Cornfield 型敏感性分析的贝叶斯重新诠释：从阈值到概率

Tommaso Costa [1,2,3,*]

[1] GCS-fMRI、Koelliker 医院与都灵大学心理学系，10124 意大利都灵；xmarferr@gmail.com[2] FOCUS 实验室 都灵大学心理学系，10124 意大利都灵[3] 都灵神经科学研究所（NIT），10124 意大利都灵 * 通讯作者：tommaso.costa@unito.it

摘要

观察性研究中针对未测量混杂的敏感性分析通常基于阈值量，例如 Cornfield 条件或 E 值，用以量化一个混杂因素必须有多强，才能解释掉观察到的关联。然而，这些方法并未回答一个根本的推断问题：这种混杂因素存在的可能性有多大?

在本研究中，我们提出了 Cornfield 型敏感性分析的贝叶斯重新表述，将未测量混杂的强度视为随机变量。在这一框架下，E值被重新解释为一个阈值，而核心推断量则变为混杂超过该阈值的后验概率。这使敏感性分析从描述性诊断转变为对结果稳健性的概率评估。

我们构建了一个简单的生成模型，将观测到的效应估计值与真实因果效应及混杂偏倚联系起来，并指定了反映合理混杂机制的先验分布。由此得到的框架能够产生对证据脆弱性的后验度量，这些度量易于解释，并适用于汇总层面数据。

基于实证案例研究的示例表明，所提出的方法在保留 E值可解释性的同时，能够更加细致且与决策相关地刻画稳健性。更广泛而言，该框架使敏感性分析与不确定性下推断的贝叶斯原则相一致，为纯粹基于阈值的推理提供了一种连贯的替代方案。

引言

观察性研究中的因果推断不可避免地与未测量混杂问题交织在一起。即使对观察到的协变量进行了广泛调整，某个未观察到的因素可能解释观察到的关联这一可能性，仍然在结构上无法得到解决。这不仅仅是技术上的局限，更是一个根本性的认识论问题。最终受到质疑的，不仅是估计效应的大小，还有因果主张本身的可信度。在实践中，研究人员常常在两种都不令人满意的立场之间摇摆：将调整后的关联视为近似因果关系，或因潜在隐性偏倚的存在而认为这些关联本质上不可靠。

在这一背景下，敏感性分析应运而生，旨在以有原则的方式刻画而非消除这种不确定性。Cornfield及其同事（Cornfield等人，1959年）提出的论证，是这一方向上的奠基性贡献，最初形成于吸烟与肺癌之争的背景下。他们的洞见看似简单：只有当某个未测量混杂因素同时与暴露和结局具有足够强的关联时，观察到的关联才可能被其完全解释。这一推理将问题从混杂的单纯可能性，转变为对混杂必须达到何种强度、才能使观测效应失效的定量评估。

这一推理路径通过引入E值（VanderWeele和Ding，2017年）在现代流行病学中得到了形式化。E值以一个标量概括：某个未测量混杂因素必须分别与暴露和结局具有多强的关联，才能使观察到的效应完全减弱。它的吸引力在于简洁、易于解释，并且适用于汇总层面数据。因此，它迅速成为应用研究中的标准工具。

近期进展已将敏感性分析拓展至新的计算领域。尤其是，近期一项研究探讨了如何运用大语言模型，将 Cornfield 类型的推理付诸实践，并在一系列案例研究中计算 E值，同时公开整理后的数据集，以供方法学研究重复使用（Xiang 等人，2026 年）。这项研究的意义不仅在于其方法上的创新，也在于它为评估新的敏感性分析方法提供了标准化的实证基础。

尽管取得了这些进展，一个根本性的局限依然存在。Cornfield论证和 E值都给出了阈值：它们指出，要解释掉某种关联，混杂因素必须有多强。然而，它们并未回答一个更为重要的问题：这种混杂因素存在的可能性有多大？在应用场景中，这一缺口通常由非正式推理来弥补。研究者可能会将某个 E值判断为“较大”或“较小”，但此类判断本质上取决于具体情境，而且往往缺乏正式的概率论基础。因此，敏感性分析仍主要停留在描述层面，而非推断层面。

已有多项方法学研究尝试解决这一局限。概率偏倚分析（Greenland，1996年、2005年；Lash等人，2009年）提出了为偏倚参数指定分布，并通过模拟传播不确定性的思路。贝叶斯方法更进一步，在层级框架中明确对未测量混杂因素进行建模（McCandless等人，2007年；Gustafson，2010年）。尽管这些方法提供了一种连贯的概率解释，但通常需要较强的建模假设，并且必须掌握详细的数据结构，因此在许多真实情境中的适用性受到限制。

Costa（2025）提出了一种互补视角，引入了一个基于模拟的框架，用于量化足以

解释观察到的关联的未测量混杂因素可能存在的概率。通过将 Cornfield 型阈值与概率推理联系起来，这种方法是迈向对敏感性分析作出更具推断性的解读的重要一步。然而，它仍以蒙特卡洛估计为基础，且未提供明确完整的混杂强度生成模型。

本研究对这一研究方向作出自然延伸：对 Cornfield 型敏感性分析及其对应的 E值进行完整的贝叶斯重新表述。我们不再询问混杂因素必须有多强，而是将混杂强度视为随机变量，并通过先验分布对不确定性进行建模。在这一框架下，E值成为一项边界条件，而核心推断量则是混杂超过这一阈值的后验概率。

这一转变具有重要的概念意义。它将敏感性分析从基于阈值的诊断方法转变为连贯的推断程序。同时，它也使分析与贝叶斯解释相契合，即将概率视为不确定情形下理性信念的度量，从而能够明确纳入而非隐含体现关于合理混杂机制的先验科学知识。

本研究有三个目标。首先，我们在统一偏倚模型中形式化表述 Cornfield 不等式与 E值之间的关系。其次，我们引入贝叶斯框架，将混杂强度视为随机变量，并推导观察到的关联的稳健性后验度量。第三，我们通过实证案例研究展示这一方法，着重说明对混杂的概率评估如何在不同推断机制下带来定性上不同的解释。

更广泛地说，本研究旨在弥合文献中长期存在的缺口。现有方法虽然阐明了需要多大程度的混杂才能推翻某项关联，却没有提供一种有原则的方法来评估这种混杂发生的可能性。通过从阈值转向概率、从描述性诊断转向连贯的推断，所提出的框架旨在推动观察性研究中因果推理的方法与实践发展。

现有方法虽然量化了解释掉某项关联所需的混杂强度，却没有回答一个推断层面的问题：这种混杂有多大可能性。本研究将敏感性分析重新界定为对这种可能性的概率评估。

Data

本研究的实证部分基于一个结构化数据集，其中收录了观察性效应估计。这些数据由近期一项关于因果推断中敏感性分析的研究公开发布（Xiang 等人，2026 年）。数据取自该研究相关的开放获取代码库（https://github.com/qingyan16/LLMs-for-sensitivity-analysis），确保本文所述分析完全透明且可复现。

数据集包含流行病学和临床研究中常见的汇总层面指标，包括效应量——通常以相对风险、优势比或风险比表示——以及在可获得时与之相关的不确定性。这样的选择既出于实践考虑，也有方法论上的依据。在许多应用场景中，无法获取个体层面的数据，而汇总统计量则是常规报告的内容。因此，在这一层面开展工作的框架符合实践中进行因果推断时经常面临的限制。

数据集中的每项关联都附有已报告的E值，或足以重新计算E值的信息。若未明确提供E值，则使用VanderWeele和Ding（2017年）提出的公式进行计算，从而确保所有观测结果的一致性。这一步使每个实证案例都能映射到由Cornfield 型推理界定的共同推断尺度上，在这一尺度上，可以量化足以解释掉某一效应所需的未测量混杂强度。

我们从该数据集中选取了一组具有代表性的案例研究，以说明所提出的贝叶斯框架。选取时考虑了覆盖性质不同的推断机制。具体而言，我们重点关注效应量相对较大的关联；要解释掉这类观察到的关系，需要存在强烈的未测量混杂。同时，我们也关注效应量较为温和的关联，在这类情况下，即使是弱混杂也可能足以解释掉观察到的关系。这种对比对于评估所提出的模型如何将观察证据转化为对稳健性的概率评估至关重要。

出于建模目的，我们将所有效应指标转换到对数尺度上，这为乘法模型提供了自然的参数化方式，也便于指定先验分布。在有置信区间的情况下，我们根据标准近似方法推导出标准误。对于未明确报告不确定性指标的情况，我们根据类似研究中通常观察到的数值，采用了保守的近似方法。尽管这会引入一定程度的近似，但不会影响本分析的概念性目标，即探究关于未测量混杂的不确定性如何在推断框架中传递。

需要强调的是，实证部分的目的并非重新评估所用数据来源研究的实质性结论，而是将该数据集用作方法学测试平台。重点在于，观测效应量与关于未测量混杂的假设相结合后，如何在所提出的贝叶斯框架下转化为后验概率。

所有分析均基于以下假设：在给定已测量协变量的条件下，报告的效应估计值近似无偏；任何残余偏倚均可通过一个表征未测量混杂强度的单一参数来表示。尽管这一假设不可避免地有所简化，但它反映了原始 Cornfield论证（Cornfield等人，1959年）及其通过 E 值（ VanderWeele 和 Ding，2017 年）形成的现代表述所依据的结构，从而能够与现有方法进行直接且一致的比较。

方法

概念框架

所提出的方法在完全概率化的框架下，重新构建了针对未测量混杂的敏感性分析。我们不将混杂视为未知但固定的量，而是将其强度建模为一个随机变量，并通过先验分布量化不确定性。

令 RR_{obs} 表示在调整已测量协变量后，以风险比尺度表示的暴露与结局之间的观察到的关联。我们引入参数 $\Gamma \geq 1$,

该参数表示未测量混杂因素引起的乘性偏倚。在这一表示下，真实因果效应可写为：

$$RR_{true} = \frac{RR_{obs}}{\Gamma}$$

其中 Γ的值越大，表示混杂越强。

参数 Γ 可解释为未测量混杂因素对暴露和结局的综合乘法效应，这与标准的敏感性分析公式一致。

在此框架中，经典的 Cornfield 型推理定义了一个阈值 Γ^*，使得 $\Gamma \geq \Gamma^*$的值足以完全解释观察到的关联。E值可直接估计这一阈值。

似然设定

为引入概率结构，我们在对数尺度上进行分析。令：

$$\theta_{obs} = \log(RR_{obs})$$

取对数后，乘法分解转化为加法结构，从而可以将混杂效应建模为对数尺度上的偏移。我们假设：

$$\theta_{obs} \sim \mathcal{N}(\theta_{true} + \log\ \Gamma, s^2)$$

其中：

$$\theta_{true} = \log(RR_{true})$$

且 s 是观测估计值的标准误

该表述反映了这样一种观点：观察到的关联因混杂偏倚项 $\log\ \Gamma$ 的影响而偏离真实的因果效应。

需要注意的是，仅凭数据无法识别该模型，因为真实效应和混杂偏倚无法分别估计。因此，推断必然受到关于未测量混杂程度的先验假设驱动。

先验分布

我们 spe 为真实因果效应和混杂强度分别指定先验分布 .

对于因果效应：

$$\theta_{true} \sim \mathcal{N}(0, \sigma_\theta^2)$$

其中 这反映了这样一种假设：较大的效应是可能的，但发生的可能性并不高 .

对于混杂强度：

$$\log\ \Gamma \sim \text{Half-Normal}(0, \sigma_\Gamma^2)$$

确保 $\Gamma \geq 1$。尺度参数 σ_Γ 控制着对强混杂可能性的先验信念。这一表述体现了这样的理念：弱混杂很常见，而强混杂则需要各种因素越来越不可信的恰好一致。较小的取值表示对较大混杂效应持怀疑态度，而较大的取值则允许更极端的情形。

例如，$\sigma_\Gamma = 0.5$ 意味着，混杂风险比大于约 2–3 的情况在先验上越来越不可能。

与 E值的联系

E值定义了一个确定性阈值 Γ*，对应于解释掉观察到的关联所需的最小混杂强度。在这一表述中，E值对应于满足以下条件的阈值 Γ^*：

$$\Gamma^* = E(RR_{obs})$$

我们不单独报告这一阈值，而是评估后验概率：

$$P(\Gamma \geq \Gamma^* \mid \theta_{obs})$$

该概率量化了未测量混杂强到足以解释掉观察到的关联这一情况的可信程度。

后验推断

后验推断通过贝叶斯定理得到：

$$p(\theta_{true}, \Gamma \mid \theta_{obs}) \propto p(\theta_{obs} \mid \theta_{true}, \Gamma)\, p(\theta_{true})\, p(\Gamma)$$

基于这一联合后验分布，我们推导出真实效应的后验分布、混杂强度的后验分布以及概率 $P(\Gamma \geq \Gamma^*)$。这些量为评估结果对未测量混杂的稳健性提供了连贯的概率框架。特别是，概率 $P(\Gamma \geq \Gamma^*)$ 直接以概率形式表述了观察到的关联的稳健性。

与先前表述的关系

这种表述等同于基于乘性偏倚因子的偏倚表示法，即将观测效应分解为真实效应和混杂部分。具体而言，参数 Γ 的作用与 Cornfield 型敏感性分析中使用的传统偏倚因子相同，但在这里，它被视为完全贝叶斯框架中的随机变量。

结果

我们将所提出的敏感性分析贝叶斯重新表述应用于来自四个观察性领域的 11 个暴露–结局关联：吸烟、背痛、阿尔茨海默病和环境卫生。对于每个案例，我们计算了未测量混杂强到足以解释观察到的关联的概率，即 $P(\Gamma \geq \Gamma^*)$，其中 Γ^*由 E值给出。结果是在对混杂强度采用中等先验的条件下呈现的。

E值的大小与相应的充分混杂概率之间呈现出清晰的单调关系。如图 1 所示，E值越大，相应概率就显著越低，这表明在假定的先验下，强混杂变得越来越不可能。反之，E值较小时，相应概率则明显较高，反映出观察到的关联更容易受到未测量混杂的影响。

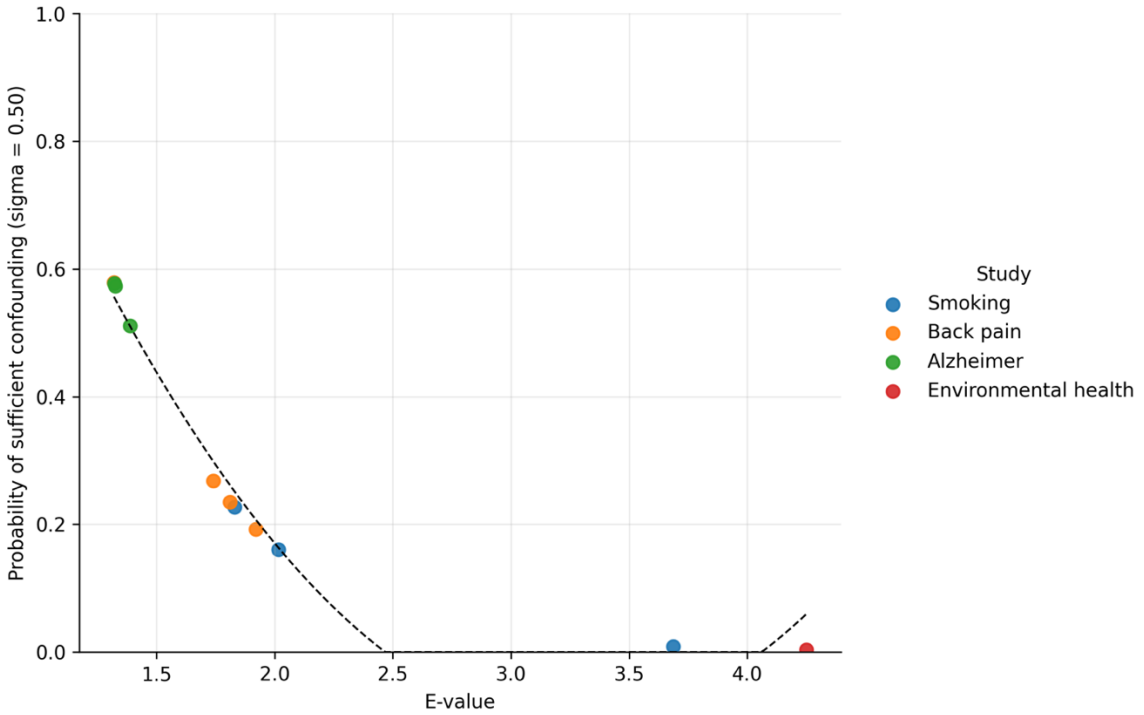


图 1.在中等先验（σ = 0.5）下，E值与未测量混杂强到足以解释掉观察到的关联这一后验概率之间的关系。每个点代表一个案例研究，颜色表示不同的研究领域。虚线表示非参数平滑趋势。

在个别案例研究层面，结果呈现出明显的稳健性梯度。E值最大（$E = 4.25$）的环境健康研究显示，未测量混杂完全解释观察到的关联的概率极低（$P \approx 0.004$）。同样，吸烟研究中最强的关联（$E = 3.686$）对应的概率也较低（$P \approx 0.03$），表明该关联对未测量混杂具有较强的稳健性。

相比之下，E值较小的关联表现出明显不同的特征。阿尔茨海默病案例研究的E值范围为1.32–1.39，其概率约为0.5，这表明在假定的先验下，强度足够的混杂不仅有可能存在，而且相对可信。背痛研究也观察到类似但程度较轻的模式，其中概率约为0.15至0.58，具体取决于效应大小。

在吸烟研究中，考虑较弱的暴露对比时可观察到中间结果。例如，母亲吸烟和家庭吸烟对应的概率约为0.16–0.20，这表明虽然强混杂不太可能，但仍不能有把握地排除。这些案例凸显了所提出的框架能够在单项研究内部进行区分，而不是给出一概而论的定性判断。

从研究层面进行总结，进一步印证了这些模式。环境健康研究表现出很强的稳健性，其次是稳健性中等的吸烟研究。背痛研究处于中间位置，而阿尔茨海默病研究似乎对未测量混杂相当敏感

。这一排序与 E 值所暗示的顺序一致，但提供了更易于理解的概率尺度。

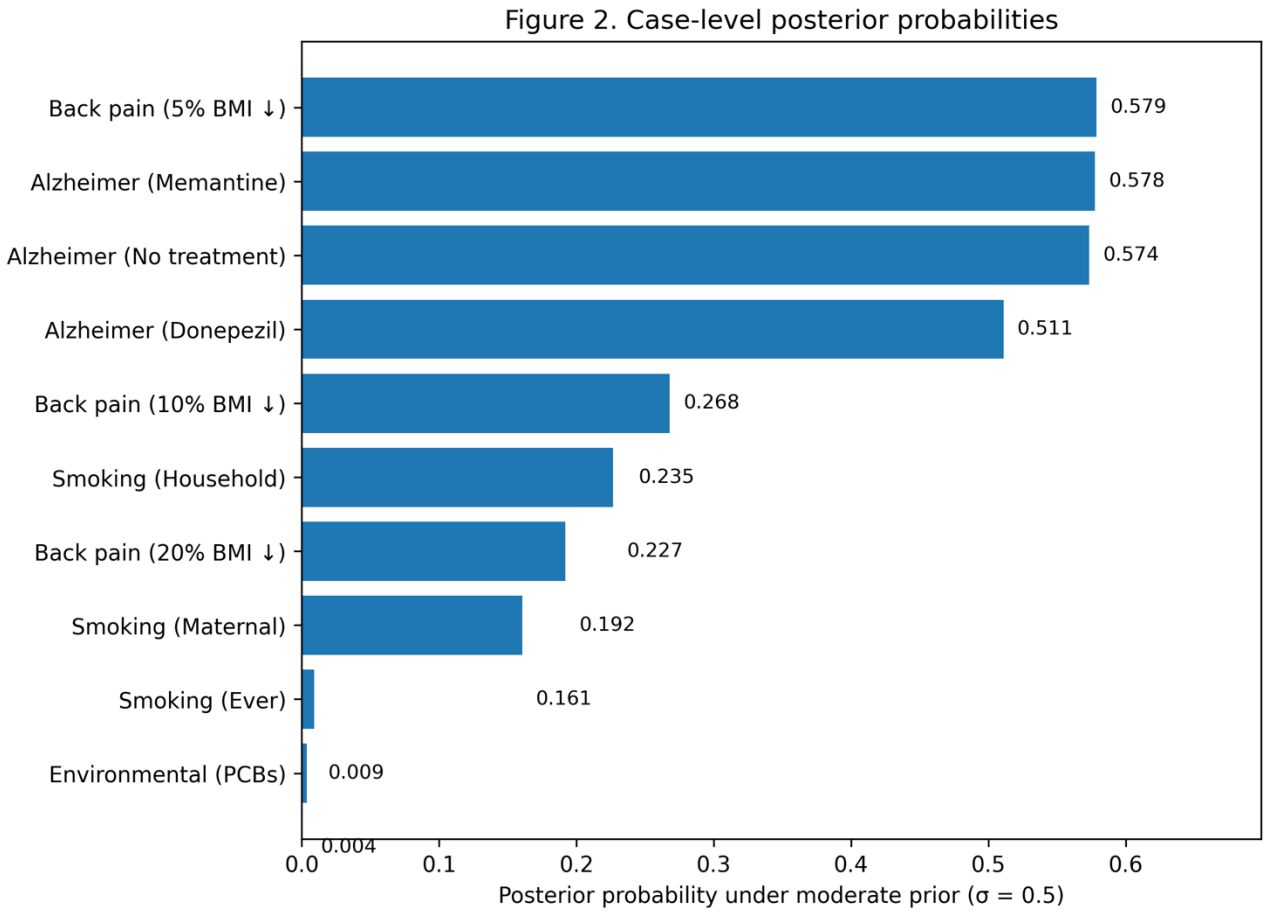


图2. 在混杂强度的中等先验（$\sigma = 0.5$）下，各案例中未测量混杂超过E值阈值的后验概率。案例按后验概率从高到低排列。为便于阅读，图中使用了缩写标签，同时保留原有的研究领域和暴露对比。该图凸显了领域内部及领域之间存在显著异质性：有些关联仍极易受到合理混杂的影响，而另一些则显得稳健得多。

图2进一步展示了各个案例的概率分布，清晰呈现了研究内部及研究之间的异质性。尤其是，高E值与低E值情景之间的差异在图中一目了然，这支持了这样一种解读：E值在连续的概率图景中充当边界条件。

最后，通过改变控制混杂强度分布的尺度参数，考察了对先验假设的敏感性。如图3所示，对于E值较大的情形，在广泛的先验范围内，定性结论保持稳定；而对于E值较小的情形，结论则更为敏感。这种不对称性体现了该框架的一个关键特征：强观测效应在广泛的假设类别下仍然稳健，而弱效应则天然更依赖于关于混杂的先验信念。

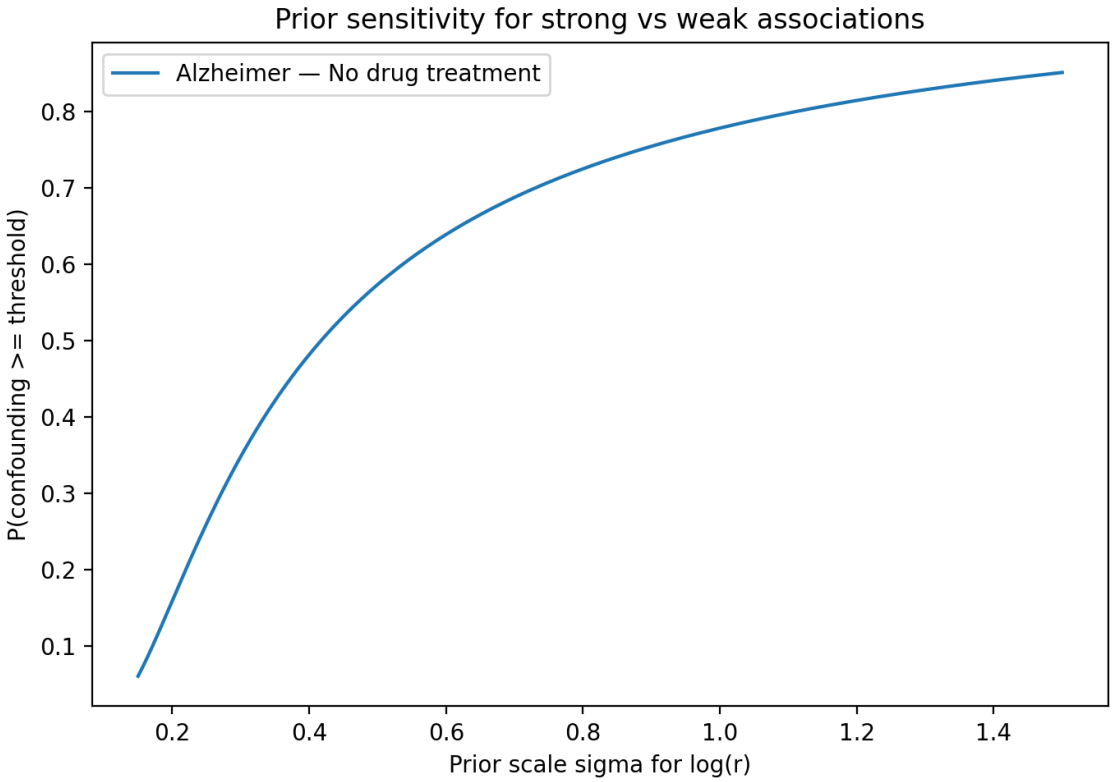


图3。 后验概率对混杂强度先验假设的敏感性。该图比较了不同先验尺度（σ = 0.25、0.5、1.0）下选取的案例研究，表明E值较大时结果稳定，而E值较小时则对先验设定更敏感。

综上，这些结果表明，所提出的贝叶斯框架在将E值扩展到完全概率化领域的同时，保留了其可解释性。该方法不再依据启发式阈值将关联归类为稳健或不稳健，而是直接量化未测量混杂解释掉观测效应的可能性。

重要的是，E值与后验概率之间的关系并非单纯由机械规则决定，而是取决于对混杂强度所假定的先验分布，这凸显了先验知识在敏感性分析中的作用。

这表明，E值本身并不能作为稳健性的推断性度量，而所提出的概率则明确量化了未测量混杂所带来的不确定性。

讨论

本研究提出了一种概念和方法上的转变，旨在改变未测量混杂敏感性分析的实施方式。我们不再将稳健性视为一种基于阈值的

属性——即根据假设的混杂因素原则上能否解释掉观察到的关联来界定——而是用概率术语重新表述这一问题。具体而言，我们量化混杂强度足以超过 E值阈值的概率，从而将确定性诊断转变为连贯的推断性陈述。

这一转变回应了现有方法中一个公认的局限。Cornfield论证（Cornfield等人，1959年）及其通过E值实现的现代版本（VanderWeele和Ding，2017年），为使关联失效所需的混杂强度提供了清晰且易于解释的基准。然而，它们并未回答实践中往往最为关键的问题：这种程度的混杂存在的可能性有多大？在应用场景中，这个问题通常依靠非正式推理来解决，依据的是领域专业知识，而缺少正式的概率结构。

所提出的概率可被解释为一种与决策相关的证据脆弱性度量，用以量化未测量混杂足以改变因果结论的可能性。通过为混杂强度引入先验分布，本框架使这一评估变得明确且具有形式化依据。由此得到的量 $P(\Gamma \geq \Gamma^{*})$，直接以概率形式表述了观察到的关联有多稳健。这种解释与推断的贝叶斯观点自然契合：不确定性以信念程度而非长期频率来表达，从而将敏感性分析从描述性诊断扩展为连贯的推断工具。

实证结果展示了这一重新表述在实践中的意义。尽管各项研究的稳健性排序总体上与E值所暗示的排序一致，但概率尺度提供了更细致的解读。具体而言，该框架不仅区分不同研究，还能揭示同一项研究内部的差异，从而展现被单一的定性摘要所掩盖的异质性。此外，结果凸显了一种重要的不对称性：较大的E值对应着在广泛的先验假设下，混杂足以产生影响的概率始终较低；而较小的E值则会使结论对先验设定更加敏感。这并非该方法的局限，而是弱观测效应所固有不确定性的体现。

所提出的框架与此前将敏感性分析概率化的研究密切相关。概率偏倚分析（Greenland，1996年、2005年；Lash等人，2009年）和针对未测量混杂的贝叶斯模型（McCandless等人，2007年；Gustafson，2010年）已经纳入了偏倚参数的不确定性。然而，这些方法通常需要详细的建模假设，并且在只能获得汇总数据时往往难以应用。相比之下，本方法直接基于E值等量开展分析，因此适用于只能获取聚合信息的情形。

该方法也拓展了 Costa（2025）的近期研究。该研究从基于模拟的视角探讨了足以解释观察到的关联的混杂是否合理。尽管该框架初步搭建了基于阈值的推理与概率解释之间的桥梁，但它依赖蒙特卡洛估计，且没有明确的生成模型。本研究在贝叶斯框架内将这一思路形式化，使先验假设能够得到清晰界定并接受系统探索。

需要承认若干局限性。首先，该框架依赖于对混杂强度先验分布的设定，这不可避免地会引入主观性。然而，这种主观性本身并非缺点；相反，它反映了背景

知识在评估未测量混杂的合理性时所发挥的不可避免的作用。重要的是，针对先验的敏感性分析能够使这种依赖关系变得透明。

其次，该模型通过单一参数对混杂进行简化表述。尽管这种抽象有助于解释和计算，但现实中的混杂结构可能更为复杂，涉及多个变量和交互作用。将该框架扩展为对混杂进行更丰富的表述，是未来研究的自然方向。

第三，实证分析基于数量有限的案例研究。尽管这些示例涵盖多个领域，并展示了该方法的关键特征，但若能在更大规模的观察性研究语料中开展更广泛的评估，将有助于考察其普适性。

更广泛而言，该框架使敏感性分析与不确定性下推断的贝叶斯原则相一致，以概率陈述取代基于阈值的推理，从而能够直接为科学与临床判断提供依据。

尽管存在这些局限，所提出的框架仍为现有方法提供了概念上连贯且实用的扩展。它保留了E值的可解释性，同时解决了其核心局限：缺乏概率解释。

结论

本研究提出了一种针对未测量混杂的敏感性分析贝叶斯重新表述，其中将混杂的强度视为随机变量，而非未知常数。通过将E值与概率度量$P(\Gamma \geq \Gamma^*)$联系起来，该方法直接量化了未测量混杂解释掉观察到的关联的可能性。

这一视角将敏感性分析从基于阈值的诊断转变为完整的推断程序。研究者不再只是询问是否可能存在强度足够大的混杂因素，而是可以在明确假设的基础上评估此类混杂因素出现的可能性。由此，该方法弥合了方法论工具与应用研究中产生的解释性问题之间的关键鸿沟。

更广泛地说，这一框架推动因果推断的逻辑从二元判断转向概率推理。通过明确假设并量化其影响，它使敏感性分析与一套连贯的不确定性下推断理论相契合。这种契合在观察性研究中尤为重要，因为未测量混杂无法消除，只能通过推理加以考量。

未来的研究可以从多个方向拓展这一方法，包括开发用于处理混杂的层级模型、将其与个体层面的贝叶斯分析相结合，以及应用于大规模证据综合。归根结底，目标不是消除不确定性，而是以一种支持透明且理性的科学判断的形式来呈现不确定性。

参考文献

Cinelli, C.，& Hazlett, C.（2020）。理解敏感性：扩展遗漏变量偏差。英国皇家统计学会杂志：*B*辑，82(1)，39–67。https://doi.org/10.1111/rssb.12348 Cornfield, J.，Haenszel, W.，Hammond, E. C.，Lilienfeld, A. M.，Shimkin, M. B.，& Wynder, E. L.（1959年）。吸烟与肺癌：近期证据及对若干问题的讨论。美国国家癌症研究所杂志，*22*（1），173–203。Costa, T.（2025）。不可能的混杂因素：量化新冠疫苗有效性替代性解释的界限。*PLOS ONE*，*20*（12），e0336063。https://doi.org/10.1371/journal.pone.0336063Greenland, S.（1996年）。偏倚敏感性分析的基本方法。国际流行病学杂志，*25*（6），1107–1116。Greenland, S.（2005年）。用于观察性数据分析的多重偏倚建模。英国皇家统计学会杂志：*A*辑，168(2)，267–306。Gustafson, P.（2010年）。部分识别模型的贝叶斯推断。国际生物统计学杂志，*6*（2）。

Lash, T. L., Fox, M. P., & Fink, A. K. (2009年). 将定量偏倚分析应用于流行病学数据. 施普林格。

McCandless, L. C., Gustafson, P., & Levy, A. R. (2007年). 观察性研究中未测量混杂的贝叶斯敏感性分析. 医学统计学, *26*(11), 2331–2347.

Rosenbaum, P. R. (2002). 观察性研究（第 2 版）。施普林格。Rosenbaum, P. R., & Rubin, D. B. (1983). 倾向评分在观察性研究中的核心作用。 *Biometrika, 70*(1), 41–55. VanderWeele, T. J., & Ding, P. (2017年). 观察性研究中的敏感性分析：引入 E值。 《内科学年鉴》, *167*(4), 268–274. Xiang, Q., Zhang, J., & Feng, B. (2026 年). 在因果推断中使用大语言模型进行敏感性分析：*Cornfield* 不等式和 *E*值的案例研究. arXiv. https://arxiv.org/abs/2603.14273