后估计的中间95%取值也涵盖真实平均处理效应，而未调整估计值则没有（图4）。

图5 比较了在 1000 次模拟中，Dahabreh 等人 Dahabreh 等人(2020b)；Lesko 等人(2017) 的结果代理盲法与我们所提出的方法在估计误差及相关推断的敏感性方面的表现。

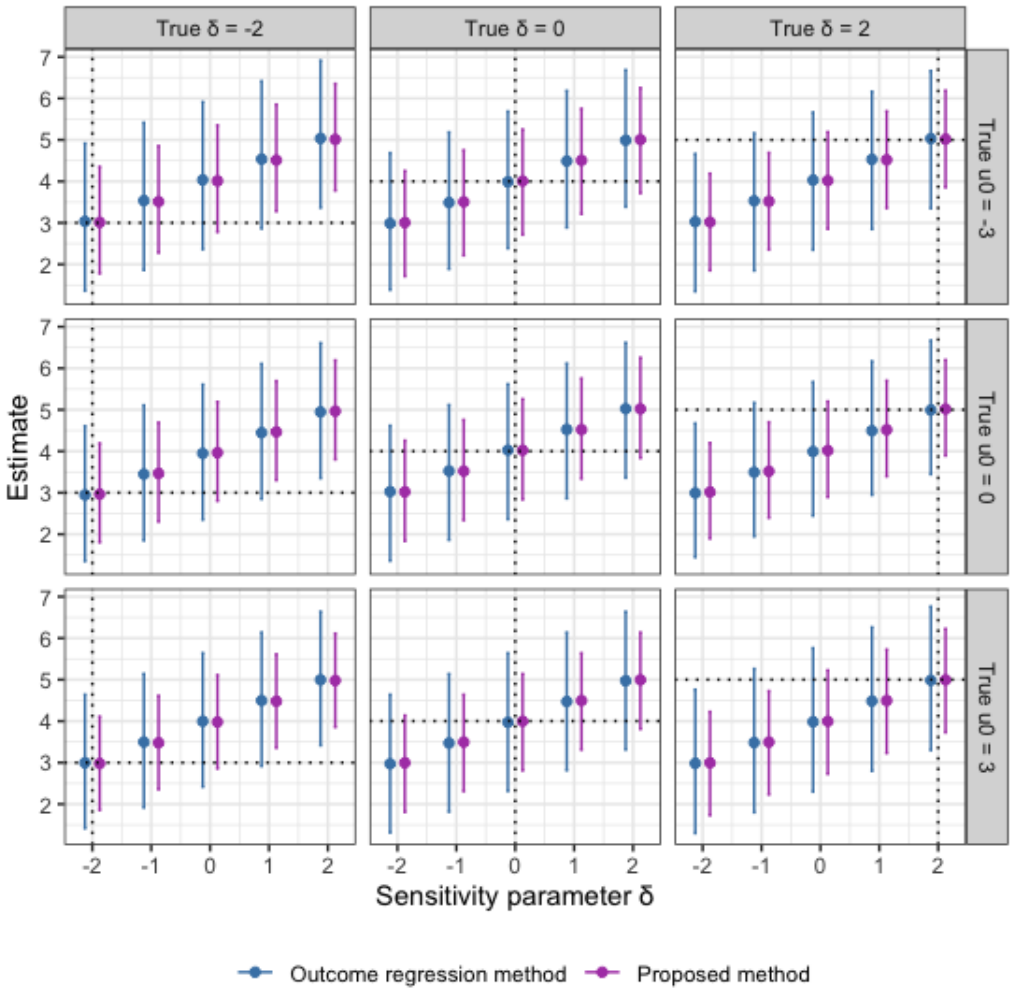

图5：使用我们所提出的方法和结局代理盲法得到的经敏感性参数调整的平均处理效应估计；展示了 1000 次模拟所得的均值、第 2.5 和第 97.5 分位数。当敏感性参数为 $\delta = 0$ 时，调整后估计对应于未调整估计值。水平点线表示给定真实 δ 时的真实总体平均处理效应；垂直点线表示敏感性参数 δ 等于真实 δ

两种方法所得估计的分布均以真实参数为中心。不过，利用结果代理变量的信息时，估计往往更加精确（表现为 2.5 至 97.5 分位数范围更窄）。效率提升会影响敏感性分析，因为在违反识别假设的程度相近时，所得推断的敏感性较低 4。

假设 4 意味着 u_0 和 u_1 均为 0。因此，真实的 δ 也等于 0。这表明，可以迁移条件潜在结果均值，进而

条件平均处理效应的迁移无需引入偏倚（图 3中间列）。我们还观察到，当 δ 为 0 时，无论 u_0 （以及u_1）取何值，未调整估计量也不存在偏倚（图 3中间列）。在这两种情况下，都不需要进行偏差校正，而纳入非零的 δ 敏感性参数反而会使估计产生偏倚。

6.3.2 作为有界函数的偏差项

当敏感性参数 γ 大于或等于真实函数界的最大值 $\{\gamma_0, \gamma_1\}$ ，且 γ_0 和 γ_1的偏差函数均受 γ_0 和 γ_1 界定时，界限始终包含真实平均处理效应（图 6 ）。

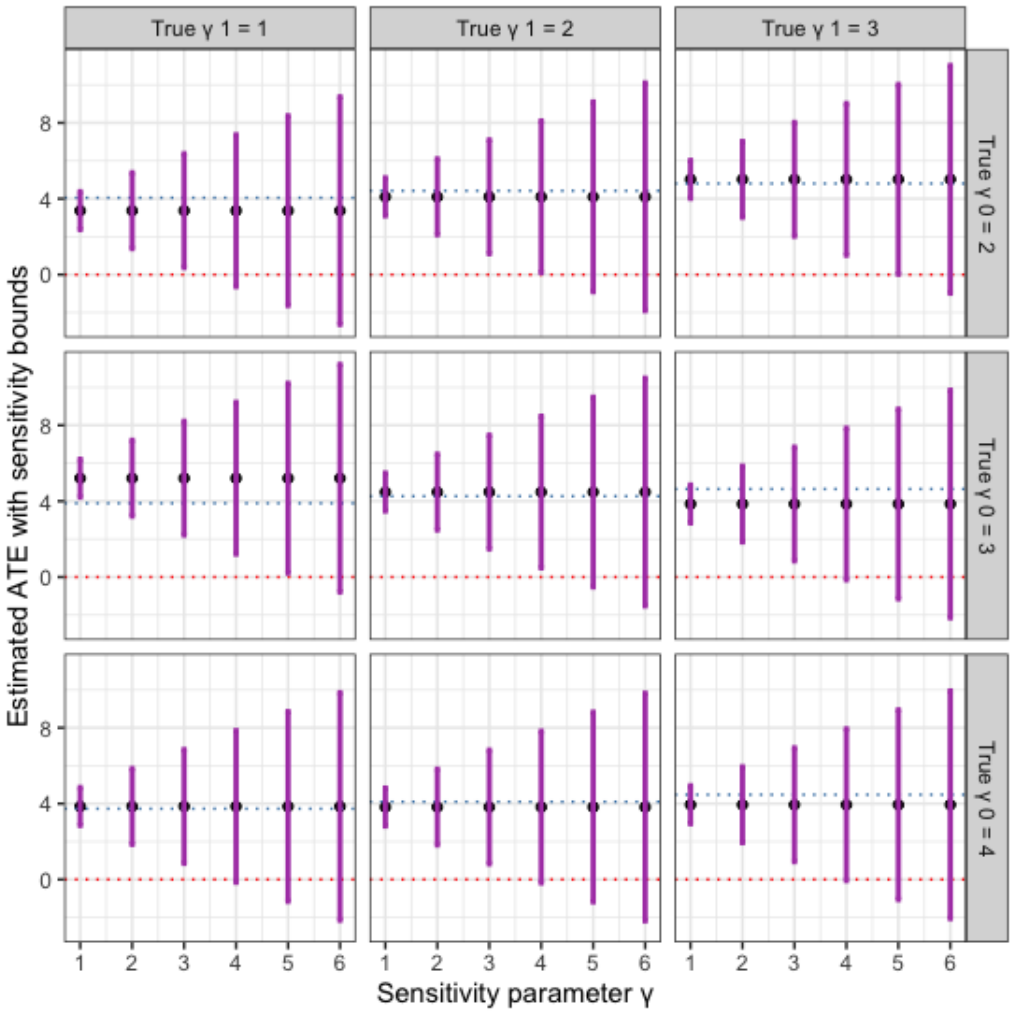

图6：不同真实偏倚和敏感性参数取值下，带有敏感性界限的平均处理效应估计与真实总体平均处理效应的对比。当敏感性参数为 $\gamma = 0$ 时，界限收敛为一个点估计。蓝色水平点线表示在给定真实偏差函数时的真实研究特异性平均处理效应

尽管这种方法对偏差函数形式只需极少的假设，但也可能较为保守，因为真实偏差函数不太可能在函数定义域内处处达到界限。例如，图 6 的下方三个面板显示，当敏感性参数 γ 大于或等于 max(true γ_0, true γ_1) 时，虽然估计值的界限包含真实平均处理效应，但也包含零假设值零。另一方面，这些界限不依赖于常数偏差函数这一假设，而我们往往没有理由相信该假设成立。在此，我们通过

模拟表明，即使采用较宽松且更可信的假设进行敏感性分析，仍能为关注参数提供有用的信息。然而，当界限过窄或过宽时，使用有界偏倚函数进行敏感性分析可能分别不准确（即未包含真实参数）或没有用处（即当真实值不为零时却包含零假设值）。

7 讨论

本文讨论了一种数据整合方法，该方法利用随访时测量的关注结局的可用代理变量中的信息来提升效率。随后，我们针对识别假设违背时的因果效应迁移，提出了两种适用于此方法的敏感性分析策略。我们对式（1）中平均处理效应的识别所作的修正，使得在结局代理变量足够强的情况下能够获得效率更高的估计量。因此，我们的偏差函数与Dahabreh 和 Hern´an （2019年）的偏差函数有相似但又不同的解释。

当假设偏差项为常数时，我们可以根据对敏感性参数的设定，得到不同的偏差调整点估计。此外，通过获得偏差调整估计的95%自助法置信区间，我们可以考察在假设违背程度不同的情况下，使用本方法所得推断的稳健性。具体而言，当敏感性参数超过某些取值时，95%置信区间将跨越零假设值0。这些违背程度可能影响推断（即95%置信区间表明结果可能从显著变为不显著）。

此外，当人们认为固定值偏差项的假设过于严格时，我们还提出了采用有界偏倚函数进行敏感性分析，作为一种替代方案。这种方法只需对未观测到的偏差函数作出极少假设，便可进行推断，同时仍能为关注参数提供有用信息。由于所作假设较少，结果更加保守、稳健，因此也更合理可信。

具体而言，尽管我们无法获得点估计，采用有界偏倚函数进行敏感性分析仍然具有参考价值，因为它能够提供有关关注参数总体方向（有益或有害）的信息。如果函数的界限并不接近其极端值、偏差函数通常并不接近其极端值，或者两个偏差函数的极值相差较大，这种方法通常会更加保守。

正确设定偏差函数有助于更精确、有效地估计真实平均处理效应。然而，由于偏差函数通常未知，且无法根据观测数据估计，敏感性分析通常是更切实可行的做法。

开展敏感性分析时，分析者可以先为敏感性参数设定一个宽泛的网格，并考察点估计和95%置信区间的变化（第一种方法），以及估计值周围的界限（第二种方法）。随后，可以寻找仍然支持拒绝零假设的“临界”敏感性参数，即95%置信区间（第一种方法）和估计值周围的界限（第二种方法）均不包含0的情形。利用关于数据生成机制的背景知识，或进一步对该机制作出假设，可以判断更大的偏倚是否合理。如果几乎没有证据表明真实偏差函数超过这些临界敏感性参数，那么我们就更有把握得出结论：观测到的效应及相关推断对于违反可迁移性假设是稳健的（Ding 和 VanderWeele，2016；Cornfield 等人，1959）。

参考文献

Bareinboim, E. and Pearl, J. (2016). 因果推断与数据融合问题。 美国国家科学院院刊, 113(27):7345–7377. 1

Barlati, S., Deste, G., De Peri, L., Ariu, C. 和 Vita, A.（2013）。认知矫正

精神分裂症中的认知矫正：当前状况与未来展望。《精神分裂症研究与治疗》，2013：156084。7

Combs, D. R., Tosheva, A., Penn, D. L., Basso, M. R., Wanner, J. L., 和 Laib, K.（2008）。注意塑造作为改善精神分裂症患者情绪感知缺陷的一种手段。《精神分裂症研究》，105(1-3):68–77。7

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., 和 Wynder, E. L.（1959）。吸烟与肺癌：最新证据及若干问题的讨论。《美国国家癌症研究所杂志》，22(1):173–203。26

Dahabreh, I., Robins, J., Haneuse, S., Robertson, S., Steingrimsson, J., 和 Hern´an, M.（2022年）。将随机试验的推断推广至目标人群的研究中的全局敏感性分析。*arXiv* 预印本 *arXiv:2207.09982*。2

Dahabreh, I. J. 和 Hern´an, M. A.（2019）。将随机试验的推断扩展至目标人群。《欧洲流行病学杂志》，34(8)：719–722。2，6，9，25

Dahabreh, I. J., Petito, L. C., Robertson, S.E., Hern´an, M. A. 和 Steingrimsson, J. A.（2020a）。迈向因果可解释的荟萃分析：将多项随机试验的推断迁移至新的目标人群。《*Epidemiology*》，31(3)：334–344。2

Dahabreh, I. J., Robertson, S. E., Steingrimsson, J. A., Stuart, E. A., and Hernan, M. A.（2020b）。将随机试验的推断扩展至新的目标人群。《*Statistics in Medicine*》，39(14)：1999–2014。2，21

Degtiar, I. 和 Rose, S.（2023年）。可推广性和可迁移性综述。《统计学及其应用年度评论》，10(1)。1

Ding, P.and VanderWeele, T.J.(2016). 无假设敏感性分析。《流行病学》（剑桥，马萨诸塞州），27(3):368。26

Dong, L., Yang, S., Wang, X., Zeng, D., and Cai, J. (2020). 随机临床试验与真实世界证据研究的整合分析。 *arXiv* 预印本 *arXiv:2003.01242*。2

Gechter, M.（2015）。社会实验结果的推广：来自墨西哥和印度的理论与证据。波士顿大学经济学系经济发展研究所。1

H¨unermund, P. 和 Bareinboim, E.（2019）。计量经济学中的因果推断与数据融合。*arXiv* 预印本 *arXiv:1912.09104*。2

Kennedy-Martin, T.、Curtis, S.、Faries, D.、Robinson, S. 和 Johnston, J.（2015）。关于随机对照试验样本代表性的文献综述，以及对试验结果外部效度的启示。《*Trials*》，16(495)。1

Lesko, C.、Buchanan, A.、Westreich, D.、Edwards, J.、Hudgens, M. 和 Cole, S.（2017）。推广研究结果：潜在结果视角。《流行病学》（马萨诸塞州剑桥），28(4):553。2，6，21

McGurk, S., Twamley, E., Sitzer, D., McHugo, G., and Mueser, K.（2007年）。精神分裂症认知矫正的荟萃分析。《美国精神病学杂志》，164(12):1791–1802。7

Nguyen, T., Ebnesajjad, C., Cole, S., and Stuart, E.（2017）。从随机对照试验向目标人群推广治疗效应时，针对未观测调节变量的敏感性分析。《应用统计年鉴》，11(1):225 –247。2

Pearl, J. 和 Bareinboim, E.（2014）。外部效度：从 do-演算到跨人群的可迁移性。载于《概率与因果推断：*Judea Pearl* 文选》，第 451–482 页。2

Tan, Y.、Papez, V.、Chang, W.、Mueller, S.、Denaxas, S. 和 Lai, A.（2022年）。比较临床试验人群代表性与真实世界人群：一项外部效度分析，涵盖英格兰的 43,895 项试验、5,685,738 名个体、989 种独特药物和 286 种疾病。《柳叶刀·健康长寿》，3(10)：e674–e689。1

Tipton, E. (2014). 您的实验具有多强的普适性？用于比较实验样本与人群的指数。《教育与行为统计学杂志》,39(6):478–501. 1

Westreich, D., Edwards, J. K., Lesko, C. R., Stuart, E., and Cole, S. R. (2017). 使用抽样逆优势权重提高试验结果的可迁移性。《美国流行病学杂志》,186(8):1010–1014. 2

Wykes, T., Huddy, V., Cellard, C., McGurk, S., and Czobor, P. (2011). 精神分裂症认知矫正的荟萃分析：方法学与效应量。《美国精神病学杂志》,168(5):472–485. 7

Wykes, T., Reeder, C., Landau, S., Everitt, B., Knapp, M., Patel, A., and Romeo, R. (2007). 精神分裂症认知矫正治疗：随机对照试验。《英国精神病学杂志》,190(5):421–427. 7

A 补充图

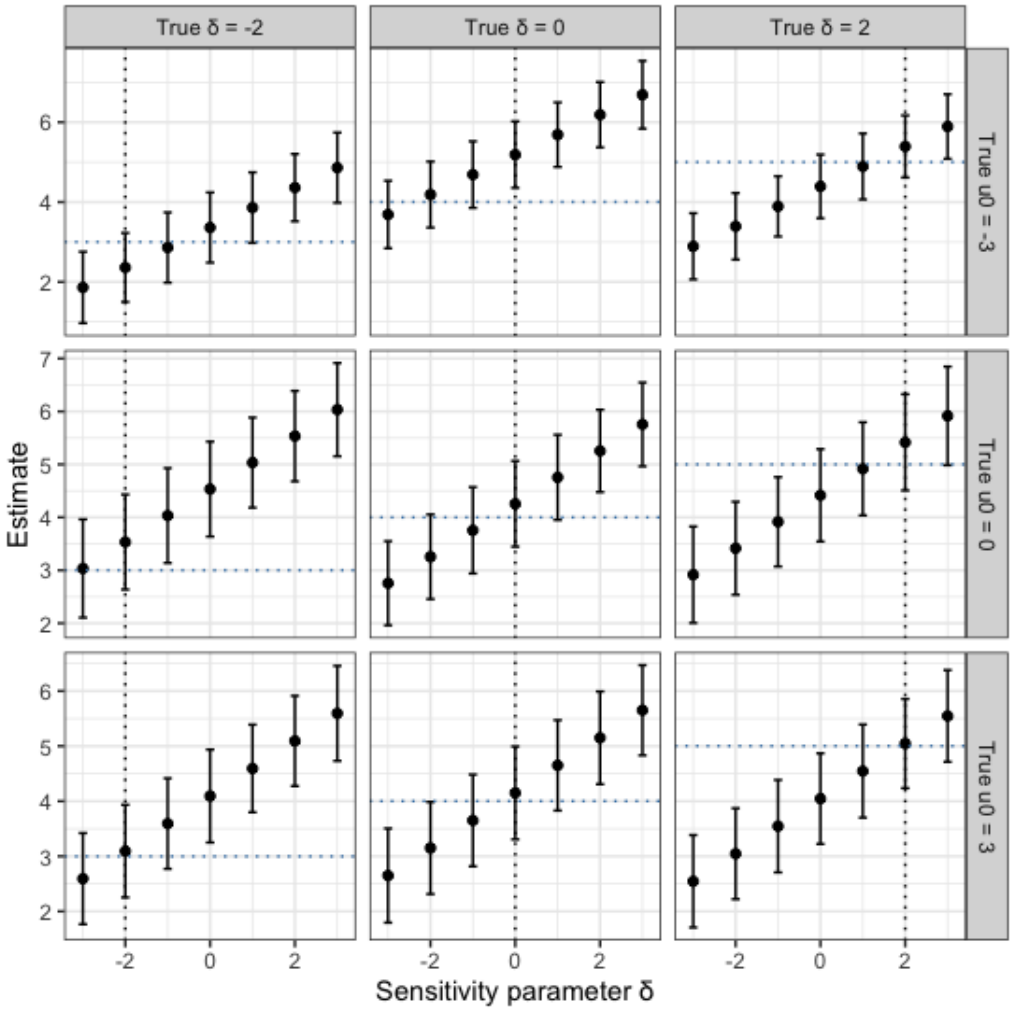

图 A.1：不同真实偏倚和敏感性参数取值下，经敏感性参数调整的平均处理效应估计与真实总体平均处理效应的对比；每项研究的样本量为 n=200，95%置信区间由 1000 个自助抽样样本构建。当敏感性参数 $\delta = 0$ 时，调整后估计与未调整估计值相对应。水平点线表示给定真实 δ 时的真实总体平均处理效应；垂直点线表示敏感性参数 δ 等于真实 δ

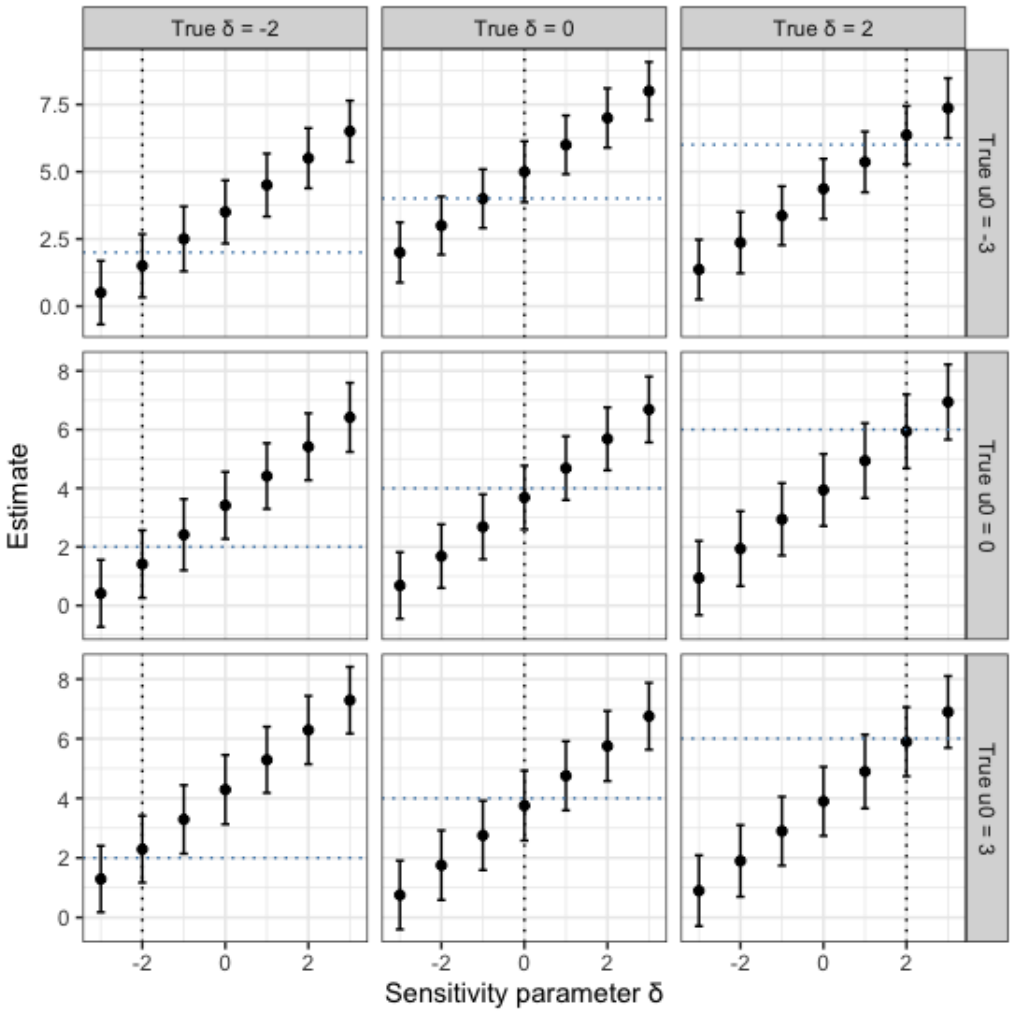

图 A.2：在不同真实偏倚和敏感性参数取值下，结局缺失研究中经敏感性参数调整的平均处理效应估计与真实平均处理效应的对比；每项研究的 n=200 ，95%置信区间根据 1000 个自助抽样样本构建。当敏感性参数 $\delta = 0$时，调整后估计对应于未调整估计值。水平点线表示给定真实 δ 时研究特异性平均处理效应的真实值；垂直点线表示敏感性参数 δ 等于真实 δ

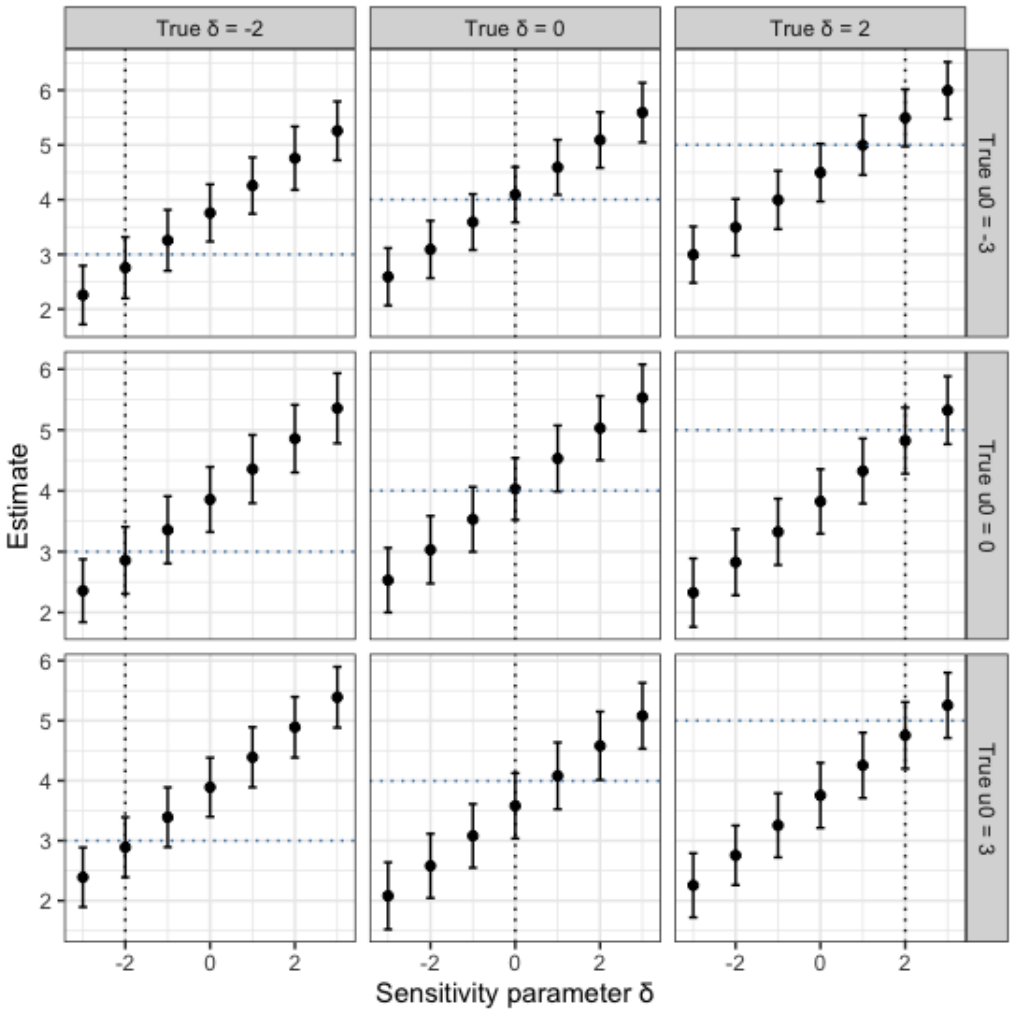

图 A.3：在不同真实偏倚和敏感性参数取值下，经敏感性参数调整的平均处理效应估计与真实总体平均处理效应的对比；每项研究的样本量为 n=500，95%置信区间由 1000 个自助抽样样本构建。当敏感性参数为 $\delta = 0$ 时，调整后估计对应于未调整估计值。水平点线表示给定真实 δ 时的真实总体平均处理效应；垂直点线表示敏感性参数 δ 等于真实 δ

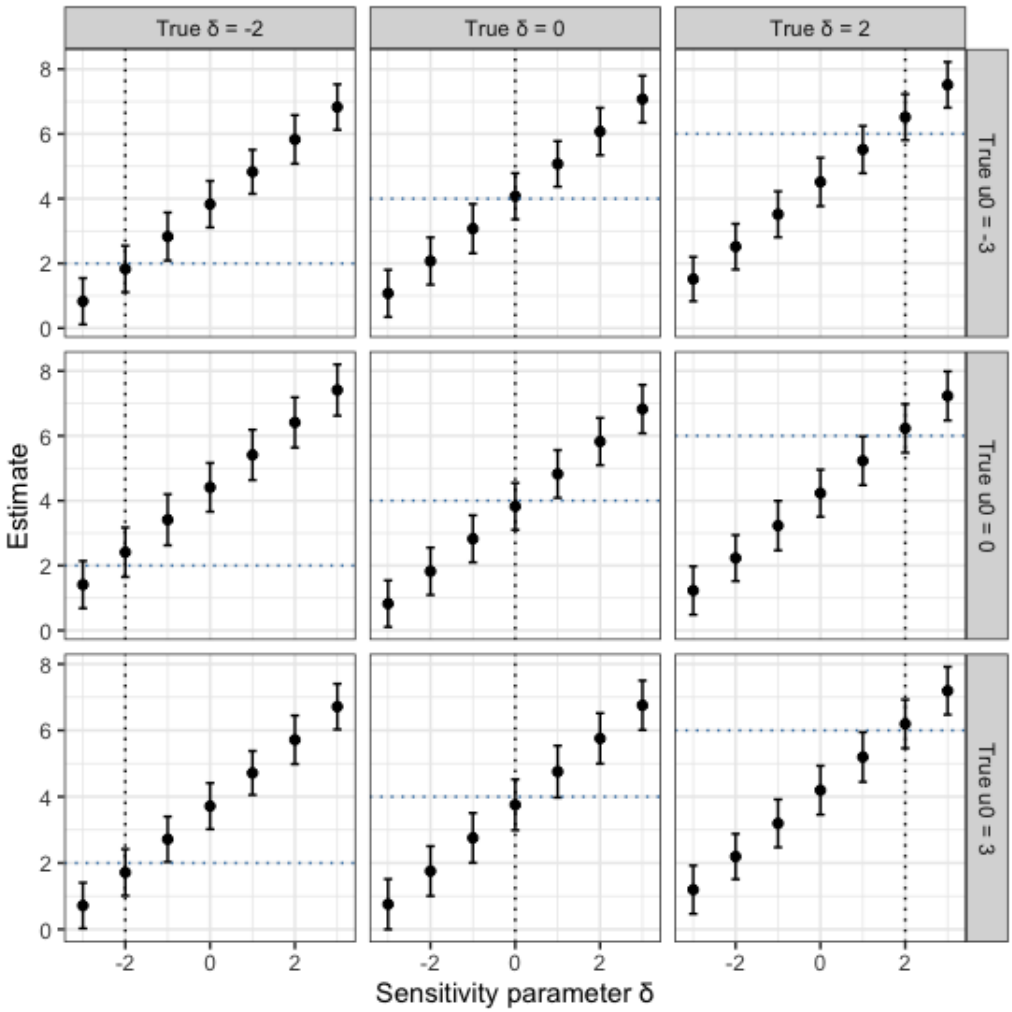

图 A.4：在不同真实偏倚和敏感性参数取值下，经敏感性参数调整的平均处理效应估计与结局缺失研究中的真实平均处理效应的对比；每项研究的 n=500 ，95%置信区间根据 1000 个自助抽样样本构建。当敏感性参数 $\delta = 0$时，调整后估计对应于未调整估计值。水平点线表示给定真实 δ 时研究特异性平均处理效应的真实值；垂直点线表示敏感性参数 δ 等于真实 δ

B 违反假设时敏感性分析公式的推导4

$$
\begin{aligned}
\mathrm{ATE} = & \sum_{s=1}^{s^*} \pi_s E\{E(Y \mid W, A=1, S=s) - E(Y \mid W, A=0, S=s) \mid S=s\} \\
& + \sum_{s=s^*+1}^{S} \pi_s E\left[E\left\{E\left(Y \mid \mathcal{T}_s, W, A=1, S=s\right) \mid W, A=1, S=s\right\}\right. \\
& \qquad \left. - E\left\{E\left(Y \mid \mathcal{T}_s, W, A=0, S=s\right) \mid W, A=0, S=s\right\} \mid S=s\right] \\
= & \sum_{s=1}^{s^*} \pi_s E\{E(Y \mid W, A=1, S=s) - E(Y \mid W, A=0, S=s) \mid S=s\} \\
& + \sum_{s=s^*+1}^{\mathcal{S}} \pi_s E\left[E\left\{E\left(Y \mid \mathcal{T}_s, W, A=1, S \in \sigma_s\right) + u\left(A=1, \mathcal{T}_s, W\right) \mid W, A=1, S=s\right\}\right. \\
& \qquad \left. - E\left\{E\left(Y \mid \mathcal{T}_s, W, A=0, S \in \sigma_s\right) + u\left(A=0, \mathcal{T}_s, W\right) \mid W, A=0, S=s\right\} \mid S=s\right] \\
= & \sum_{s=1}^{s^*} \pi_s E\{E(Y \mid W, A=1, S=s) - E(Y \mid W, A=0, S=s) \mid S=s) \\
& + \sum_{s=s^*+1}^{\mathcal{S}} \pi_s E\left[E\left\{E\left(Y \mid \mathcal{T}_s, W, A=1, S \in \sigma_s\right) \mid W, A=1, S=s)\right\}\right. \\
& \qquad \left. - E\left\{E\left(Y \mid \mathcal{T}_s, W, A=0, S \in \sigma_s\right) \mid W, A=0, S=s)\right\} \mid S=s\right] \\
& + \sum_{s=s^*+1}^{\mathcal{S}} \pi_s E[E\left\{u\left(A=1, \mathcal{T}_s, W\right) \mid W, A=1, S=s\right\} \\
& \qquad - E\left\{u\left(A=0, \mathcal{T}_s, W\right) \mid W, A=0, S=s\right\} \mid S=s]
\end{aligned}
$$