Probabilities of Causation: Bounds and Identification

Jin Tian and Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles, CA 90024
{*jtian, judea* }*@cs.ucla.edu*

Abstract

This paper deals with the problem of estimating the probability that one event was a cause of another in a given scenario. Using structural-semantical definitions of the probabilities of necessary or sufficient causation (or both), we show how to optimally bound these quantities from data obtained in experimental and observational studies, making minimal assumptions concerning the data-generating process. In particular, we strengthen the results of Pearl (1999) by weakening the data-generation assumptions and deriving theoretically sharp bounds on the probabilities of causation. These results delineate precisely how empirical data can be used both in settling questions of attribution and in solving attribution-related problems of decision making.

1 Introduction

Assessing the likelihood that one event ***was the cause*** of another guides much of what we understand about (and how we act in) the world. For example, few of us would take aspirin to combat headache if it were not for our conviction that, with high probability, it was aspirin that "actually caused" relief in previous headache episodes. [Pearl, 1999] gave counterfactual definitions for the probabilities of ***necessary*** or ***sufficient*** causation (or both) based on structural model semantics, which defines counterfactuals as quantities derived from modifiable sets of functions [Galles and Pearl, 1997, Galles and Pearl, 1998, Halpern, 1998, Pearl, 2000, chapter 7].

The central aim of this paper is to estimate probabilities of causation from frequency data, as obtained in experimental and observational statistical studies. In general, such probabilities are ***non-identifiable***, that is, non-estimable from frequency data alone. One factor that hinders identifiability is confounding – the cause and the effect may both be influenced by a third factor. Moreover, even in the absence of confounding, probabilities of causation are sensitive to the data-generating process, namely, the functional relationships that connect causes and effects [Robins and Greenland, 1989, Balke and Pearl, 1994]. Nonetheless, useful information in the form of ***bounds*** on the probabilities of causation can be extracted from empirical data without actually knowing the data-generating process. We show that these bounds improve when data from observational and experimental studies are combined. Additionally, under certain assumptions about the data-generating process (such as exogeneity and monotonicity), the bounds may collapse to point estimates, which means that the probabilities of causation are ***identifiable*** – they can be expressed in terms of probabilities of observed quantities. These estimates often appear in the literature as measures of ***attribution***, and our analysis thus explicates the assumptions that must be ascertained before those measures can legitimately be interpreted as probabilities of causation.

The analysis of this paper extends the results reported in [Pearl, 1999] [Pearl, 2000, pp. 283-308]. Pearl derived bounds and identification conditions under certain assumptions of exogeneity and monotonicity, and this paper narrows his bounds and weakens his assumptions. In particular, we show that for most of Pearl's results, the assumption of strong exogeneity can be replaced by weak exogeneity (to be defined in Section 3.3). Additionally, we show that the point estimates that Pearl obtained under the assumption of monotonicity (Definition 6) constitute valid lower bounds when monotonicity is not assumed. Finally, we prove that the bounds derived by Pearl, as well as those provided in this paper are ***sharp***, that is, they cannot be improved without strengthening the assumptions. We illustrate the use of our results in the context of legal disputes (Section 4) and personal

因果概率：界限与识别

金天和朱迪亚·佩尔
认知系统实验室 计算机科学系 加州大学
洛杉矶分校，加利福尼亚州 90024

{tian, judea } @cs. ucla. edu

摘要

本文讨论了在给定场景中估计一个事件是否是另一个事件的因果的概率问题。使用结构语义定义的必要或充分因果的概率（或两者兼而有之），我们展示了如何从实验和观察研究中获得的数据中最优地界定这些量，对数据生成过程做出最小假设。特别是，我们通过放宽数据生成假设并推导出理论上精确的因果概率界限，加强了 Pearl（1999）的结果。这些结果精确地说明了如何使用经验数据来解决归因问题以及解决与归因相关的决策问题。

1 引言

评估一个事件是否是另一个事件的因果概率，引导我们理解世界（以及我们如何行动）的许多方面。例如，如果我们没有确信，在先前的头痛发作中，阿司匹林“实际上导致了”缓解，那么我们中很少有人会服用阿司匹林来对抗头痛。[Pearl，1999]基于结构模型语义给出了必要或充分因果的概率（或两者兼而有之）的反事实定义，该语义将反事实定义为从可修改的函数集合中导出的量[Galles 和 Pearl，1997，Galles 和 Pearl，1998，Halpern，1998，Pearl，2000，第 7 章]。

本文的核心目标是估计因果概率，这些概率是从实验和观察性统计研究中获得的频率数据中得到的。一般来说，这些概率是不可识别的也就是说，仅从频率数据中无法估计。阻碍可识别性的一个因素是，原因和效果可能都受到第三个因素的影响。此外，即使在没有混杂的情况下，因果概率对数据生成过程也很敏感，即连接原因和效果的功能关系[Robins 和 Greenland，1989，Balke 和 Pearl，1994]。尽管如此，我们可以在不知道数据生成过程的情况下，从经验数据中提取关于因果概率界限的有用信息。我们表明，当将观察性和实验研究的数据结合起来时，这些界限会得到改善。此外，在关于数据生成过程的某些假设下（如外生性和单调性），界限可能会缩小到点估计，这意味着因果概率是可识别的，它们可以用观察量的概率来表示。这些估计通常在文献中作为归因的度量出现，因此我们的分析解释了在那些度量可以合法地解释为因果概率之前必须确定的前提。

本文的分析扩展了[Pearl，1999] [Pearl，2000，第 283-308 页]中报告的结果。Pearl 在一定的外生性和单调性假设下推导了界限和识别条件，本文缩小了他的界限并放宽了他的假设。特别是，我们表明，对于 Pearl 的大多数结果，强外生性的假设可以被弱外生性（将在第 3.3 节中定义）所取代。此外，我们表明，Pearl 在单调性假设下（定义 6）获得的点估计在单调性不被假设的情况下构成了有效的下界。最后，我们证明，Pearl 推导出的界限以及本文提供的界限都是尖锐的，也就是说，在不加强假设的情况下无法改进。我们以法律纠纷（第 4 节）和个人

decision making (Section 5).

2 Probabilities of Causation: Definitions

In this section, we present the definitions for the three aspects of causation as defined in [Pearl, 1999]. We use the language of counterfactuals in its structural model semantics, as given in Balke and Pearl (1995), Galles and Pearl (1997, 1998), and Halpern (1998). We use $Y_x = y$ to denote the counterfactual sentence "Variable Y would have the value y, had X been x." The structural model interpretation of this sentence reads: "Deleting the equation for X from the model and setting the value of X to a constant x will yield a solution in which variable Y will take on the value y."

One property that the counterfactual relationships satisfy is the *consistency condition* [Robins, 1987]:

$$(X = x) \Rightarrow (Y_x = Y) \tag{1}$$

stating that if we intervene and set the experimental conditions $X = x$ equal to those prevailing before the intervention, we should not expect any change in the response variable Y. This property will be used in several derivations of this section and Section 3. For detailed exposition of the structural account and its applications see [Pearl, 2000, chapter 7]. For notational simplicity, we limit the discussion to binary variables; extension to multi-valued variables are straightforward (see Pearl 2000, p. 286, footnote 5).

Definition 1 (*Probability of necessity* (*PN*))
Let X and Y be two binary variables in a causal model M, let x and y stand for the propositions $X = true$ and $Y = true$, respectively, and x' and y' for their complements. The probability of necessity is defined as the expression

$$\begin{aligned} PN &\triangleq P(Y_{x'} = false \mid X = true, Y = true) \\ &\triangleq P(y'_{x'} | x, y) \end{aligned} \tag{2}$$

In other words, PN stands for the probability that event y would not have occurred in the absence of event x, $y'_{x'}$, given that x and y did in fact occur.

Note that lower case letters (e.g., x, y) stand for propositions (or events). Note also the abbreviations y_x for $Y_x = \text{true}$ and y'_x for $Y_x = \text{false}$. Readers accustomed to writing "$A > B$" for the counterfactual "B if it were A" can translate Eq. (2) to read $PN \triangleq P(x' > y' | x, y)$.

PN has applications in epidemiology, legal reasoning, and artificial intelligence (AI). Epidemiologists have long been concerned with estimating the probability that a certain case of disease is *attributable* to a particular exposure, which is normally interpreted counterfactually as "the probability that disease would not have occurred in the absence of exposure, given that disease and exposure did in fact occur." This counterfactual notion is also used frequently in lawsuits, where legal responsibility is at the center of contention (see Section 4).

Definition 2 (*Probability of sufficiency* (*PS*))

$$PS \triangleq P(y_x | y', x') \tag{3}$$

PS finds applications in policy analysis, AI, and psychology. A policy maker may well be interested in the dangers that a certain exposure may present to the healthy population [Khoury *et al.*, 1989]. Counterfactually, this notion is expressed as the "probability that a healthy unexposed individual would have gotten the disease had he/she been exposed." In psychology, PS serves as the basis for Cheng's (1997) causal power theory [Glymour, 1998], which attempts to explain how humans judge causal strength among events. In AI, PS plays a major role in the generation of explanations [Pearl, 2000, pp. 221-223].

Definition 3 (*Probability of necessity and sufficiency* (*PNS*))

$$PNS \triangleq P(y_x, y'_{x'}) \tag{4}$$

PNS stands for the probability that y would respond to x both ways, and therefore measures both the sufficiency and necessity of x to produce y.

Although none of these quantities is sufficient for determining the others, they are not entirely independent, as shown in the following lemma.

Lemma 1 *The probabilities of causation satisfy the following relationship* [Pearl, 1999] *:*

$$PNS = P(x,y)PN + P(x',y')PS \tag{5}$$

Since all the causal measures defined above invoke conditionalization on y, and since y is presumed affected by x, the antecedent of the counterfactual y_x, we know that none of these quantities is identifiable from knowledge of frequency data alone, even under condition of no confounding. However, useful information in the form of bounds may be derived for

以及本文提供的那些都是尖锐的决策（第 5 节）。

2 因果概率：定义

在本节中，我们介绍了[Pearl，1999]中定义的因果关系的三个方面的定义。我们使用 Balke 和 Pearl（1995）、Galles 和 Pearl（1997、1998）以及 Halpern（1998）给出的结构模型语义中的反事实语言。我们用 Yx = y 表示反事实句子“如果变量 X 是 x，那么变量 Y 将具有值 y。”这个句子的结构模型解释为：“从模型中删除 X 的方程，并将 X 的值设置为常数 x，将得到一个解，其中变量 Y 将取值 y。”

流行病学家长期以来一直关注估计某种疾病病例是否归因于特定的暴露，这通常被反事实地解释为“在疾病和暴露确实发生的情况下，没有暴露疾病就不会发生的概率。”这种反事实观念也经常在诉讼中使用，其中法律责任是争议的核心（见第 4 节）。

定义 2（充分性概率（PS））

PS P(Y|xiY',x') (3)

概率因果律在政策分析、人工智能和心理学中都有应用。政策制定者可能会对某种暴露对健康人群可能带来的危险感兴趣[Khoury 等人，1989 年]。反事实地，这个概念可以表达为“一个未暴露的健康个体如果暴露了会患上疾病的概率。”在心理学中，概率因果律是 Cheng（1997）因果力量理论[Glymour，1998]的基础，该理论试图解释人类如何判断事件之间的因果强度。在人工智能中，概率因果律在解释生成中起着重要作用[Pearl，2000，第 221-223 页]。

定义 3（必要性和充分性概率（PNS））

$$PNS \triangleq P(y_x, y'_{x'}) \tag{4}$$

定义 1（必要性概率（PN））设 X 和 Y 是因果模型 M 中的两个二元变量，x 和 y 分别表示 X 为真和 Y 为真的命题，x'和 y'分别表示它们的补。必要性概率被定义为以下表达式

PN = P(Yx' =f al se I X=true, Y =true)
≜ P(y��lx,y) (2)

PNS 代表 y 对 x 双向响应的概率，因此衡量了 x 对产生 y 的充分性和必要性。

虽然这些量中的任何一个都不足以确定其他量，但它们并不是完全独立的，如下面的引理所示。

因果概率满足以下关系[Pearl，1999]：

PNS P(x,y)PN + P(x',y')PS (5)

换句话说，PN 表示在没有事件 x 的情况下，事件 y 不会发生，给定 x 和 y 实际上已经发生。

注意，小写字母（例如，x，y）代表命题（或事件）。还要注意缩写 Yx 表示 Yx= true，y� 表示 Yx = false。习惯于写作 "A B" 表示 "如果 A 则 B" 的读者可以将公式（2）翻译为 PN=P(x' > y'lx,y)。

概率因果边界在流行病学、法律推理和人工智能（AI）中都有应用

由于上述所有因果度量都涉及对 y 的条件化，并且由于假设受 x 影响，反事实 Yx 的前件，我们知道仅从频率数据的知识中无法识别这些量，即使在无混杂的情况下也是如此。然而，可以推导出有用的信息形式为界限。

these quantities from frequency data, especially when knowledge about *causal effects* $P(y_x)$ and $P(y_{x'})$ is also available[1]. Moreover, under some general assumptions about the data-generating process, these quantities may even be identified.

3 Bounds and Conditions of Identification

In this section we will assume that experimental data will be summarized in the form of the causal effects $P(y_x)$ and $P(y_{x'})$, and nonexperimental data will be summarized in the form of the joint probability function: $P_{XY} = \{P(x,y), P(x',y), P(x,y'), P(x',y')\}$.

3.1 Linear programming formulation

Since every causal model induces a joint probability distribution on the four binary variables: X, Y, Y_x and $Y_{x'}$, specifying the sixteen parameters of this distribution would suffice for computing the PN, PS, and PNS. Moreover, since Y is a deterministic function of the other three variables, the problem is fully specified by the following set of eight parameters:

$$\begin{aligned}
p_{111} &= P(y_x, y_{x'}, x) &&= P(x, y, y_{x'}) \\
p_{110} &= P(y_x, y_{x'}, x') &&= P(x', y, y_x) \\
p_{101} &= P(y_x, y'_{x'}, x) &&= P(x, y, y'_{x'}) \\
p_{100} &= P(y_x, y'_{x'}, x') &&= P(x', y', y_x) \\
p_{011} &= P(y'_x, y_{x'}, x) &&= P(x, y', y_{x'}) \\
p_{010} &= P(y'_x, y_{x'}, x') &&= P(x', y, y'_x) \\
p_{001} &= P(y'_x, y'_{x'}, x) &&= P(x, y', y'_{x'}) \\
p_{000} &= P(y'_x, y'_{x'}, x') &&= P(x', y', y'_x)
\end{aligned}$$

where we have used the consistency condition Eq. (1). These parameters are further constrained by the probabilistic equality

$$\sum_{i=0}^{1}\sum_{j=0}^{1}\sum_{k=0}^{1} p_{ijk} = 1$$
$$p_{ijk} \geq 0 \text{ for } i,j,k \in \{0,1\} \qquad (6)$$

In addition, the nonexperimental probabilities P_{XY} impose the constraints:

$$\begin{aligned}
p_{111} + p_{101} &= P(x,y) \\
p_{011} + p_{001} &= P(x,y') \\
p_{110} + p_{010} &= P(x',y)
\end{aligned} \qquad (7)$$

[1]The causal effects $P(y_x)$ and $P(y_{x'})$ can be estimated reliably from controlled experimental studies, and from certain observational (i.e., nonexperimental) studies which permit the control of confounding through adjustment of covariates [Pearl, 1995].

and the causal effects, $P(y_x)$ and $P(y_{x'})$, impose the constraints:

$$\begin{aligned}
P(y_x) &= p_{111} + p_{110} + p_{101} + p_{100} \\
P(y_{x'}) &= p_{111} + p_{110} + p_{011} + p_{010}
\end{aligned} \qquad (8)$$

The quantities we wish to bound are:

$$PNS = p_{101} + p_{100} \qquad (9)$$
$$PN = p_{101}/P(x,y) \qquad (10)$$
$$PS = p_{100}/P(x',y') \qquad (11)$$

Optimizing the functions in (9)–(11), subject to equality constraints, defines a linear programming (LP) problem that lends itself to closed-form solution. Balke (1995, Appendix B) describes a computer program that takes symbolic descriptions of LP problems and returns symbolic expressions for the desired bounds. The program works by systematically enumerating the vertices of the constraint polygon of the dual problem. The bounds reported in this paper were produced (or tested) using Balke's program, and will be stated here without proofs; their correctness can be verified by manually enumerating the vertices as described in [Balke, 1995, Appendix B]. These bounds are guaranteed to be sharp because the optimization is global.

3.2 Bounds with no assumptions

3.2.1 Given nonexperimental data

Given P_{XY}, constraints (6) and (7) induce the following upper bound on PNS:

$$0 \leq PNS \leq P(x,y) + P(x',y'). \qquad (12)$$

However, PN and PS are not constrained by P_{XY}.

These constraints also induce bounds on the causal effects $P(y_x)$ and $P(y_{x'})$:

$$\begin{aligned}
P(x,y) &\leq P(y_x) \leq 1 - P(x,y') \\
P(x',y) &\leq P(y_{x'}) \leq 1 - P(x',y')
\end{aligned} \qquad (13)$$

3.2.2 Given causal effects

Given constraints (6) and (8), the bounds induced on PNS are:

$$\max[0, P(y_x) - P(y_{x'})] \leq PNS \leq \min[P(y_x), P(y'_{x'})] \qquad (14)$$

with no constraints on PN and PS.

3.2.3 Given both nonexperimental data and causal effects

Given the constraints (6), (7) and (8), the following bounds are induced on the three probabilities of cau-

这些量来自频率数据，尤其是在已知因果效应 P(yx) 和 P(Yx') 的情况下。此外，在关于数据生成过程的一些一般假设下，这些量甚至可以被识别。

3 识别的界限和条件

在本节中，我们将假设实验数据将以因果效应 P(yx) 和 P(Yx') 的形式总结，而非实验数据将以联合概率函数的形式总结：Pxy

= {P(x, y), P(x', y), P(x,y'), P(x', y')}

3.1 线性规划公式

由于每个因果模型都会在四个二元变量 X、Y、Yx 和 Yx'上诱导一个联合概率分布，因此指定这个分布的十六个参数就足以计算 PN、PS 和 PNS。此外，由于 Y 是其他三个变量的确定性函数，因此问题完全由以下八个参数集定义：

Plll - P(x, Y, Yx')
Pllo= - P(x', y, Yx)
P101= - P(x, y, y',)
PlOO= - P(x', y', yx)
Poll= =P(x,y', Yx')
Pow= =P(x', y, y',)
POOl= =P(x, y', y□,)
Pooo - P(x', y', y□)

其中我们使用了一致性条件方程（1）。我们知道，仅从频率数据中无法识别这些量。这些参数进一步受到概率等式的约束

1 1
I: I: L:Pijk=
i=0 j=0 k=0
Pijk

此外，非实验概率 Pxy 施加以下约束：

Plll + P101 P(x,y)
民调+Po0 P(x, y') (7)
Pno+Pow= P(x', y)

1 原因效应 P(yx) 和 P(Yx') 可以从受控实验研究中可靠地估计，也可以从某些允许通过调整协变量控制混杂因素的观察性（即非实验性）研究中估计 [Pearl, 1995]。

并且原因效应，P(yx) 和 P(Yx') 施加以下约束：

P(y_x) = Plll+Pno+P101+Pwo
P(y_x') = P111+Pno+Poll +Pow (8)

我们希望约束的量是：

PNS = PlOl + PlOO (9)
PN = P101 (10)
PS = Pioo/ (11)

优化(9)-(11)中的函数，受限于等式约束，定义了一个线性规划（LP）问题，该问题适合于闭式解。Balke（1995 年，附录 B）描述了一个计算机程序，该程序接受 LP 问题的符号描述，并返回所需的界限的符号表达式。该程序通过系统地枚举对偶问题的约束多边形的顶点来工作。本文中报告的界限是通过（或测试）Balke 的程序产生的，此处将不提供证明；其正确性可以通过手动枚举顶点来验证，如[Balke, 1995, 附录 B]中所述。这些界限保证是尖锐的，因为优化是全局的。

3.2 无假设的界限

3.2.1 给定非实验数据

给定 Pxy，约束(6)和(7)诱导出以下 PNS 的上界：

0:::; PNS :::; P(x,y) +P(x', y'). (12)

然而，PN 和 PS 并不受 Pxy 的约束。

这些约束还导致因果效应 P(yx) 和 P(Yx') 的界限。

P(x, y) P(yx) ::=; 1- P(x, y')
P(x',y) P(Yx') :::; 1- P(x', y') (13)

3.2.2 因果效应的概率

给定约束（6）和（8），对 PNS 诱导的界限是：

max[O, P(yx)- P(Yx')] :::; PNS :::; min[J-'(Yx), P(y□,)] (14)

对 PN 和 PS 没有任何约束。

3.2.3 给定非实验数据和因果效应

在约束（6）、（7）和（8）下，对三个因果概率产生了以下界限

sation:

$$\max\left\{\begin{array}{c}0\\P(y_x)-P(y_{x'})\\P(y)-P(y_{x'})\\P(y_x)-P(y)\end{array}\right\}\leq PNS \qquad (15)$$

$$PNS\leq\min\left\{\begin{array}{c}P(y_x)\\P(y'_{x'})\\P(x,y)+P(x',y')\\P(y_x)-P(y_{x'})+P(x,y')+P(x',y)\end{array}\right\} \qquad (16)$$

$$\max\left\{\begin{array}{c}0\\\frac{P(y)-P(y_{x'})}{P(x,y)}\end{array}\right\}\leq PN\leq\min\left\{\begin{array}{c}1\\\frac{P(y'_{x'})-P(x',y')}{P(x,y)}\end{array}\right\} \qquad (17)$$

$$\max\left\{\begin{array}{c}0\\\frac{P(y_x)-P(y)}{P(x',y')}\end{array}\right\}\leq PS\leq\min\left\{\begin{array}{c}1\\\frac{P(y_x)-P(x,y)}{P(x',y')}\end{array}\right\} \qquad (18)$$

Thus we see that some information about PN and PS can be extracted without making any assumptions about the data-generating process. Furthermore, combined data from both experimental and nonexperimental studies yield information that neither study alone can provide.

3.3 Bounds under exogeneity (no confounding)

Definition 4 (*Exogeneity*)
A variable X is said to be exogenous for Y in model M iff

$$P(y_x)=P(y|x) \quad \textit{and} \quad P(y_{x'})=P(y|x'). \qquad (19)$$

In words, the way Y would potentially respond to experimental conditions x or x' is independent of the actual value of X.

Eq. (19) is also known as "no-confounding" [Robins and Greenland, 1989], "as if randomized," or "weak ignorability" [Rosenbaum and Rubin, 1983].

Combining Eq. (19) with the constraints of (6)–(8), the linear programming optimization (Section 3.1) yields the following results:

Theorem 1 *Under condition of exogeneity, the three probabilities of causation are bounded as follows:*

$$\max[0,P(y|x)-P(y|x')]\leq PNS\leq\min[P(y|x),P(y'|x')] \qquad (20)$$

$$\frac{\max[0,P(y|x)-P(y|x')]}{P(y|x)}\leq PN\leq\frac{\min[P(y|x),P(y'|x')]}{P(y|x)} \qquad (21)$$

$$\frac{\max[0,P(y|x)-P(y|x')]}{P(y'|x')}\leq PS\leq\frac{\min[P(y|x),P(y'|x')]}{P(y'|x')} \qquad (22)$$

[Pearl, 1999] derived Eqs. (20)-(22) under a stronger condition of exogeneity (see Definition 5). We see that under the condition of no-confounding the lower bound for PN can be expressed as

$$PN\geq 1-\frac{1}{P(y|x)/P(y|x')}\triangleq 1-\frac{1}{RR} \qquad (23)$$

where $RR\triangleq P(y|x)/P(y|x')$ is called *relative risk* in epidemiology. Courts have often used the condition $RR>2$ as a criterion for legal responsibility [Bailey *et al.*, 1994]. Eq. (23) shows that this practice represents a conservative interpretation of the "more probable than not" standard (assuming no confounding); PN must indeed be higher than 0.5 if RR exceeds 2.

3.3.1 Bounds under strong exogeneity

The condition of exogeneity, as defined in Eq. (19) is testable by comparing experimental and nonexperimental data. A stronger version of exogeneity can be defined as the joint independence $\{Y_x,Y_{x'}\}\perp\!\!\!\perp X$ which was called "strong ignorability" by Rosenbaum and Rubin (1983). Though untestable, such joint independence is implied when we assert the absence of factors that simultaneously affect exposure and outcome.

Definition 5 (*Strong Exogeneity*)
A variable X is said to be strongly exogenous for Y in model M iff $\{Y_x,Y_{x'}\}\perp\!\!\!\perp X$, that is,

$$\begin{aligned}P(y_x,y_{x'}|x)&=P(y_x,y_{x'})\\P(y_x,y'_{x'}|x)&=P(y_x,y'_{x'})\\P(y'_x,y_{x'}|x)&=P(y'_x,y_{x'})\\P(y'_x,y'_{x'}|x)&=P(y'_x,y'_{x'})\end{aligned} \qquad (24)$$

Remarkably, the added constraints introduced by strong exogeneity do not alter the bounds of Eqs. (20)–(22). They do, however, strengthen Lemma 1:

Theorem 2 *If strong exogeneity holds, the probabilities PN, PS, and PNS are constrained by the bounds of Eqs. (20)–(22), and, moreover, PN, PS, and PNS are related to each other as follows* [Pearl, 1999]*:*

$$PN=\frac{PNS}{P(y|x)} \qquad (25)$$

$$PS=\frac{PNS}{P(y'|x')} \qquad (26)$$

3.4 Identifiability under monotonicity

Definition 6 (*Monotonicity*)
A variable Y is said to be monotonic relative to variable X in a causal model M iff

$$y'_x\wedge y_{x'}=\textit{false} \qquad (27)$$

站

$$\max\left\{\begin{array}{c}0\\P(yx)-P(Yx')\\P(y)-P(Yx')\\P(yx)-P(y)\end{array}\right\}\leq PNS \qquad (15)$$

$$PNS\text{ 小于}\left\{\begin{array}{c}P(yx)\\P(y\text{◆})\\P(x,y)+P(x',y')\\-P(Y|x')+P(x,y')+\\P(x',y)\end{array}\right. \qquad (16)$$

$$\text{最大}\left\{\begin{array}{c}0\\P(y)-P(Yx)\end{array}\right\}\leq PN\leq\text{最小}\left\{\begin{array}{c}1\\P(x,y)\end{array}\right. \qquad (17)$$

$$\text{最大}\left\{\begin{array}{c}0\\\frac{P(y_x)-P(y)}{P(x',y')}\end{array}\right\}\leq PS\leq\min\left\{\begin{array}{c}1\\\frac{P(yx)-P(x,y)}{P(x',y')}\end{array}\right\} \qquad (18)$$

因此，我们可以看到，在没有任何关于数据生成过程的假设的情况下，可以提取关于 PN 和 PS 的一些信息。此外，来自实验性和非实验性研究的综合数据提供了单独研究无法提供的信息。

外生性下的界限（无混杂）

定义 4（外生性）
在模型 Miff 中，变量 X 对 Y 是外生的

P(yx) P(yjx) 和 P(Yx') P(yjx')。（19）

用话来说，Y 在实验条件 x 或 x'下的潜在反应方式与 X 的实际值无关。

式(19)也被称为“无混杂”[Robins 和 Greenland, 1989]，“如同随机化”或“弱可忽略”[Rosenbaum 和 Rubin, 1983]。
将式(19)与约束(6)-(8)相结合，线性规划优化（第 3.1 节）得到以下结果：

定理在自生性条件下，三个因果概率如下界定：

$$\max(O,P(yjx)-P(yjx')]:::;\ PNS:::;\ \min(P(yjx),P(y'jx')] \qquad (20)$$

$$\frac{\max(O,P(yjx)-P(yjx')]}{P(yjx)-P(yjx)}\leq PN\leq\frac{\min[P(yjx),P(y'jx')]}{} \qquad (21)$$

$$\frac{\max(O,P(yjx)-P(yjx')]}{P(Y)\ P(Y')}\leq PS\leq\frac{\min(P(yjx),P(y'jx'))}{Yx} \qquad (22)$$

[Pearl, 1999] 在更强的外生性条件下推导了公式(20)-(22)（参见定义 5）。我们看到

在无混杂条件下，PN 的下限可以表示为

$$PN\geq 1-\frac{1}{P(yjx)/P(yjx')\ RR}-\frac{1}{} \qquad (23)$$

其中，RR 表示相对风险，即 P(yjx)/P(yjx')
在流行病学中，法院通常将条件 RR>2 作为法律责任的标准[Bailey 等人，1994]。公式(23)表明，这种做法代表了“更有可能”标准的保守解释（假设没有混杂因素）；如果 RR 超过，PN 必须确实高于 0.5。

3.3.1 强外生性下的界限

外生性条件，如公式(19)中定义的，可以通过比较实验数据和非实验数据来检验。可以定义一个更强版本的外生性，即{Yx, Yx'}对 X 的联合独立性，这被 Rosenbaum 和 Rubin（1983）称为“强不可识别性”。尽管不可检验，但当我们断言不存在同时影响暴露和结果的因素时，这种联合独立性是隐含的。

定义 5（强外生性）如果变量 X 在模型 M 中对于 Y 是强外生的，即{Yx, Yx'}对 X 的联合独立性，

P(yx, Yx'jx) P(yx, Yx') P(Yx, y◆,jx) P(yx, y◆,)
P(y◆, Yx'lx) P(y◆, Yx') (24) P(y◆, y◆,jx)
P(y◆, y◆,)
=
=

非常明显，强外生性引入的附加约束并没有改变方程(20)和(22)的界限。然而，它们确实加强了引理 1：

定理 2 如果强外生性成立，概率 PN、PS 和 PNS 受方程{20}-{22}的界限约束，并且，此外，PN、PS 和 PNS 之间的关系如下 [Pearl, 1999]：

$$PN=\frac{PNS}{P(yjx)}- \qquad (25)$$

$$PS=\frac{PNS}{P(y'|x')} \qquad (26)$$

max(O,P(yjx)- P(yjx')]:::; PNS:::; min(P(yjx),P(y'jx')] (20) max(O, P(ylx)-P(ylx')]
P(ylx) - - P(ylx) PN min[P(ylx),P(y'lx')] (21)
max(O, P(ylx)- P(ylx')]
P('I ') P('I ') PS min(P(ylx), P(y' lx')) (22)

3.4 单调性下的可识别性

D e f i n i t i o n 6 (M o n o t o n i c y
一个变量 Y 被称为相对于变量单调
在因果模型 M 中，X 是 Y 的充分条件

Y◆ Yx' 错误（27）

Monotonicity expresses the assumption that a change from X = false to X = true cannot, under any circumstance make Y change from *true* to *false*. In epidemiology, this assumption is often expressed as "no prevention," that is, no individual in the population can be helped by exposure to the risk factor.

In the linear programming formulation of Section 3.1, monotonicity narrows the feasible space to the manifold:

$$\begin{aligned} p_{011} &= 0 \\ p_{010} &= 0 \end{aligned} \tag{28}$$

3.4.1 Given nonexperimental data

Under the constraints (6), (7), and (28), we find the same bounds for PNS as the ones obtained under no assumptions (Eq. (12)). Moreover, there are still no constraints on PN and PS. Thus, with nonexperimental data alone, the monotonicity assumption does not provide new information.

However, the monotonicity assumption induces sharper bounds on the causal effects $P(y_x)$ and $P(y_{x'})$:

$$\begin{aligned} P(y) &\le P(y_x) \le 1 - P(x, y') \\ P(x', y) &\le P(y_{x'}) \le P(y) \end{aligned} \tag{29}$$

Compared with Eq. (13), the lower bound for $P(y_x)$ and the upper bound for $P(y_{x'})$ are tightened. The importance of Eq. (29) lies in providing a simple necessary test for the commonly made assumption of "no-prevention." These inequalities are sharp, in the sense that every combination of experimental and nonexperimental data that satisfy these inequalities can be generated from some causal model in which Y is monotonic in X. Alternatively, if the no-prevention assumption is theoretically unassailable, the inequalities of Eq. (29) can be used for testing the compatibility of the experimental and non-experimental data, namely, whether subjects used in clinical trials were sampled from the same target population, characterized by the joint distribution P_{XY}.

3.4.2 Given causal effects

Constraints (6), (8), and (28) induce no constraints on PN and PS, while the value of PNS is fully determined:

$$PNS = P(y_x, y'_{x'}) = P(y_x) - P(y_{x'})$$

That is, under the assumption of monotonicity, PNS can be determined by experimental data alone, although the joint event $y_x \wedge y'_{x'}$ can never be observed.

3.4.3 Given both nonexperimental data and causal effects

Under the constraints (6)–(8) and (28), the values of PN, PS, and PNS are all determined precisely.

Theorem 3 *If Y is monotonic relative to X, then PNS, PN, and PS are given by*

$$PNS = P(y_x, y'_{x'}) = P(y_x) - P(y_{x'}) \tag{30}$$

$$PN = P(y'_{x'} | x, y) = \frac{P(y) - P(y_{x'})}{P(x, y)} \tag{31}$$

$$PS = P(y_x | x', y') = \frac{P(y_x) - P(y)}{P(x', y')} \tag{32}$$

Eqs. (30)–(32) are applicable to situations where, in addition to observational probabilities, we also have information about the causal effects $P(y_x)$ and $P(y_{x'})$. Such information may be obtained either directly, through separate experimental studies, or indirectly, from observational studies in which certain identifying assumptions are deemed plausible (e.g., assumptions that permits identification through adjustment of covariates) [Pearl, 1995].

3.5 Identifiability under monotonicity and exogeneity

Under the assumption of monotonicity, if we further assume exogeneity, then $P(y_x)$ and $P(y_{x'})$ are identified through Eq. (19), and from theorem 3 we conclude that PNS, PN, and PS are all identifiable.

Theorem 4 (*Identifiability under exogeneity and monotonicity*)
If X is exogenous and Y is monotonic relative to X, then the probabilities PN, PS, and PNS are all identifiable, and are given by

$$PNS = P(y|x) - P(y|x') \tag{33}$$

$$PN = \frac{P(y) - P(y|x')}{P(x, y)} = \frac{P(y|x) - P(y|x')}{P(y|x)} \tag{34}$$

$$PS = \frac{P(y|x) - P(y)}{P(x', y')} = \frac{P(y|x) - P(y|x')}{P(y'|x')} \tag{35}$$

These expressions are to be recognized as familiar measures of attribution that often appear in the literature. The r.h.s. of (33) is called "risk-difference" in epidemiology, and is also misnamed "attributable risk" [Hennekens and Buring, 1987, p. 87]. The probability of necessity, PN, is given by the *excess-risk-ratio* (ERR)

$$PN = \frac{P(y|x) - P(y|x')}{P(y|x)} = 1 - \frac{1}{RR} \tag{36}$$

单调性表示假设从 X 假到 X 真的变化在任何情况下都不能使 Y 从真变为假。在流行病学中，这个假设通常被表达为“无预防”，即人群中没有任何个体可以通过接触风险因素而得到帮助。

在第 3.1 节的线性规划公式中，单调性将可行空间缩小到单形：

$$\begin{aligned} P_{on\,0} &= \\ P_{01o\,0} &= \end{aligned} \tag{28}$$

3.4.1 给定非实验数据

在约束（6）、（7）和（28）下，我们找到了与无假设下（公式（12））相同的 PNS 界限。此外，PN 和 PS 仍然没有约束。因此，仅凭非实验数据，单调性假设并不提供新的信息。

然而，单调性假设使得因果效应 P(yx)和 P(Yx')的界限更加精确：

$$\begin{aligned} P(y) &\quad S\ P(yx) \quad :S \quad 1\text{-}\ P(x, y') \\ P(x', y) &\quad S\ P(Yx\bullet) \quad :S \quad P(y) \end{aligned} \tag{29}$$

与公式（13）相比，P(Yx)的下界和 P(Yx')的上界得到了收紧。公式（29）的重要性在于提供了一个简单的必要测试，用于检验“无预防”的常见假设。这些不等式是尖锐的，即在满足这些不等式的所有实验和非实验数据组合都可以从某个 Y 在 X 中单调的因果模型中生成。或者，如果“无预防”假设在理论上是不可辩驳的，那么公式（29）的不等式可以用来检验实验和非实验数据的兼容性，即临床试验中使用的受试者是否来自具有联合分布 Pxy 的同一目标人群。

3.4.2 给定因果效应

约束（6）、（8）和（28）对 PN 和 PS 没有约束，而 PNS 的值完全确定：

$$PNS = P(y_x, y'_{x'}) = P(y_x) - P(y_{x'})$$

即，在单调性的假设下，PNS 可以仅通过实验数据来确定，尽管联合事件 Yx y. 永远无法观察到。

3.4.3 给定非实验数据及因果效应

在约束条件（6）-（8）和（28）下，PN、PS 和 PNS 的值都得到了精确确定。

定理 3 如果 Y 相对于 X 是单调的，那么 PNS、PN 和 PS 由以下公式给出

$$PNS\ P(yx\text{-}Y\bullet) P(yx) - P(Yx') \tag{30}$$

$$PN\ P(y\mp x,\ y) \qquad \frac{P(y) - P(Y \mid x')}{P(x, y)} \tag{31}$$

$$PS\ P(y\models x \mid x',\ y') \qquad \frac{P(y \mid x) - P(y)}{P(x', y')} \tag{32}$$

公式（30）-（32）适用于以下情况：除了观测概率外，我们还有关于因果效应 P(yx)和 P(Yx')的信息。此类信息可以通过直接方式获得，例如通过单独的实验研究，或者通过观察研究间接获得，在这些观察研究中，某些识别假设被认为是合理的（例如，允许通过协变量调整进行识别的假设）[Pearl, 1995]。

3.5 单调性和外生性下的可识别性

在单调性的假设下，如果我们进一步假设外生性，那么 P(yx)和 P(Yx')将通过公式（19）被识别，并且根据定理 3，我们可以得出 PNS、PN 和 PS 都是可识别的。

定理 4（在外生性和单调性下的可识别性）如果 X 是外生的，Y 相对于 X 是单调的，那么 PN、PS 和 PNS 的概率都是可识别的，并且由以下公式给出

$$PNS = P(y|x) - P(y|x') \quad (33)$$

$$PN = \frac{P(y) - P(y|x')}{P(x, y)} \frac{P(y|x) - P(y|x')}{P(y|x)} \quad (34)$$

$$PS = \frac{P(yjx) - P(y) - P(yjx)}{P(x', y')} - P(yjx') \quad (35)$$

这些表达式应被视为熟悉的归因度量，常出现在文献中。公式(33)的右侧在流行病学中被称为“风险差异”，也被误称为“归因风险” [Hennekens and Buring, 1987, p. 87]。必要性概率 PN 由超额风险比（ERR）给出

$$PN \qquad \frac{P(y|x) - P(y|x')}{P(y|x)\ RR} = 1 \qquad 1 \tag{36}$$

often misnamed as the *attributable fraction*, *attributable-rate percent*, *attributed fraction for the exposed* [Kelsey *et al.*, 1996, p. 38], or *attributable proportion* [Cole, 1997]. The reason we consider these labels to be misnomers is that ERR invokes purely statistical relationships, hence it cannot in itself serve to measure attribution, unless fortified with some causal assumptions. Exogeneity and monotonicity are the causal assumptions that endow ERR with attributional interpretation, and these assumptions are rarely made explicit in the literature on attribution.

The expression for PS is likewise quite revealing

$$PS = [P(y|x) - P(y|x')]/[1 - P(y|x')], \qquad (37)$$

as it coincides with what epidemiologists call the "relative difference" [Shep, 1958], which is used to measure the *susceptibility* of a population to a risk factor x. It also coincides with what Cheng calls "causal power" (1997), namely, the effect of x on y after suppressing "all other causes of y." See Pearl (1999) for additional discussions of these expressions.

To appreciate the difference between Eqs. (31) and (36) we can rewrite Eq. (31) as

$$\begin{aligned} PN &= \frac{P(y|x)P(x) + P(y|x')P(x') - P(y_{x'})}{P(y|x)P(x)} \\ &= \frac{P(y|x) - P(y|x')}{P(y|x)} + \frac{P(y|x') - P(y_{x'})}{P(x,y)} \end{aligned} \qquad (38)$$

The first term on the r.h.s. of (38) is the familiar ERR as in (36), and represents the value of PN under exogeneity. The second term represents the correction needed to account for X's non-exogeneity, i.e. $P(y_{x'}) \neq P(y|x')$. We will call the r.h.s. of (38) by corrected excess-risk-ratio (CERR).

From Eqs. (33)–(35) we see that the three notions of causation satisfy the simple relationships given by Eqs. (25) and (26) which we obtained under the strong exogeneity condition. In fact, we have the following theorem.

Theorem 5 *Monotonicity (27) and exogeneity (19) together imply strong exogeneity (24).*

3.6 Summary of results

Table 1 lists the best estimate of PN under various assumptions and various types of data—the stronger the assumptions, the more informative the estimates. We see that the excess-risk-ratio (ERR), which epidemiologists commonly identify with the probability of causation, is a valid measure of PN only when two assumptions can be ascertained: exogeneity (i.e., no confounding) and monotonicity (i.e., no prevention). When monotonicity does not hold, ERR provides merely a lower bound for PN, as shown in Eq. (21). (The upper bound is usually unity.) In the presence of confounding, ERR must be corrected by the additive term $[P(y|x') - P(y_{x'})]/P(x,y)$, as stated in (38). In other words, when confounding bias (of the causal effect) is positive, PN is higher than ERR by the amount of this additive term. Clearly, owing to the division by $P(x,y)$, the PN bias can be many times higher than the causal effect bias $P(y|x') - P(y_{x'})$. However, confounding results only from association between exposure and other factors that affect the outcome; one need not be concerned with associations between such factors and susceptibility to exposure, as is often assumed in the literature [Khoury *et al.*, 1989, Glymour, 1998].

Table 1: PN as a function of assumptions (exogeneity or monotonicity) and available data (experimental or nonexperimental or both). ERR stands of the excess-risk-ratio and CERR is given in Eq. (38). The non-entries (—) represent vacuous bounds, that is, $0 \leq PN \leq 1$.

Assumptions		Data Available		
Exo.	Mono.	Exp.	Non-exp.	Combined
+	+	ERR	ERR	ERR
+	−	bounds	bounds	bounds
−	+	—	—	CERR
−	−	—	—	bounds

The last two rows in Table 1 correspond to no assumptions about exogeneity, and they yield vacuous bounds for PN when data come from either experimental or observational study. In contrast, informative bounds (17) or point estimates (38) are obtained when data from experimental and observational studies are combined. Concrete use of such combination will be illustrated in Section 4.

4 Example 1: Legal Responsibility

A lawsuit is filed against the manufacturer of drug x, charging that the drug is likely to have caused the death of Mr. A, who took the drug to relieve symptom S associated with disease D.

The manufacturer claims that experimental data on patients with symptom S show conclusively that drug x may cause only minor increase in death rates. The plaintiff argues, however, that the experimental study is of little relevance to this case, because it represents the effect of the drug on *all* patients, not on patients like Mr. A who actually died while using drug x. Moreover, argues the plaintiff, Mr. A is unique in that he used the drug on his own voli-

常被误称为归因分数、归因率百分比、暴露归因分数[凯利等，1996，第 38 页]或归因比例[科尔，1997]。我们认为这些标签是误称的原因在于 ERR 仅涉及纯粹统计

统计关系，因此它本身不能用来衡量归因，除非与某些因果假设相结合。外生性和单调性是赋予 ERR 归因解释的因果假设，而这些假设在归因文献中很少被明确提出。

该表达式对于 PS 同样具有很高的揭示性

PS [P(yix) - P(yix')]/[1- P(yix')], (37) 翻译为：PS [P(yix) - P(yix')]/[1- P(yix')], (37)

与流行病学家所说的“相对差异”[Shep, 1958]相一致，该差异用于衡量人群对风险因素 x 的易感性。它也与程所说的“因果力”（1997）相一致，即在抑制“y 的所有其他原因”之后，x 对 y 的影响。参见 Pearl（1999）对这些表达式的进一步讨论。

要理解方程（31）和（36）之间的区别，我们可以将方程（31）重写为 PN

$$\begin{aligned} &= \frac{P(y|x)P(x)\ P(y|x')P(x') - P(Y|x')}{P(y|x)P(x)} \\ &= \frac{P(yix) - P(yix')}{P(yix)\,P(x,y)} + \frac{P(yix') - P(Yx')}{} \end{aligned} \qquad (38)$$

（38）式的右侧的第一个项是熟悉的 ERR，正如（36）式中所示，它代表了在外生性条件下的 PN 的值。第二个项代表了为了解释 X 的非外生性所需的校正，即 P(Y|x') = I P(y|x')。我们将（38）式的右侧称为校正后的超额风险比（CERR）。

从方程（33）-（35）我们可以看出，三种因果关系的概念满足我们在强外生性条件下获得的方程（25）和（26）给出的简单关系。事实上，我们有一个以下定理。

表 1：PN 作为假设（外生性或单调性）和可用数据（实验性或非实验性或两者）的函数。ERR 代表超额风险比，CERR 由公式（38）给出。非条目（-）表示空集界限，即 0 < PN < 1。假设 外生性 单调性

		数据可用		
		实验	非专家组合	
++		ERR	错误 错误	
+	−	范围	界限界限	
−	+	—	—	CERR
−	−	—	—	界限

当单调性不成立时，ERR 仅提供 PN 的下界，如公式（21）所示。（上界通常是 1。）在存在混杂因素的情况下，ERR 必须通过加性项[P(yix') - P(Yx')]/P(x, y)进行校正，如（38）所述。换句话说，当因果效应的混杂偏差（正值）时，PN 比 ERR 高这个加性项的量。显然，由于除以 P(x, y)，PN 偏差可以比因果效应偏差 P(yIx') - P(Yx')高许多倍。然而，混杂仅来自暴露与其他影响结果的因素之间的关联；不需要担心这些因素与暴露易感性之间的关联，正如文献中经常假设的那样（Khoury 等，1989，Glymour，1998）。

表 1 的最后两行对应于对外生性的无假设，并且当数据来自实验性或观察性研究时，PN 的界限是空洞的。相比之下，当实验性和观察性研究的数据结合时，可以获得信息界限（17）或点估计（38）。这种组合的具体应用将在第 4 节中说明。

4 例子 1：法律责任

定理 5 单调性（27）和外生性（19）一起意味着强外生性（24）。

3.6 结果总结

表 1 列出了在各种假设和各种类型的数据下 PN 的最佳估计值。我们注意到，流行病学家通常将超额风险比（ERR）与病因概率等同起来，只有当两个假设可以确定时，ERR 才是 PN 的有效度量：外生性（即，无混杂）和单调性（即。

一起诉讼被提起，指控药物 x 的制造商，称该药物可能导致 A 先生死亡，A 先生服用该药物是为了缓解与疾病 D 相关的症状 S。

制造商声称，关于有症状 S 的患者的实验数据明确表明，药物 x 可能只会导致死亡率的轻微增加。然而，原告认为，这项实验研究对这个案件几乎没有相关性，因为它代表了药物对所有患者的影响，而不是对像 A 先生这样的实际使用药物 x 而死亡的患者的影响。此外，原告认为，A 先生是独一无二的，因为他自愿使用该药物。

Table 2: Frequency data (hypothetical) obtained in experimental and nonexperimental studies, comparing deaths (in thousands) among drug users (x) and non-users (x').

	Experimental		Nonexperimental	
	x	x'	x	x'
Deaths(y)	16	14	2	28
Survivals(y')	984	986	998	972

tion, unlike subjects in the experimental study who took the drug to comply with experimental protocols. To support this argument, the plaintiff furnishes non-experimental data indicating that most patients who chose drug x would have been alive if it were not for the drug. The manufacturer counter-argues by stating that: (1) counterfactual speculations regarding whether patients would or would not have died are purely metaphysical and should be avoided, and (2) nonexperimental data should be dismissed a priori, on the ground that such data may be highly biased; for example, incurable terminal patients might be more inclined to use drug x if it provides them greater symptomatic relief. The court must now decide, based on both the experimental and non-experimental studies, what the probability is that drug x was in fact the cause of Mr. A's death.

The (hypothetical) data associated with the two studies are shown in Table 2. The experimental data provide the estimates

$$\begin{aligned} P(y_x) &= 16/1000 = 0.016 \\ P(y_{x'}) &= 14/1000 = 0.014 \\ P(y'_{x'}) &= 1 - P(y_{x'}) = 0.986 \end{aligned}$$

The non-experimental data provide the estimates

$$\begin{aligned} P(y) &= 30/2000 = 0.015 \\ P(x,y) &= 2/2000 = 0.001 \\ P(x',y') &= 972/2000 = 0.486 \end{aligned}$$

Since both the experimental and nonexperimental data are available, we can obtain bounds on all three probabilities of causation through Eqs. (15)–(18) without making any assumptions about the underlying mechanisms. The data in Table 2 imply the following numerical results:

$$0.002 \leq PNS \leq 0.016 \quad (39)$$
$$1.0 \leq PN \leq 1.0 \quad (40)$$
$$0.002 \leq PS \leq 0.031 \quad (41)$$

These figures show that although surviving patients who didn't take drug x have only less than 3.1% chance to die had they taken the drug, there is 100% assurance (barring sample errors) that those who took the drug and died would have survived had they not taken the drug. Thus the plaintiff was correct; drug x was in fact responsible for the death of Mr. A.

If we assume that drug x can only cause, but never prevent, death, Theorem 3 is applicable and Eqs. (30)–(32) yield

$$PNS = 0.002 \quad (42)$$
$$PN = 1.0 \quad (43)$$
$$PS = 0.002 \quad (44)$$

Thus, we conclude that drug x was responsible for the death of Mr. A, with or without the no-prevention assumption.

Note that a straightforward use of the experimental excess-risk-ratio would yield a much lower (and incorrect) result:

$$\frac{P(y_x) - P(y_{x'})}{P(y_x)} = \frac{0.016 - 0.014}{0.016} = 0.125 \quad (45)$$

Evidently, what the experimental study does not reveal is that, given a choice, terminal patients stay away from drug x. Indeed, if there were any terminal patients who would choose x (given the choice), then the control group (x') would have included some such patients (due to randomization) and so the proportion of deaths among the control group $P(y_{x'})$ would have been higher than $P(x', y)$, the population proportion of terminal patients avoiding x. However, the equality $P(y_{x'}) = P(y, x')$ tells us that no such patients were present in the control group, hence (by randomization) no such patients exist in the population at large and therefore none of the patients who freely chose drug x was a terminal case; all were susceptible to x.

The numbers in Table 2 were obviously contrived to show the usefulness of the bounds in Eqs. (15)-(18). Nevertheless, it is instructive to note that a combination of experimental and non-experimental studies may unravel what experimental studies alone will not reveal.

5 Example 2: Personal Decision Making

Consider the case of Mr. B, who is one of the surviving patients in the observational study of Table 2. Mr. B wonders how safe it would be for him to take drug x, given that he has refrained thus far from taking the drug and that he managed to survive the disease. His argument for switching to the drug rests on the observation that only 2 out of 1000 drug users died in

表 2：实验性和非实验性研究中获得的数据频率（假设性），比较药物使用者（x）和非使用者（x'）中的死亡人数（以千计）

	实验		非实验	
	x	x'	x	x'
死亡人数(y)	16	14	2	28
存活人数(y')	984	986	998	972

诉讼，与实验研究中为了遵守实验方案而服用药物的受试者不同。为了支持这一论点，原告提供了非实验数据，表明大多数选择药物 x 的患者如果没有服用该药物本应存活。制造商反驳说：（1）关于患者是否死亡的假设性推测纯粹是形而上学的，应予以避免；（2）非实验数据应被先验地排除，理由是此类数据可能存在高度偏差；例如，无法治愈的晚期患者如果药物 x 能提供更大的症状缓解，他们可能更倾向于使用药物 x。现在，法院必须根据实验和非实验研究，确定药物 x 是否实际上是导致 A 先生死亡的原因。

与两项研究相关的（假设性）数据如表 2 所示。实验数据提供了估计

P(yx) = 16/1000 = 0.016
P(Yx') =14/1000 =0.014
P(y,) =1- P(Yx') =0.986

非实验数据提供了估计值

P(y) =30/2000 =0.015
P(x, y) =2/2000 =0.001
P(x', y') =972/2000 =0.486

由于实验数据和非实验数据都可用，我们可以通过公式（15）至（18）获得所有三个因果概率的界限，而不需要对潜在机制做出任何假设。表 2 中的数据表明以下数值结果：

0.002 ::; PNS ::; 0.016 (39)
1.0 ::; PN ::; 1.0 (40)
0.002 ::; p s ::; 0.031 (41)

此外，这些数字显示，尽管未服用药物 x 的存活患者数量不到 3

如果他们服用了这种药物，死亡的概率为 1%，但如果没有服用药物，那些服用药物后死亡的人 100%（除非样本误差）会存活。因此，原告是正确的；实际上药物是导致 A 先生死亡的原因。

假设药物 x 只能导致死亡，而不能预防死亡，定理 3 适用，方程（30）和（32）得出

PNS = 0.002 (42)
PN = 1.0 (43)
PS = 0.002 (44)

因此，我们得出结论，药物 x 是导致 A 先生死亡的原因，无论是否有不预防的假设。

注意，直接使用实验的过量风险比会得出一个更低（且不正确）的结果：

$$\frac{P(yx) - P(Yx')}{P(yx)} = \frac{0.016 - 0.014}{0.016} = 0\_125 \quad (45)$$

显然，实验研究没有揭示的是，如果给以选择，晚期患者会远离药物 x。事实上，如果有任何晚期患者会选择 x（如果给予选择），那么对照组（x'）就会包括一些这样的患者（由于随机化），因此对照组中死亡的比例 P(Yx') 就会高于 P(x', y)，即避免 x 的晚期患者的总体比例。然而，P(Yx') P(y, x')的等式告诉我们，对照组中没有这样的患者，因此（由于随机化）总体中也不存在这样的患者，因此没有任何自由选择药物 x 的患者是晚期病例；他们都是对 x 有反应的。

表 2 中的数字明显是编造的，以显示方程(15)－(18)中界限的有用性。然而，值得注意的是，结合实验和非实验研究可能会揭示仅靠实验研究无法揭示的内容。

5 例子 2：个人决策

考虑表 2 中的病例，B 先生是其中一位幸存的患者。B 先生想知道，鉴于他迄今为止一直未服用药物 x，并且成功战胜了疾病，他开始服用这种药物是否安全。他转向该药物的理由是，在 1000 名药物使用者中，只有 2 人死亡。

the observational study, which he considers a rather small risk to take, given the effectiveness of the drug as a pain killer.

Conventional wisdom instructs us to warn Mr. B against consulting a nonexperimental study in matters of decisions, since such studies are marred with uncontrolled factors, which tend to bias effect estimates. Specifically, the death rate of 0.002 among drug users may be indicative of low tolerance to discomfort, or of membership in a medically-informed socio-economic group. Such factors do not apply to Mr. B, who did not use the drug in the past (be it by choice, instinct or ignorance), and who is now considering switching to the drug by rational deliberation. Conventional wisdom urges us to refer Mr. B to the randomized experimental study of Table 2, from which the death rate under controlled administration of the drug was evaluated to be $P(y_x) = 0.016$, eight times higher than 0.002.

What would his risk of death be, if Mr. B decides to start taking the drug? 0.2 percent or 1.6 percent?

The answer is that neither number is correct. Mr. B cannot be treated as a random patient in either study, because his history of not using the drug and his survival thus far puts him in a unique category of patients, for which the effect of the drug was not studied.[2] These two attributes provide extra evidence about Mr. B's sensitivity to the drug. This became clear already in Example 1, where we discovered definite relationships among these attributes – for some obscure reasons, terminal patients chose not to use the drug.

To properly account for this additional evidence, the risk should be measured through the counterfactual expression $PS = P(y_x|x', y')$; the probability that a patient who survived with no drug would have died had he/she taken the drug. The appropriate bound for this probability is given in Eq. (41):

$$0.002 \leq PS \leq 0.031$$

Thus, Mr. B's risk of death (upon switching to drug usage) can be as high as 3.1 percent; more than 15 times his intuitive estimate of 0.2 percent, and almost twice the naive estimate obtained from the experimental study.

However, if the drug can safely be assumed to have no death-preventing effects, then monotonicity applies, and the appropriate bound is given by Eq. (44), $PS = 0.002$, which coincides with Mr. B's intuition.

[2]The appropriate experimental design for measuring the risk of interest is to conduct a randomized clinical trial on patients in the category of Mr. B, that is, to subject a random sample of non-users to a period of drug treatment and measure their rate of survival.

6 Conclusion

This paper shows how useful information about probabilities of causation can be obtained from experimental and observational studies, with weak or no assumptions about the data-generating process. We have shown that, in general, bounds for the probabilities of causation can be obtained from combined experimental and nonexperimental data. These bounds were proven to be sharp and, therefore, they represent the ultimate information that can be extracted from statistical methods. We clarify the two basic assumptions – exogeneity and monotonicity – that must be ascertained before statistical measures such as excess-risk-ratio could represent attributional quantities such as probability of causation.

One application of this analysis lies in the automatic generation of verbal explanations, where the distinction between necessary and sufficient causes has important ramifications. As can be seen from the definitions and examples discussed in this paper, necessary causation is a concept tailored to a specific event under consideration (singular causation), whereas sufficient causation is based on the general tendency of certain event *types* to produce other event types. Adequate explanations should respect both aspects. Clearly, some balance must be made between the necessary and the sufficient components of causal explanation, and the present paper illuminates this balance by formally explicating the basic relationships between the two components. In Pearl (2000, chapter 10) it is further shown that PN and PS are too crude for capturing probabilities of causation in multi-stage scenarios, and that the structure of the intermediate process leading from cause to effect must enter the definitions of causation and explanation. Such considerations will be the subject of future investigation (See [Halpern and Pearl, 2000]).

Another important application of probabilities of causation is found in decision making problems. As was pointed out in Pearl (2000, pp. 217-219) and illustrated in Section 5, the counterfactual "y would have been true if x were true" can often be translated into a conditional action claim "given that currently x and y are false, y will be true if we do x." The evaluation of such conditional predictions, and the probabilities of such predictions, are commonplace in decision making situations, where actions are brought into focus by certain eventualities that demand remedial correction. In troubleshooting, for example, we observe undesirable effects $Y = y$ that are potentially caused by other conditions $X = x$ and we wish to predict whether an action that brings about a change in X would remedy the situation. The information provided by the

观察性研究，他认为这是一个相对较小的风险，因为药物作为止痛药的有效性。

传统智慧教导我们，在决策问题上，要警告 B 先生不要咨询非实验性研究，因为这类研究受到未受控制的因素的影响，这些因素往往会导致效果估计偏差。具体来说，药物使用者的死亡率 0.002 可能表明对不适的耐受性低，或者属于一个医学知识丰富的社会经济群体。这些因素并不适用于 B 先生，他过去没有使用过药物（无论是出于选择、本能还是无知），现在他正在通过理性思考考虑改用该药物。传统智慧敦促我们向 B 先生推荐表 2 中的随机实验研究，该研究评估了药物在受控给药下的死亡率 P(yx)为 0.016，是 0.002 的八倍。

如果 B 先生决定开始服用该药物，他的死亡风险是多少？0.2%还是 1.6%?

答案是这两个数字都不正确。B 先生不能被视为两个研究中的随机患者，因为他的不使用药物的历史以及至今的生存状况使他成为一类独特的患者，该药物对这些患者的效果并未进行研究。这两个属性为 B 先生对药物敏感性的额外证据。这已经在示例 1 中变得明显，我们发现这些属性之间存在明确的关系——出于一些神秘的原因，晚期患者选择不使用该药物。

为了正确考虑这一额外证据，风险应通过反事实表达式 PS P(yx jx1, y1)来衡量；即一个未使用药物而幸存的患者在服用药物后可能会死亡的概率。这个概率的适当界限由公式(41)给出：

0.002 :;; p s :;; 0.031

因此，B 先生的死亡风险（在转为使用药物后）可能高达 3.1%，比他直观估计的 0.2%高出 15 倍，几乎是实验研究中得到的朴素估计的两倍。

然而，如果可以安全地假设药物没有预防死亡的效果，那么单调性适用，适当的界限由公式(44)给出，PS 0.002，这与 B 先生的直觉相符。

2 测量所需风险的适当实验设计是在 B 先生所在的类别患者中进行随机临床试验，即对非使用者进行一段时间的药物治疗，并测量他们的存活率。

6 结论

本文展示了如何从实验和观察研究中获得关于因果概率的有用信息，对数据生成过程几乎没有或没有假设。我们已经证明，通常可以从结合实验和非实验数据中获得因果概率的界限。这些界限已被证明是尖锐的，因此它们代表了可以从统计方法中提取的最终信息。我们阐明了在统计度量（如超额风险比）能够代表归因量（如因果概率）之前必须确定的两个基本假设——外生性和单调性。

这种分析的一个应用在于自动生成口头解释，其中必要原因和充分原因之间的区别具有重要意义。正如本文中讨论的定义和例子所示，必要原因是针对考虑的特定事件（单一原因）的概念，而充分原因则是基于某些事件类型产生其他事件类型的一般趋势。适当的解释应尊重这两个方面。显然，在因果解释的必要和充分成分之间必须取得某种平衡，而本文通过正式阐述这两个成分之间的基本关系来阐明这种平衡。在 Pearl（2000 年，第 10 章）中进一步表明，PN 和 PS 对于捕捉多阶段场景中的因果概率过于粗糙，并且导致从原因到效果的中介过程的结构必须进入因果和解释的定义。这些问题将是未来调查的主题（参见[Halpern 和 Pearl，2000]）。

因果概率的另一个重要应用在于决策问题。正如 Pearl（2000 年，第 217-219 页）所指出的，并在第 5 节中举例说明，反事实“如果 x 为真，y 就会为真”通常可以翻译为条件行动主张“如果目前 x 和 y 都是假的，那么如果我们做 x，y 就会为真。”在决策情境中，这种条件预测的评价以及这种预测的概率是常见的，在这些情境中，行动被某些需要补救纠正的偶然事件所关注。例如，在故障排除中，我们观察到由其他条件 X 引起的潜在不希望的效果 Y，我们希望预测是否采取改变 X 的行动会解决这种情况。提供的信息

evidence y and x is extremely valuable, and it must be processed before we can predict the effect of any action[3]. Thus, the expressions developed in this paper constitute bounds on the effectiveness of pending policies, when full knowledge of the state of affairs is not available, yet the pre-action states of the decision variable (X) and the outcome variable (Y) are known.

For these bounds to be valid in policy making, the data generating model must be time-invariant, that is, all probabilities associated with the model should represent epistemic uncertainty about static, albeit unknown boundary conditions $U = u$. The constancy of U is well justified in the control and diagnosis of physical systems, where U represents fixed, but unknown physical characteristics of devices or subsystems. The constancy approximation is also justified in the health sciences where patients' genetic attributes and physical characteristics can be assumed relatively constant between observation and treatment. For instance, if a patient in the example of Section 5 wishes to assess the risk of switching from no drug to drug, it is reasonable to assume that this patient's susceptibility to the drug remains constant through the interim period of analysis. Therefore, the risk associated with this patient's decision will be well represented by the counterfactual expression $PS = P(y_x|x', y')$, and should be assessed by the bounds in Eq. (41).

The constancy assumption is less justified in economic systems, where agents are bombarded by rapidly fluctuating streams of external forces ("shocks" in econometric terminology) and inter-agents stimuli. These forces and stimuli may vary substantially during the policy making interval and they require, therefore, detailed time-dependent analysis. The canonical violation of the constancy assumption occurs, of course, in quantum mechanical systems, where the indeterminism is "intrinsic" and memory-less, and where the existence of a deterministic relationship between the boundary conditions and measured quantities is no longer a good approximation. A method of incorporating such intrinsic indeterminism into counterfactual analysis is outlined in Pearl (2000, p. 220).

Acknowledgements

We thank the referees for making useful suggestions on the first draft. Full version of this paper will appear in the Annals of Mathematics and AI, 2000. This research was supported in parts by grants from NSF, ONR and AFOSR and by a Microsoft Fellowship to the first author.

[3]Such processing have been applied indeed to the evaluation of economic policies [Balke and Pearl, 1995] and to repair-test strategies in troubleshooting [Breese and Heckerman, 1996]

References

[Bailey *et al.*, 1994] L. A. Bailey, L. Gordis, and M. Green. Reference guide on epidemiology. *Reference Manual on Scientific Evidence*, 1994. Federal Judicial Center. Available online at http://www.fjc.gov/EVIDENCE/science/sc.ev_sec.html.

[Balke and Pearl, 1994] A. Balke and J. Pearl. Probabilistic evaluation of counterfactual queries. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, volume Volume I, pages 230–237. MIT Press, Menlo Park, CA, 1994.

[Balke and Pearl, 1995] A. Balke and J. Pearl. Counterfactuals and policy analysis in structural models. In P. Besnard and S. Hanks, editors, *Uncertainty in Artificial Intelligence 11*, pages 11–18. Morgan Kaufmann, San Francisco, 1995.

[Balke, 1995] A. Balke. *Probabilistic Counterfactuals: Semantics, Computation, and Applications*. PhD thesis, Computer Science Department, University of California, Los Angeles, CA, November 1995.

[Breese and Heckerman, 1996] J.S. Breese and D. Heckerman. Decision-theoretic troubleshooting: A framework for repair and experiment. In E. Horvitz and F. Jensen, editors, *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, pages 124–132. Morgan Kaufmann, San Francisco, CA, 1996.

[Cheng, 1997] P.W. Cheng. From covariation to causation: A causal power theory. *Psychological Review*, 104(2):367–405, 1997.

[Cole, 1997] P. Cole. Causality in epidemiology, health policy, and law. *Journal of Marketing Research*, 27:10279–10285, 1997.

[Galles and Pearl, 1997] D. Galles and J. Pearl. Axioms of causal relevance. *Artificial Intelligence*, 97(1-2):9–43, 1997.

[Galles and Pearl, 1998] D. Galles and J. Pearl. An axiomatic characterization of causal counterfactuals. *Foundations of Science*, 3(1):151–182, 1998.

[Glymour, 1998] C. Glymour. Psychological and normative theories of causal power and the probabilities of causes. In G.F. Cooper and S. Moral, editors, *Uncertainty in Artificial Intelligence*, pages 166–172. Morgan Kaufmann, San Francisco, CA, 1998.

[Halpern and Pearl, 2000] J. Y. Halpern and J. Pearl. Causes and explanations: A structural-model approach. Technical Report R-266, Cognitive System Laboratory, Department of Computer Science, University of California, Los Angeles, March, 2000.

证据 y 极其宝贵，我们必须在预测任何行动的效果之前对其进行处理 3。因此，本文中开发的表达式构成了在无法获得事态全貌的情况下，决策变量（X）和结果变量（Y）的预行动状态已知时，待定政策的有效性的界限。

为了使这些界限在政策制定中有效，数据生成模型必须是时间不变的，即与模型相关的所有概率都应该代表关于静态（尽管未知）边界条件 U 的认知不确定性。在控制和诊断物理系统时，U 代表设备或子系统的固定但未知的物理特性，这种恒定性假设是合理的。在健康科学中，患者的遗传特性和生理特征在观察和治疗之间可以假设相对恒定，这种恒定性近似也是合理的。例如，如果第 5 节中的患者想评估从无药到用药的风险，可以合理地假设该患者对药物的反应性在分析期间保持不变。因此，与该患者决策相关的风险将由反事实表达式 PS P(Yxlx',y')很好地表示，并且应通过方程（41）中的界限进行评估。

在经济系统中，这种恒定性假设的合理性较低，因为代理人会受到快速波动的外部力量（计量经济学术语中的“冲击”）和代理人之间刺激的冲击。这些力量和刺激在政策制定期间可能会发生很大变化，因此需要详细的时间依赖性分析。当然，恒定性假设的典型违反发生在量子力学系统中，那里的不确定性是“内在的”和“无记忆的”，并且边界条件和测量量之间的确定性关系不再是一个好的近似。在 Pearl（2000，第 220 页）中概述了一种将这种内在的不确定性纳入反事实分析的方法。

致谢

我们感谢审稿人对我们第一稿提出的宝贵建议。本文的完整版本将于 2000 年在《数学与人工智能年鉴》上发表。本研究部分得到了 NSF、ONR 和 AFOSR 的资助，以及第一作者获得的微软奖学金。

确实，此类处理已应用于经济政策的评估[Balke 和 Pearl，1995]以及故障排除中的修复-测试策略[Breese 和 Heckerman，1996]。

参考文献

[Bailey 等，1994] L. A. Bailey，L. Gordis，和 M. Green. 流行病学参考指南
ogy. 科学证据参考手册，
1994 年。联邦司法中心。可在 http:/jwww.fjc.gov/EVIDENCE/sciencejsc_ev _sec.html 在线获取。

[Balke 和 Pearl，1994] A. Balke 和 J. Pearl. 对反事实查询的概率评估。在
《第十二届全国人工智能会议论文集》
人工智能，第 I 卷，第 230-237 页。
麻省理工学院出版社，加利福尼亚州门洛帕克，1994 年。

[Balke 和 Pearl，1995] A. Balke 和 J. Pearl. 结构模型中的反事实与政策分析。
在 P. Besnard 和 S. Hanks 编辑的《不确定性》中。
在人工智能 11 卷，第 11-18 页。摩根
贾夫曼，旧金山，1995 年。

[Balke，1995] A. Balke. 概率反事实：语义、计算和应用。博士论文
计算机科学系，加州大学洛杉矶分校，加利福尼亚州，
洛杉矶，1995 年 11 月。

[布里斯和贝克曼，1996] J.S. 布里斯和 D. 贝克曼. 决策理论故障排除
框架：修复和实验。在 E. Horvitz 和 F. Jensen 编辑的《会议论文集》中。
第十二次人工智能不确定性会议
智能，第 124-132 页。摩根考夫曼出版社，旧金山，
加利福尼亚州，1996 年。

[程，1997] P.W. 程著. 从协变关系到因果关系：因果力理论。心理评论，104(2)：367-405，1997。

[科尔，1997] 科尔。流行病学中的因果关系
政策，和法律。市场研究杂志，
27:10279-10285, 1997.

[加勒斯和佩尔，1997] D. 加勒斯和 J. 佩尔。因果相关公理。人工智能，97(1-2):9-43，1997。

[加勒斯和佩尔，1998] D. 加勒斯和 J. 佩尔。因果反事实的公理化描述。科学基础，3(1):151-182，1998。

[Glymour，1998] C. Glymour. 心理与规范理论中的因果力及其原因的概率。在 G.F. Cooper 和 S. Moral 编著的《》中。
人工智能中的确定性，第 166-172 页。
摩根考夫曼，旧金山，加利福尼亚州，1998 年。

[Halpern 和 Pearl，2000] J. Halpern 和 J. Pearl.
原因与解释：一种结构模型方法。技术报告 R-266，认知系统实验室，加州大学洛杉矶分校计算机科学系，
2000 年 3 月。

[Halpern, 1998] J.Y. Halpern. Axiomatizing causal reasoning. In G.F. Cooper and S. Moral, editors, *Uncertainty in Artificial Intelligence*, pages 202–210. Morgan Kaufmann, San Francisco, CA, 1998.

[Hennekens and Buring, 1987] C.H. Hennekens and J.E. Buring. *Epidemiology in Medicine*. Brown, Little, Boston, 1987.

[Kelsey *et al.*, 1996] J.L. Kelsey, A.S. Whittemore, A.S. Evans, and W.D. Thompson. *Methods in Observational Epidemiology*. Oxford University Press, New York, 1996.

[Khoury *et al.*, 1989] M.J. Khoury, W.D. Flanders, S. Greenland, and M.J. Adams. On the measurement of susceptibility in epidemiologic studies. *American Journal of Epidemiology*, 129(1):183–190, 1989.

[Pearl, 1995] J. Pearl. Causal diagrams for experimental research. *Biometrika*, 82:669–710, December 1995.

[Pearl, 1999] J. Pearl. Probabilities of causation: three counterfactual interpretations and their identification. *Synthese*, 121(1-2):93–149, November 1999.

[Pearl, 2000] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, NY, 2000.

[Robins and Greenland, 1989] J.M. Robins and S. Greenland. The probability of causation under a stochastic model for individual risk. *Biometrics*, 45:1125–1138, 1989.

[Robins, 1987] J.M. Robins. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *Journal of Chronic Diseases*, 40(Suppl 2):139S–161S, 1987.

[Rosenbaum and Rubin, 1983] P. Rosenbaum and D. Rubin. The central role of propensity score in observational studies for causal effects. *Biometrica*, 70:41–55, 1983.

[Shep, 1958] M.C. Shep. Shall we count the living or the dead? *New England Journal of Medicine*, 259:1210–1214, 1958.

[Halpern, 1998] J.Y. Halpern. 建立因果推理公理。在 G.F. Cooper 和 S. Moral 主编，
人工智能中的不确定性，第 202- 页。
210. 摩根考夫曼出版社，旧金山，加利福尼亚州，1998 年。

[Hennekens and Buring, 1987] C.H. Hennekens 和 J.E. Buring. 医学流行病学。布朗，利特
标题，波士顿，1987 年。

[凯利等，1996] J.L. 凯利，A.S. 威特莫尔，A.S. 埃文斯，和 W.D. 汤普森。观察方法
观察流行病学。牛津大学出版社
纽约，1996 年。

[库雷等，1989] M.J. 库雷，W.D. 弗兰德斯，S. 格林兰德，和 M.J. 亚当斯。关于度量...
确定流行病学研究中易感性的测量。
美国流行病学杂志，129(1):183-190
1989.

[Pearl, 1995] J. Pearl. 因果图在实验研究中的应用。Biometrika，82:669-710，1995 年 12 月。

[Pearl, 1999] J. Pearl. 因果概率：三种反事实解释及其识别。Synthese，121(1-2):93-149，1999 年 11 月。

[Pearl, 2000] J. Pearl. 因果性：模型、推理，
和推断。剑桥大学出版社，纽约，2000 年。

[Robins and Greenland, 1989] J.M. Robins 和 S. Greenland. 在个体风险随机模型下因果概率。生物统计学，45:1125-1138，1989。

[Robins, 1987] J.M. Robins. 在具有持续暴露期的死亡率研究中识别和估计因果参数的图形方法。慢性病杂志，40(Suppl 2):139S-161S，1987。

sure periods. Journal of Chronic Diseases, 40(Suppl
2):139S-161S, 1987。

[Rosenbaum and Rubin, 1983] P. Rosenbaum 和 D. Rubin. 倾向性评分在观察性研究中因果效应的核心作用。生物统计学，70:41-55，1983。

[Shep, 1958] M.C. Shep. 我们该计算活人
还是死者？新英格兰医学杂志，
259:1210-1214, 1958.