arXiv:2602.09235v1 [cs.LG] 9 Feb 2026

RAPID：合成微观数据中的属性预测诱发披露风险

Matthias Templ *† Oscar Thees † Roman Müller †

摘要

统计数据匿名化日益依赖完全合成的微观数据；对于此类数据，经典的身份披露度量所提供的信息，不如对手从发布数据中推断敏感属性的能力更有参考价值。我们提出 RAPID（属性预测引发的披露风险），这是一种在现实攻击模型下直接量化推断脆弱性的披露风险度量。对手仅使用发布的合成数据训练预测模型，并将其应用于真实个人的准标识符。对于连续型敏感属性，RAPID 报告预测值落在指定相对误差容限内的记录比例。对于分类属性，我们提出一种基线归一化置信度分数，用以衡量攻击者对真实类别的置信程度比仅依据类别流行率所预期的高出多少，并将风险概括为超过策略定义阈值的记录所占比例。该构造产生一种可解释、有界的风险度量，对类别不平衡具有稳健性，不依赖于任何特定合成器，并可与任意学习算法配合使用。我们利用模拟与真实数据，展示阈值校准、不确定性量化以及对合成数据生成器的比较评估。我们的结果表明，RAPID 为属性推断披露风险提供了一个实用且符合现实攻击者情境的上界，可补充现有的效用诊断与披露控制框架。

关键词： 合成数据 ，披露风险、属性推断、统计 披露控制、隐私 y

1 引言

开放研究数据（ORD）日益被视为实现科学透明度和可重复性的关键[Nosek 等人， 2015]。然而，共享数据集的比例仍然偏低：例如，目前瑞士国家科学基金会资助的项目中，只有 23% 提供开放研究数据（ORD） [瑞士国家科学基金（SNSF）， 2024]。法律限制和合同使用限制常常阻碍微观数据的发布，而对于许多研究团队而言，匿名化的技术障碍仍然相当大。

完全合成的微观数据 [Rubin，1993] ——即所有记录均为模拟生成，而非对真实个人记录进行扰动后得到的副本——为解决这一问题提供了一种很有前景的方案。然而，合成数据也让数据管理者面临一个迫切的问题：当主要威胁不再是重新识别，而是对手利用发布数据推断敏感属性的能力时，应如何量化披露风险？

1.1 用户视角：数据效用

从用户视角来看，合成数据必须满足两项核心要求：与原始数据具有统计相似性，以及具备结构合理性。除了再现边际分布和关联关系外，合成数据还必须遵循领域特定约束，例如合理的家庭构成、非负支出（如药物支出）、内部一致的人口统计特征（例如未成年人不能有成年子女），以及有效的事件序列（例如先获得硕士学位，再获得博士学位）。即使分布相似性很高，违反这些约束也会严重削弱合成数据集的可信度和可用性。

图1展示了生成合成数据并评估其适用性的一般工作流程。分析人员通常旨在：（i）通过拟合统计模型或机器学习模型来回答预先设定的研究问题；（ii）探索数据，以识别新的模式和关系。只有当合成数据在分布和结构上都近似于原始数据时，才能实现这些目标。

*通讯作者：matthias.templ@fhnw.ch
†西北瑞士应用科学与艺术大学（FHNW）商学院，瑞士奥尔滕

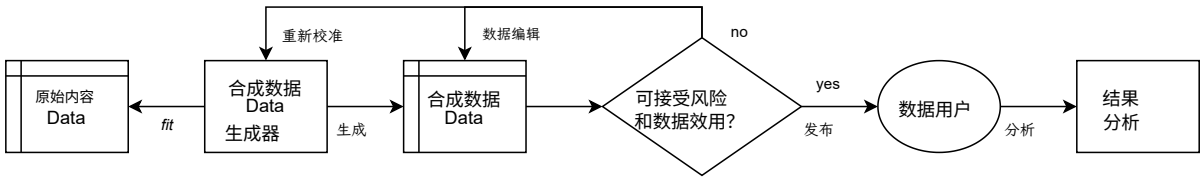

图1：合成数据生成与评估工作流程。使用原始数据训练合成数据生成器，以生成合成记录。随后对这些记录的隐私风险和统计效用进行评估。如果结果可接受，则将合成数据发布给数据用户进行分析；否则，重新校准生成器或编辑数据。

然而，仅有较高的分析效用并不能保证数据发布安全。从数据提供方的角度来看，提高效用往往需要保留变量之间的强依赖关系——而这类信息恰恰可能使对手能够推断敏感属性。因此，随着合成数据效用提升，披露风险未必会降低；事实上，风险甚至可能上升。在实践中，人们通常借助风险—效用（RU）图 [Duncan 等人， 2001] 以及近期提出的多变量扩展 [Thees 等人， 2025], 来概念化这种权衡，以便就分析价值与披露保护之间可接受的取舍作出明智决策。如何以与现代高效用合成数据相称的方式评估推断披露风险，仍是一项重要的方法学挑战。

1.2 合成数据生成方法概述

自 Rubin [1993] 将合成数据引入为从预测分布中模拟抽样所得的数据以来，各种生成方法相继涌现。条件建模 方法按顺序合成变量，每个变量都以此前生成的变量为条件；synthpop [Nowok 等人， 2016]和 simPop [Templ 等人， 2017] 等实现采用基于树的方法、回归或其他学习器。联合建模 方法（包括生成对抗网络和深度生成模型）试图同时捕捉完整的联合分布，但需要大型数据集、仔细调优和大量计算资源 [Mekonnen，2024年，Miletic 和 Sariyar， 2024], 且可能难以处理复杂的表格结构 [Thees 等，2024年，Ward 等人， 2025]。

关键在于，使合成数据具有分析用途的同等保真度，也保留了能够促成属性推断攻击的预测关系。高效用的合成数据会忠实再现协变量与结果之间的关联；攻击者可利用这些关联，根据准标识符推断敏感属性。效用与隐私之间的这种矛盾，正是我们提出该度量的动因。

尽管本文聚焦于合成数据，但推断性披露风险框架同样适用于传统的匿名化数据或任何已发布的数据集：给定已发布的数据，对手能够根据准标识符以多高的准确度推断敏感属性?

1.3 披露风险度量

经典的统计披露控制（SDC）区分：（i）身份披露（即将记录关联到特定个人）；（ii）属性披露（即正确获知敏感值）；以及（iii）成员披露（即获知某个人的数据是否属于某个数据集） [Hundepool 等人， 2012]。

对于完全合成的微观数据，身份披露风险通常较低 [Templ，2014年，El Emam 等人， 2020],因为合成记录并非对个人的直接呈现。因此，关注重点大多转向分析有效性和属性披露风险。

现有属性披露诊断。
现有属性披露诊断大致分为两类:

- 基于匹配的度量评估合成值与原始值完全匹配的频率 [Taub 等人，2018]。例如 DiSCO [Raab 等人， 2025], 该方法会将记录标记为存在风险，条件是原始数据中的某个准标识符组合在合成数据中出现，且对应的敏感值相同。这类方法透明，且能自然地融入经典的统计披露控制（SDC）概念，但可能会受到类别不平衡的影响，也不会直接对推断过程建模。

- 基于模型的度量会在合成数据上训练预测模型，并在原始数据上评估其表现。Hittmeir 等人 [2020] 提出将预测准确率与代表攻击者先验知识的基线进行比较（例如边际类别频率），从而量化合成数据发布带来的额外风险。这一基线比较原则——衡量攻击者利用发布数据后比不利用时能做得好多少——是本文所提出方法的核心。

攻击者场景。
对于基于模型的度量，威胁模型假设攻击者会利用发布的合成数据训练预测模型，并将其应用于准标识符已知的个人。图2展示了这一场景。如果仅使用合成数据训练的模型能够可靠地预测真实数据中的机密属性，那么此次合成数据发布就存在隐私风险 [Taub 等，2018，Barrientos 等人，2018，Kwatra 和 Torra， 2024]。

推断性披露。
此处提出的风险度量聚焦于推断性披露风险——也称为预测性披露[Willenborg 和 de Waal， 2001] ——当微观数据的发布使人们能够比不发布数据时更准确或更有把握地推断敏感属性时，就会发生这种情况 [Duncan 和 Lambert，1989年，Hundepool 等人， 2012]。这一概念建立在 Dalenius 的基础性定义之上：

> “如果发布统计量 S 使得能够比不访问 S时更准确地确定敏感属性] 的值 [，那么就发生了披露。” [Dalenius，1977年，第 432] 页

重要的是，推断性披露可能影响并不属于原始数据集的个人，并且可能涉及受影响个人甚至并不知晓的信息。例如，如果发布数据揭示了生活方式指标与疾病风险之间的强关联，攻击者就可能推断出数据集之外某个相似个人的风险升高。

尽管完全消除推断性披露需要破坏敏感变量与非敏感变量之间所有有意义的关系 [Dwork 和 Naor， 2010], 我们认为，在大数据和人工智能时代，推断风险值得更多关注。Mühlhoff [2021] 指出，当敏感信息在违背个人意愿的情况下被统计估计，且这些预测导致影响其福祉或自由的差别待遇时，预测性隐私即遭到侵犯。

记录级评估的必要性。
尽管现有的基于模型的度量通常提供聚合数据集级风险摘要，但我们认为，记录级评估对于有针对性的风险缓解至关重要。识别哪些具体个人面临升高的披露风险，使数据托管方能够采取选择性保护措施，而不是可能不必要地降低效用的统一措施。这促使我们提出 RAPID 度量，我们将在讨论差分隐私的作用之后介绍该度量。

1.4 差分隐私与属性推断

一个自然的问题是，差分隐私（DP）是否已经解决了我们旨在衡量的披露风险。我们认为，差分隐私和 RAPID 关注的是互补的问题。

差分隐私 [Dwork 等人， 2006] 提供了输出稳定性的形式化保证：当将单个个体的数据添加到数据集中或从中移除时，任何特定输出出现的概率变化幅度至多为乘法因子 $\exp(\varepsilon)$ [Domingo-Ferrer 等人， 2021]。这一保证独立于入侵者的背景知识，也不依赖于关于数据分布或攻击者能力的假设。

关键在于，差分隐私并未声称攻击者无法获知有关某个个体的敏感信息。相反，它保证无论某个特定个体的数据是否包含在数据集中，攻击者获得的信息增益都基本相同 [Muralidhar 和 Ruggles， 2024]。其依据是，如果某个个体是否被纳入数据集无关紧要，那么就不会发生针对该个体的隐私泄露。因此，差分隐私在防止成员推断攻击方面尤为有效；在这类攻击中，对手试图判断某个特定个体是否出现在训练数据中。

然而，属性推断攻击 利用的是另一种机制。攻击者不是问“这个人是否出现在数据集中？”，而是问“根据发布的数据以及我对这个人的了解，这个人的敏感属性是什么？”差分隐私不会直接限制这类推断的准确率，因为风险源于数据所保留的群体层面模式，而非任何个体是否参与其中。如果存在强相关性

存在于总体中——例如年龄、教育程度与疾病状态之间的相关性——那么在差分隐私（或任何其他匿名化方法）下发布的数据仍可能编码这些关系，从而使准确的属性推断成为可能[Blanco-Justicia etal., 2022,Muralidhar and Domingo-Ferrer，2023]。

这一局限并非 DP 所独有。传统匿名化方法也面临同样的根本性矛盾：要保留分析效用，就必须保留统计关系，但这些关系同样会促成推断攻击。无论采用何种匿名化技术，发布数据是否会提高攻击者推断敏感属性的能力——相较于仅凭背景知识所能进行的推断——这一问题始终适用。此外，DP 还面临一个实际困境：严格的隐私保证（较小的 ε）需要较高的噪声水平，可能使输出在分析上失去效用；而宽松的隐私保证（较大的 ε）则几乎无法提供有意义的保护 [Domingo-Ferrer 等人， 2021]。

因此，DP（以及更广义的匿名化方法）应辅以明确的、基于场景的披露风险评估，直接衡量属性推断脆弱性。正如 Domingo-Ferrer 等人所[2025] 指出的那样，对于合成数据或受 DP 保护的数据，经验性披露风险评估仍然不可避免，正如在传统效用优先方法下那样。RAPID 通过量化攻击者从发布数据中推断敏感属性的准确程度来满足这一需求，为形式化隐私保证提供了实用的补充。

与成员披露和差分隐私基准的关系。
近期关于 成员披露 的研究，通过比较两个仅在是否包含某个个体方面存在差异的匿名化数据集——一个包含该个体的“成员”版本和一个不包含该个体的“非成员”版本——并衡量纳入该个体后推断准确率的边际提升 [例如 Francis 和 Wagner， 2025]，将属性推断风险置于差分隐私框架中考察。这些方法旨在量化因纳入个体而导致的个人隐私损失，其依据是差分隐私的稳定性保证。

RAPID 探讨的是另一个问题。基于成员身份的评估关注个体层面的稳定性（“我的数据被纳入后，风险会发生变化吗？”），而 RAPID 关注总体层面的脆弱性（“攻击者有多大概率能够有把握地作出正确推断？”）。两种方法都有助于理解披露风险，但用途各不相同：以成员身份为重点的方法适用于认证差分隐私式保证，而 RAPID 则支持对公开使用的数据发布进行实用的风险—效用评估。

RAPID 是什么，不是什么。
为避免误解，我们明确说明 RAPID 能提供什么，以及它不声称提供什么。RAPID 是一种基于场景、符合现实攻击者情境的实证诊断工具，用于量化在指定预测模型下实际发生的推断脆弱性。它衡量攻击者在多大程度上能够有把握地根据发布数据推断敏感属性，前提是这些数据保留了相应的预测结构。RAPID 不是形式化隐私保证、差分隐私类比，也不是最坏情况界。相反，RAPID 通过提供透明且可操作的风险评估来补充形式化保证，支持作出知情的发布决策——回答一个实际问题：发布的数据是否会使攻击者能够实施数据托管方认为不可接受的推断攻击。

1.5 贡献

本文作出以下贡献。

1. **基于现实威胁模型提出一种新的披露风险度量。** 我们提出 RAPID（属性预测引发的披露风险），这是一种针对匿名化微观数据的属性推断披露风险度量，重点关注完全合成数据。RAPID 量化攻击者能够有把握地推断出敏感属性的记录所占比例。与差分隐私不同，差分隐私提供的是最坏情况界，衡量某个个体是否被纳入对发布数据的影响程度；RAPID 则衡量在整个发布的数据集中推断成功的频率，是一种实用的披露脆弱性概览，数据托管方可以直接理解和比较。

   我们的威胁模型假设，入侵者只能访问发布数据和目标个人的准标识符，但无法获得用于验证或校准的真实数据辅助样本。这反映了公开使用文件外部分析人员所面临的条件，也避免了基于留出的评估在实践中的局限性；这类评估的结论高度依赖留出集的规模和代表性 [Hittmeir 等人，2020年，Platzer 和 Reutterer， 2021]。评估风险时，数据保管人采用保守的最坏情况假设：攻击者知道原始数据集中所有个人的准标识符（但

不知道他们的敏感属性）。随后，RAPID 量化攻击者仅凭发布的合成数据成功推断这些属性的频率。这种最坏情况框架意味着，如果 RAPID 表明风险可接受，那么对于任何知道人数更少的现实攻击者而言，风险也都是可接受的。这些风险估计能否从原始样本推广到更广泛的总体，是一个代表性问题——这是所有实证分析都会面临的考量，并非 RAPID 特有的局限性。RAPID 直接将攻击者的目标转化为可操作的衡量方式：使用仅基于发布数据训练的模型，有把握地推断真实个人的敏感属性。这种基于预测的脆弱性对应于经典统计披露控制分类体系中的推断性披露 [Duncan 和 Lambert，1989年，Hundepool 等人，2012]。（见第2节。）

2. **分类属性的基线归一化置信度评分**。我们提出一种归一化增益度量，将攻击者对真实类别的预测概率与根据原始数据中类别流行率确定的基线进行比较（参见 Slokom 等人，2022年）。由此得到一种经过校准的超越偶然水平的置信度概念，能够明确考虑类别不平衡，避免在偏斜分布中夸大风险。RAPID 关注置信度调整后的正确性，而非原始准确率，从而捕捉发布数据在多大程度上支持对敏感属性进行有意义的推断。（参见第2.2节和第3节。）

3. **适用于分类和连续型敏感属性的统一框架**。RAPID 为离散和连续机密变量提供了一种一致的表述。对于分类属性，推断风险通过基线归一化预测置信度进行量化；对于连续属性，则通过基于容差的相对预测误差进行定义。该统一框架无需对连续结果进行离散化、引入任意分箱，也无需构建比较成员与非成员情景的反事实数据集，即可在不同变量类型之间一致地评估推断风险。（参见第2.2–2.3节和第3节。）

4. **基于阈值、便于政策解读的风险摘要**。RAPID 将披露风险概括为超过置信度或准确率阈值的记录占比，由此得出易于沟通、可调节以适应机构或监管风险容忍度的、具有明确含义且有界的度量指标。这种表述与统计披露控制中常用的风险—效用决策框架自然契合，并支持对不同合成器、参数设置和数据发布进行一致比较。（参见算法1和第 2.6 节。）

5. **支持诊断分析的记录级风险指标**。尽管 RAPID 报告的是总体披露度量，但其构建基础是逐条记录的风险指标。这种设计支持对推断脆弱性进行细致分析，包括识别高风险记录、亚组特定风险模式，以及对披露风险产生不成比例影响的准标识符组合。这些记录级信号有助于数据管理者了解风险成因，并采取有针对性的缓解策略。（参见第 5.5 节和图 5。）

6. **实证验证与稳健性分析**。通过模拟研究和真实数据示例，我们展示了 RAPID 如何随准标识符与敏感属性之间的依赖强度变化，阈值选择如何影响风险分类，以及结果如何在多种攻击者模型下保持稳定。我们还展示了在具备辅助真实数据时，如何将 RAPID 与基于留出数据的评估结合起来，从而支持内部基准测试，同时不改变核心威胁模型。（参见第 4 节和第 5 节。）

7. **开源实现**。 RAPID 以 R 语言实现，并与现有的合成数据生成流程集成。该实现支持多种攻击者模型、自助法置信区间和诊断输出。RAPID 关注总体层面脆弱性，而非个体隐私损失，因此可作为差分隐私等形式化隐私框架的补充，尤其适用于评估公开使用的微观数据发布。（参见第3.1节）。

1.6 动机示例

为了说明 RAPID 所应对的推断威胁，我们进行了一项简单的攻击实验。我们使用 `eusilc`公开使用文件——一个基于欧盟收入与生活条件统计（EU-SILC）、贴近现实的合成人口，其中包含复杂的家庭结构以及丰富的人口统计和

社会经济变量 [Alfons 等人， 2011] ——生成了一个完全合成数据集 `synthpop` [Nowok 等人， 2016]。随后，我们在合成数据上训练了随机森林模型 [Wright 和 Ziegler， 2017]，以预测两个敏感属性：分类变量 `marital_status` 和连续变量 `income`。

将这些在合成数据上训练的模型应用于真实协变量值，揭示了具有实际意义的披露风险。对于 `marital_status`，尽管分类器在训练期间从未接触过真实标签，但它在真实数据上的准确率仍达到 82%。对于 `income`，许多预测值都接近真实值，这表明攻击者能够在具有实际意义的范围内近似推断真实收入。

这项实验表明，准确率或均方误差等标准指标虽有助于评估预测性能，却无法充分刻画披露风险。这些指标忽略了攻击者对正确预测的**置信度**，容易受到类别不平衡的影响，也无法提供可由政策解读的阈值。这些局限促使我们在下一节正式提出 RAPID，为分类变量和连续敏感属性提供统一的框架。

2 RAPID 度量

本节将正式定义 RAPID（属性预测引发的披露风险）。回顾一下，我们的威胁模型假设攻击者在发布数据上训练预测模型，并将其应用于准标识符已知的个人（参见图2）。RAPID 用于量化攻击者能够对敏感属性作出有把握且正确的 推断的频率。

关键思路是将攻击者的预测表现与完全忽略准标识符的基线进行比较。对于分类属性，该基线是原始数据中各类别的边际频率——也就是说，攻击者仅凭类别流行率、而不使用任何准标识符信息，就能为正确类别赋予的概率。对于连续属性，该基线是参考预测误差。

只有当攻击者的预测表现显著超过该基线时，才会将记录标记为存在风险，从而确保 RAPID 捕捉到的是真正的信息泄露，而非类别不平衡或分布特性造成的假象。

2.1 设置与符号

将原始微观数据记为

$$\mathbf{Z} = [\mathbf{X}_Q,\ \mathbf{X}_U,\ \mathbf{y}] \quad ,$$

其中 $\mathbf{X}_Q \in \mathbb{R}^{n \times p_Q}$ 表示攻击者已知的准标识符，$\mathbf{X}_U \in \mathbb{R}^{n \times p_U}$ 表示攻击者无法获取的其他属性，而$\mathbf{y} \in \mathbb{R}^n$ 则是机密敏感属性。

发布的数据集（例如合成数据集）为

$$\mathbf{Z}^{(s)} = [\mathbf{X}_Q^{(s)},\ \mathbf{X}_U^{(s)},\ \mathbf{y}^{(s)}] \quad .$$

攻击者观察发布数据 $\mathbf{Z}^{(s)}$ 以及目标个人的准标识符 $\mathbf{X}_Q$ （例如来自外部来源），但无法访问 $\mathbf{X}_U$ 或 $\mathbf{y}$。

攻击者基于（$\mathbf{X}_Q^{(s)}$ ， $\mathbf{y}^{(s)}$）训练预测模型 $\mathcal{M}$ ，以获得参数估计值 $\hat{}\Theta^{(s)}$，然后将该模型应用于 $\mathbf{X}_Q$ ，生成对敏感属性的预测 $\hat{}\mathbf{y}$ 。

为简化符号表示，我们将 $\mathbf{Z} = [\mathbf{X},\ \mathbf{y}]$, 记作，其中 $\mathbf{X}$ 表示准标识符，因为 $\mathbf{X}_U$ 在风险估计中不起作用。默认情况下，我们评估原始数据中所有 n 条记录的风险。不过，RAPID 的记录级设计支持对任意目标集进行风险评估——无论是特定子群体、样本，还是原始数据中不存在但已知其准标识符的个人。

图2从数据保管人的视角展示了这一威胁模型。

2.2 RAPID 对分类敏感属性的处理

当 $\mathbf{y}$ 为分类变量时，我们根据模型赋予真实类标签的概率来定义风险。对于每条记录 i，攻击者的模型赋予概率

$$g_i = \Pr(\hat{y}_i = y_i \mid \mathbf{x}_i, \hat{\Theta}^{(s)})$$

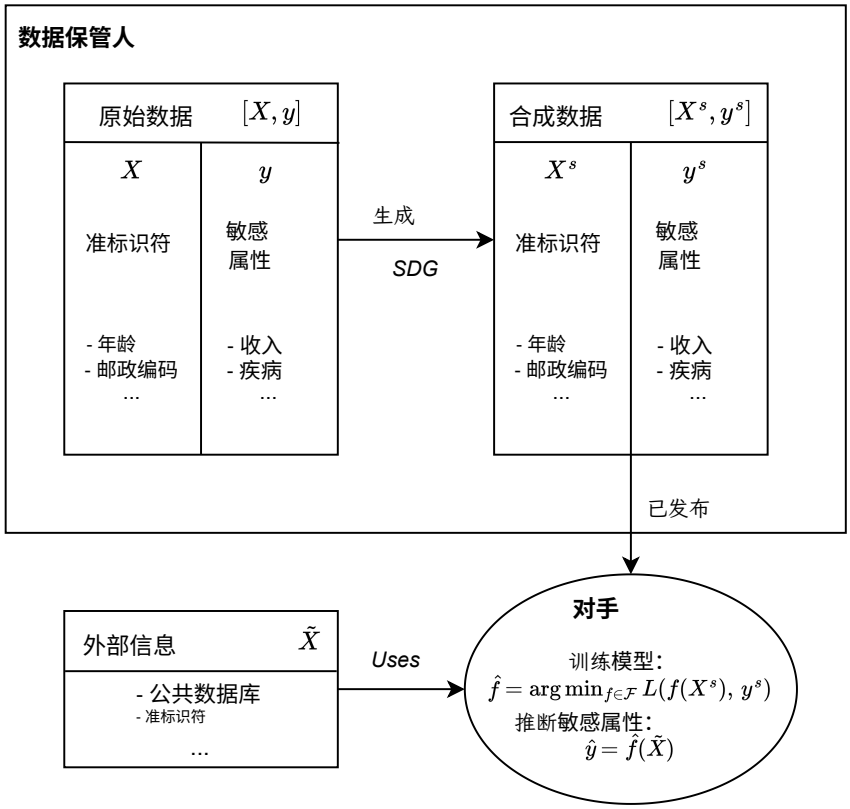

图2：推断式披露威胁模型。数据保管人发布由包含准标识符（$\mathbf{X}$）和敏感属性（$\mathbf{y}$）的原始数据生成的合成数据 $[\mathbf{X}^s$、$\mathbf{y}^s]$ 。能够访问发布的合成数据，并掌握个人准标识符外部知识（$\mathbf{X}$，其中可能包含也可能不包含原始样本中的个人）的对手，会训练预测模型来推断敏感属性。

属于真实类别 y_i。为评估该置信度是否异常偏高，我们将其与基线 b_i进行比较，该基线定义为原始数据中类别 y_i 的边际比例：

$$b_i = \frac{1}{n}\sum_{j=1}^{n}\mathbb{I}(y_j = y_i)$$

该基线表示仅根据边际类别分布进行猜测（不使用任何准标识符信息）所能达到的预测置信度。我们使用原始边际（而非合成边际），以确保基线有意义且一致：RAPID 衡量的是相对于真实总体比率的提升，而不是相对于可能存在失真的合成边际的提升。这样可使不同合成器的风险估计具有可比性，并避免将边际保持不佳与低披露风险混为一谈。

我们计算归一化增益分数，

$$r_i = \frac{g_i - b_i}{1 - b_i}$$

该分数衡量预测置信度相对于基线的提升，并以可能达到的最大提升（即达到完美置信度 $g_i = 1$）进行归一化。该分数满足：

- $r_i < 0$：模型表现低于基线
- $r_i = 0$：模型表现与基线持平（无信息增益）
- $0 < r_i < 1$：准标识符带来的部分信息增益
- $r_i = 1$：完美预测（完全披露）

若 $r_i > \tau$，则认为记录存在风险，其中 $\tau \in (0, 1)$ 是策略定义阈值。我们建议将$\tau = 0.3$ 设为默认值，这意味着当攻击者达到至少相当于基线之上最大可能提升幅度的 30% 时，就会将该记录标记出来。此阈值在敏感度（检测有意义的推断增益）和特异度（避免因微小提升而产生误报）之间取得平衡；第 5 节提供了有关阈值选择的实证指导。

分类RAPID指标如下：

$$\mathrm{RAPID}^{\mathrm{cat}}(\tau) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(r_i > \tau)$$

它表示发布数据使推断效果显著优于基线比率的记录所占比例。

2.3 连续敏感属性的 RAPID

当机密属性 $\mathbf{y}$ 为连续变量时，我们评估模型预测值 $\hat{}y_i$ 是否足够接近真实值 y_i。具体而言，我们计算预测误差 e_i，并将其与固定的全局阈值ε进行比较。若记录 i 满足 at-risk$_i = \mathbb{I}(e_i < \varepsilon)$，则视为存在风险，其中 ε 是固定的容差水平，用于反映预测值必须与真实值接近到何种程度才会构成披露风险。根据应用场景，可以采用多种误差指标。表 1 总结了适用的选择。重要的是，所有提出的误差度量都是无量纲且可比较的，因此可以对所有记录和变量一致地应用单一阈值 ε。

在所有情况下，连续型 RAPID 指标为：

$$\mathrm{RAPID}^{\mathrm{cont}}(\varepsilon) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(e_i < \varepsilon)$$

表示预测值落在指定误差容限范围内的记录比例。请注意，虽然汇总误差指标（MAE、RMSE）不足以用于披露评估——因为它们无法识别哪些记录存在风险——但它们可以通过描述典型的预测准确率，为选择 ε 提供参考。

表 1：连续型敏感属性的误差指标。如果 $e_i < \varepsilon$，则记录会被标记为存在风险，其中 ε 是用户指定的阈值。平滑常数 $\delta > 0$ 可防止除以零（默认值： $\delta = 0.01$，或根据数据范围进行缩放）。

误差指标 e_i	**描述、用法和示例**
对称相对误差（默认） $\frac{2\lvert y_i - \hat{y}_i\rvert}{\lvert y_i\rvert + \lvert\hat{y}_i\rvert + 2\delta}$	对尺度具有稳健性，在误差计算中同时考虑真实值和预测值。 示例：对于 $\varepsilon = 0.05$ (5%)的收入数据，一条记录的真实收入为 $50,000，预测收入为 $52,000，其误差为 $\approx 3.9\%$，因此被标记为存在风险。 适用情况：基于百分比的误差度量的默认选项——法。
稳定相对误差 $\frac{\lvert y_i - \hat{y}_i\rvert}{\lvert y_i\rvert + \delta}$	公式比对称型更简单。对于大多数数值，结果相近，但对于非常小的真实值，误差更大。 示例：对于 $\varepsilon = 0.10$ (10%) 的收入数据，一条记录的真实收入为 $500，预测收入为 $450，误差为 $\approx 10\%$ 并且存在风险。 适用场景：在某些领域中较为传统，或偏好使用更简单公式时适用。
绝对误差 $\lvert y_i - \hat{y}_i\rvert$	没有百分比解释；误差以原始单位衡量的 y。阈值 ε必须使用相同的尺度。 示例：对于 $\varepsilon = 2$ 年的年龄数据，真实年龄为65岁、预测年龄为66岁的记录，其误差为1岁，因此被标记为存在风险。 适用情形：有意义的阈值是“在 X 个单位以内”，而不是“在 X% 以内”（例如，年龄相差 2 年以内、温度相差 1°C 以内、日期相差 7 天以内）。

2.4 计算

算法1将 RAPID 评估协议形式化：在发布数据上训练预测模型，将其应用于目标准标识符，计算逐记录风险评分，并汇总得出 RAPID。

算法1 RAPID 评估协议

要求： 原始数据 $\mathbf{Z} = [\mathbf{X}, \mathbf{y}]$；发布数据 $\mathbf{Z}^{(s)}$
要求： 阈值 $\tau \in (0,1)$（分类变量；默认值 $\tau = 0.3$）和 $\varepsilon > 0$（连续变量；默认值 $\varepsilon = 10\%$）
在 $\mathbf{Z}^{(s)}$ 上拟合预测模型 $\mathcal{M}$，得到 $\hat{\Theta}^{(s)}$
计算预测 $\hat{\mathbf{y}} = \mathcal{M}(\mathbf{X}; \hat{\Theta}^{(s)})$
对于 $i = 1, \ldots, n$ **do**
    如果 y 是分类变量**则**
        $g_i \leftarrow \Pr(\hat{y}_i = y_i \mid \mathbf{x}_i)$
        $b_i \leftarrow n^{-1} \sum_{j=1}^{n} \mathbb{I}(y_j = y_i)$
        $r_i \leftarrow (g_i - b_i)/(1 - b_i)$
        $I_i \leftarrow \mathbb{I}(r_i > \tau)$
    else
        $e_i \leftarrow 2|y_i - \hat{y}_i|/(|y_i| + |\hat{y}_i| + 2\delta)$
        $I_i \leftarrow \mathbb{I}(e_i < \varepsilon)$
输出： 记录级风险 $\{r_i\}$或 $\{e_i\}$、指标$\{I_i\}$，以及

$$\mathrm{RAPID} \leftarrow \frac{1}{n} \sum_{i=1}^{n} I_i$$

因此，RAPID 是一个经验估计值，表示在指定的攻击者模型下，随机抽取的一条记录遭受成功属性推断的概率。

评估多个模型 $\mathcal{M} \in \mathcal{S}$ 时，同时报告平均值和保守包络：

$$\overline{\mathrm{RAPID}}(\cdot) = \frac{1}{|\mathcal{S}|} \sum_{m \in \mathcal{S}} \mathrm{RAPID}^{(m)}(\cdot), \qquad \mathrm{RAPID}_{\max}(\cdot) = \max_{m \in \mathcal{S}} \mathrm{RAPID}^{(m)}(\cdot)$$

2.5 实施注意事项

RAPID 适用于任何能够生成类别概率或点预测的算法。为体现强大的攻击者，我们建议评估能力较强的学习器（例如随机森林、梯度提升），并报告观测到的最大风险。当有多个已发布数据集可用时，可对风险取平均值，也可报告最坏情况下的数值，具体取决于所需的保守程度。阈值 τ 和 ε 应体现政策要求；我们建议默认采用 $\tau = 0.3$ 和 $\varepsilon = 10\%$，并在第 5 节通过实证进行验证。若要根据数据进行选择，可以对 $\mathbf{y}$ 进行置换，重新计算 RAPID，并选择一个阈值，使观测到的风险高于所得零分布的第 95 百分位数。可通过自助重抽样量化不确定性；或者将 RAPID 视为二项比例，使用 Wilson 评分法 [Wilson, 1927] 或 Clopper–Pearson 置信区间 [Clopper 和 Pearson, 1934]。

评估合成数据生成器。
为评估合成数据生成器（SDG）的总体披露风险，我们建议采用 k折交叉验证：将原始数据划分为 k 折（若为分类变量，则按$\mathbf{y}$ 分层），由 k-1折生成合成数据，并在留出折上评估RAPID。对各折结果进行汇总，可得到期望风险 $\mathbb{E}$[RAPID] 及其置信区间，从而对SDG的平均披露风险作出稳健评估。该方法通过 `rapid_synthesizer_cv()` 函数实现，尤其适用于在不同合成方法之间进行选择（例如，分类与回归树与参数化方法）或优化超参数。选定方法后，应对最终合成数据产品应用标准RAPID，以便在数据发布前获得记录级风险评估。

计算复杂度。
RAPID的计算成本主要由模型训练步骤（算法1第1行）决定，后续预测与风险评估仅需 $O(n)$次运算。对于随机森林，训练复杂度为 $O(n \cdot p \cdot T \cdot \log n)$，其中 n 为记录数， p 为准标识符数， T 为树的数量；预测复杂度为 $O(n \cdot T \cdot d)$，其中 d 为树深度。梯度提升机具有类似的$O(n \cdot p \cdot T)$训练复杂度。线性模型与逻辑回归通过正规方程或迭代拟合需要 $O(n \cdot p^2)$次运算，而CART树的复杂度为 $O(n \cdot p \cdot \log n)$。

在实践中，对于包含 10,000 条记录的数据集，RAPID 评估可在数秒内完成；采用默认随机森林设置（ $T = 500$ 棵树）时，对于 $n > 100{,}000$ 条记录的数据集，评估仍然可行。内存使用主要由拟合后的模型对象决定；在高维数据上使用树数量较多的随机森林，可能需要数百兆字节。当评估多个攻击者模型或自助法重复样本时，各模型间的计算可轻松并行，从而高效利用多核系统。

2.6 性质

RAPID 具有若干可直接从其定义推导出的结构性质。首先，它关注成功且置信度高的属性推断事件，而非整体预测准确率，因此其风险度量与经典统计披露控制分类体系中的推断性披露相一致。其次，按经验基准率进行归一化，确保分类风险得分不受类别不平衡影响，避免常见结果不当地抬高披露风险。

第三，RAPID 的取值有界，因为它定义为记录级披露指示变量的经验均值。这有助于比较不同数据集、合成器和攻击者模型。第四，该方法在攻击者模型方面具有灵活性：任何能够生成点预测或类别概率的预测模型均可使用，因此可以在强攻击者假设和自适应攻击者假设下评估 RAPID。

最后，RAPID 分别关于决策阈值 τ 和 ε 单调变化。提高分类阈值 τ 或降低连续容差 ε，都只会减少被判定为存在风险的记录数量。这种单调性有助于开展透明的敏感性分析，也使 RAPID 非常适用于实际的披露控制。

由于 RAPID 在记录级别运作，因此还可以通过将指标平均值限定在准标识符空间的子集内，开展针对特定子群体的条件风险分析（例如参见图 5）。

3 玩具示例

我们通过一个简单示例展示分类RAPID指标。假设有一个包含 100 条记录的数据集，其中类别 `healthy` 的流行率（边际比例）为 60%。将基于发布数据训练的模型应用于三条原始记录，这些记录实际都属于类别 `healthy`：

- 记录 1： $g_1 = \Pr(\hat{y}_1 = \text{健康} \mid \mathbf{x}_1) = 0.70$
- 记录 2： $g_2 = \Pr(\hat{y}_2 = \text{健康} \mid \mathbf{x}_2) = 0.85$
- 记录3： $g_3 = \Pr(\hat{y}_3 = \text{健康} \mid \mathbf{x}_3) = 0.55$

基线是原始数据中类别 `healthy` 的边际频率：

$$b_i = 0.60 \quad \text{for all three records (same class)}$$

我们计算每条记录的归一化增益：

$$r_1 = \frac{0.70 - 0.60}{1 - 0.60} = \frac{0.10}{0.40} = 0.25$$
$$r_2 = \frac{0.85 - 0.60}{1 - 0.60} = \frac{0.25}{0.40} = 0.625$$
$$r_3 = \frac{0.55 - 0.60}{1 - 0.60} = \frac{-0.05}{0.40} = -0.125$$

解释：

- 记录 1：较基线最高提升幅度的 25%
- 记录 2：达到最大改进幅度的 62.5%（置信度高！）
- 记录3：低于基线（模型表现不如猜测）

使用默认阈值 $\tau = 0.3$，我们会标记满足以下条件的记录： $r_i > 0.3$

$$\mathbb{I}\{r_1 > 0.3\} = 0, \quad \mathbb{I}\{r_2 > 0.3\} = 1, \quad \mathbb{I}\{r_3 > 0.3\} = 0$$

RAPID 指标为：

$$\text{RAPID}^{\text{cat}}(0.3) = \frac{1}{3}(0 + 1 + 0) = 0.33$$

这表明，33% 的记录所对应的预测明显优于基线比率。记录 2 的归一化增益为 62.5%，表明其披露风险较高：相较于仅凭边际类别频率所能实现的推断，发布数据支持更有把握的推断。此示例说明，RAPID 能识别准标识符提供了有意义的信息增益的记录，而不只是衡量预测准确率。

示例（连续属性）：

现在考虑敏感变量为连续变量的情况，例如收入。假设攻击者在发布数据上训练回归模型，并将其应用于三个人的原始协变量。真实收入和模型的预测值如下：

- 记录 1： $y_1 = 50,$ 000, $\hat{y}_1 = 47,$ 000
- 记录 2： $y_2 = 35,$ 000, $\hat{y}_2 = 39,$ 000
- 记录3： $y_3 = 80,$ 000, $\hat{y}_3 = 90,$ 000

The 稳定后的相对预测误差（作为真实值的百分比（假设 $\delta = 0$）为 e:

$$e_1 = \frac{|50{,}000 - 47{,}000|}{50{,}000} \times 100 = 6\%$$
$$e_2 = \frac{|35{,}000 - 39{,}000|}{35{,}000} \times 100 = 11.4\%$$
$$e_3 = \frac{|80{,}000 - 90{,}000|}{80{,}000} \times 100 = 12.5\%$$

使用一个 th 阈值为 $\varepsilon = 10\%$ 时，我们会标记预测值与真实值相差不超过 10% 的记录 :

$$\mathbb{I}\{e_1 < 10\%\} = 1, \quad \mathbb{I}\{e_2 < 10\%\} = 0, \quad \mathbb{I}\{e_3 < 10\%\} = 0$$

连续型 RAPID 指标为：

$$\text{RAPID}^{\text{cont}}(10\%) = \frac{1}{3}(1 + 0 + 0) = 0.33$$

这表明，对于三分之一的个人，攻击者的模型预测收入的相对误差在 10% 以内。记录 1 的误差仅为 6%，这构成了披露风险：发布数据使攻击者能够高精度地推断收入。这种属性披露与风险评估直接相关，但仅靠平均绝对误差等聚合指标无法反映这种情况。

3.1 软件与默认设置

所提出的风险度量已在 R 包 `RAPID` 中实现，该程序包可在 GitHub [上公开获取 2026]Thees 等人，2026]。该程序包基于成熟的程序库构建：用于随机森林的 `ranger` [Wright 和 Ziegler， 2017] ，它支持原生并行化和内存高效存储；用于逻辑回归的 `glm` ；以及用于梯度提升的 `xgboost` [Chen 和 Guestrin， 2016] 。该程序包提供了用于计算分类和连续敏感属性的 RAPID 的函数，包括基线归一化、阈值校准和基于自助法的不确定性量化。

除非另有说明，我们在实验中采用以下默认设置：攻击者模型 $\mathcal{M}$ 为一棵包含 500 棵树且启用概率输出的随机森林。对于分类属性，我们使用默认阈值 $\tau = 0.3$；对于连续属性，我们使用默认相对误差容限 $\varepsilon = 0.10$。我们通过对原始数据集进行非参数自助抽样（500 次重复）来量化不确定性，并计算基于百分位数的置信区间。

4 真实数据示例：UCI 成人数据集（人口普查收入）

我们以 $Adult$ 数据集的训练部分为例说明工作流程（UCI 机器学习知识库；32,561 行，15 个属性；二元机密属性 $\mathbf{y} = \mathbb{I}$（收入 $>$ \$50K）） [Becker 和 Kohavi， 1996]。协变量 $\mathbf{X}$包括年龄、教育程度、每周工作时长、婚姻状况等。我们将$\mathbf{y}$ 视为机密的敏感变量，将$\mathbf{X}$ 视为潜在已知的准标识符。

预处理。
我们移除了人口普查抽样权重（`fnlwgt`）和冗余的分类变量教育变量（保留`education.num`），并将字符变量转换为因子。

合成器。
我们考虑一种基于 CART 的表格合成器（如 `synthpop` [Nowok 等人， 2016] 中所实现），该合成器仅使用$\mathbf{Z}$ 进行训练，并生成 $M = 5$ 个合成重复样本。

攻击模型。
我们评估一个攻击者套件 $\mathcal{S}$ ={RF、GBM 和 ℓ_1-logistic}，在每个合成重复样本上训练，并在真实协变量 $\mathbf{X}$上评分。

指标与报告。
我们在一系列阈值下计算 RAPID$^{\text{cat}}(\tau)$，并将结果绘制为阈值曲线（图 3）。对于每个合成重复样本，我们对 $M = 5$ 个重复样本的 RAPID 取平均值，并报告 95% 自助法置信区间。此外，我们按真实类别 y 对 RAPID 进行分层，以揭示不同结果类别的披露风险是否存在差异——例如，高收入个人是否比低收入个人更容易被识别。

敏感性与诊断。
为了解 RAPID 对阈值选择的响应情况，我们绘制 RAPID$^{\text{cat}}(\tau)$ 在一组 τ 取值上的变化（图 3），以直观展示随着施加更严格的归一化增益阈值，披露风险如何降低。

实践者可以考虑其他诊断指标：

- **类别平衡：** 报告每个类别的基线概率 b_k ， k 可以揭示类别流行率对归一化增益的影响。

- **效用—风险联合视图：** 为披露风险提供背景信息，可以将 RAPID 与效用指标一并报告，例如在$\mathbf{Z}_{(s)}$ 上训练、但在$\mathbf{Z}$ 上评估的模型的预测准确率。

为评估属性推断风险，我们将 RAPID 应用于 UCI Adult 数据集的训练部分，并使用二元变量 `income` 作为敏感属性。我们通过 CART 合成器（`synthpop` 软件包）生成了五个合成数据集。对于每个重复样本，我们在合成数据上训练随机森林攻击者模型，以根据准标识符推断敏感属性，然后在原始数据集上评估模型的预测结果。

我们在一系列阈值 $\tau \in [0, 1]$ 下计算 RAPID，以考察披露风险如何随阈值严格程度变化。为量化不确定性，我们从原始数据集中进行 $R = 1000$ 次非参数自助法重抽样，从而得出稳健的基于百分位数的置信区间。RAPID 分数根据随机森林预测计算，反映了强大的攻击者模型。

图 3 展示了 RAPID 如何随归一化增益阈值 τ 变化。随着 τ 从 0 增加到 1，超过该阈值的记录逐渐减少，体现了第 2.6 节所讨论的单调性。该阈值曲线使数据提供方能够根据其隐私要求校准披露风险。

表 2 报告了五个合成重复样本在默认阈值 $\tau = 0.3$ 下的 RAPID 值。通过使用 $R = 500$ 个重抽样样本进行非参数自助法获得的窄置信区间表明结果稳定

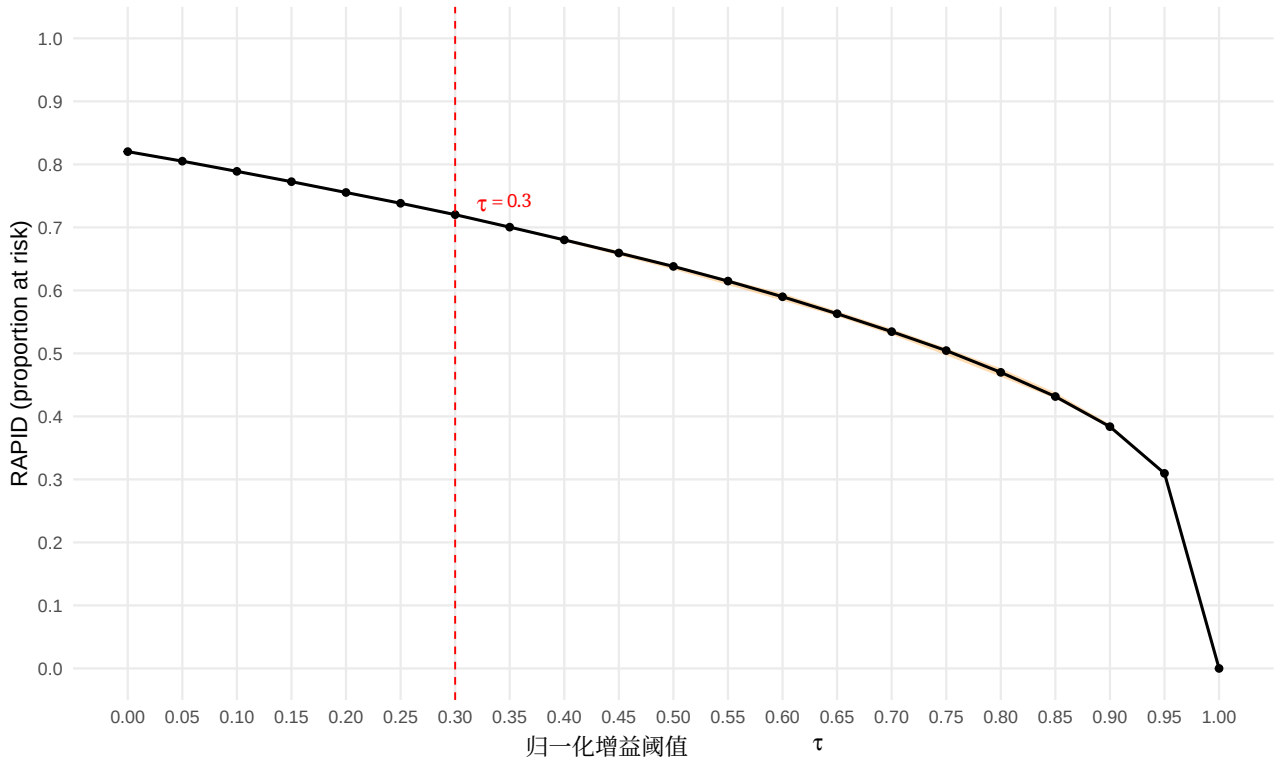

图 3：UCI Adult 数据集的阈值敏感性曲线。RAPID（风险记录数占比）随归一化增益阈值 τ变化的曲线，取通过 synthpop [采用 CART 合成生成的 5 个合成重复样本的平均值（Raab 等人， 2024]）。竖直虚线表示默认阈值（$\tau = 0.3$）。曲线显示，随着阈值设定得更严格，披露风险随之降低：在 $\tau = 0.3$ 时，约有 70% 的记录被标记，而在 $\tau = 0.9$时，不到 5% 的记录被标记。阴影区域表示不同重复样本之间的取值范围。

各合成重复样本的风险估计。约 72% 的记录的归一化增益超过阈值，这表明对于该数据集，使用 CART 合成器会带来较高的属性推断风险。

表 2：UCI Adult 数据集在默认阈值 $\tau = 0.3$ 下的 RAPID 估计值（$n = 32,561$）。每行对应一个通过 CART 生成的合成重复样本。置信区间（95%）通过非参数自助法获得（$R = 500$）。

重复样本	RAPID	95% 置信区间（下限）	95% 置信区间（上限）	n面临风险
1	0.719	0.715	0.725	23,453
2	0.720	0.714	0.723	23,454
3	0.722	0.717	0.727	23,513
4	0.720	0.715	0.725	23,447
5	0.726	0.720	0.730	23,645

72% 的属性推断风险算“太高”吗？答案取决于具体情境。成人数据集呈现出很强的预测结构：教育程度、职业、工作时长及其他准标识符能够很好地预测收入类别。任何为了保留分析效用而保留这种结构的合成器，也必然会保留使推断攻击成为可能的关系。协变量—结果关系较弱的数据集通常会得到较低的 RAPID 值。某一风险等级是否可接受，取决于该属性的敏感程度、预期应用场景以及机构的风险承受度。发现 RAPID 值较高时，数据保管人可以考虑采取几种应对措施：(i) 改用隐私保证更强的合成器；(ii) 抑制或粗化那些会影响预测准确率的准标识符（诊断工具见第 5.5 节）；(iii) 限制对合成数据的访问；或 (iv) 对于敏感属性并非特别机密的应用场景，接受这一风险。RAPID 不会规定普遍适用的可接受阈值；相反，它为知情且因情境而异的发布决策提供了实证依据。

5 模拟研究

5.1 概述与设计

为验证 RAPID 并研究影响属性推断风险的因子，我们开展了三项模拟研究:

1. **依赖强度：** 准标识符（QIs）与敏感属性之间关系的强度如何影响披露风险的尺度？（$\kappa \in [0, 100]$）

2. **阈值敏感性：** τ 如何影响不同依赖机制下的风险？（5 κ个水平 $\times$ 19 τ 个值）

3. **准标识符归因：** 哪些准标识符会导致风险？（基于回归的分析）

所有模拟均采用包含六个变量（性别、年龄、教育程度、收入、健康评分和疾病状态）的合成健康微观数据。我们通过全局参数 $\kappa \geq 0$ 控制依赖强度，详细信息见 A。

5.2 数据生成过程

我们模拟 $n = 1,000$ 条包含六个变量的独立记录：性别（G）、年龄（A）、教育程度（E）、收入（I）、健康评分（H）和疾病状态（D），其中 D 作为敏感属性。该设计通过潜在社会经济地位（SES）变量编码现实中的依赖关系，依赖强度由参数 $\kappa \geq 0$ 控制。

信号与噪声权重由 κ 得出:

$$w_{\text{signal}} = \sqrt{\frac{\kappa}{1+\kappa}}, \quad w_{\text{noise}} = \sqrt{\frac{1}{1+\kappa}}.$$

在 $\kappa = 0$ 时，关系完全由噪声驱动（$w_{\text{signal}} = 0$）；在 $\kappa \gg 1$ 时，依赖关系趋近于确定性强度（$w_{\text{signal}} \to 1$）。年龄服从截断正态分布；教育程度是由潜变量推导出的有序变量；收入采用对数线性形式；健康评分使用 Sigmoid变换；疾病状态通过多项逻辑回归生成，其系数按 κ缩放；性别为二元变量，且轻微依赖于社会经济地位。完整的数学规范见附录 A。

5.3 依赖强度

我们将依赖参数 κ 从 0 变化至 100（共 101 个取值，每个取值重复 10 次），以考察属性推断风险如何随准标识符与敏感属性之间的关系强度变化。图 4 的子图 (a) 显示了一条 S 形轨迹：从弱依赖过渡到中等依赖时，风险迅速上升；随着关系趋近确定性强度，风险则逐渐饱和。饱和水平接近 0.97 而非 1.0，反映了分类与回归树合成器采样所固有的噪声。攻击者准确率也呈现出类似的变化轨迹（从 0.70 升至 0.98），表明 RAPID 能够可靠地反映实际预测表现：RAPID 值较高的数据集确实面临更高的披露风险。

在确认风险会随依赖强度变化后，我们接下来考察阈值参数 τ 如何与这种依赖结构相互作用。模拟 2 通过研究整个依赖范围内的阈值敏感性来应对这一问题。

5.4 阈值敏感性

我们考察了归一化增益阈值 τ对 RAPID 的影响，选取五种依赖水平（$\kappa \in \{0,\ 5,\ 10,\ 20,\ 50\}$），并将 τ 从 0.05 以 0.05 为增量变化至 0.95（每种组合重复 10 次）。图 4 的子图 (b) 揭示了曲线几何形态的定性变化：在低依赖水平下（$\kappa = 0$，灰色），RAPID 呈凸形下降，从 0.50 迅速降至 $\tau = 0.60$附近的零值，表明攻击者置信度分布较分散。在高依赖水平下（$\kappa \geq 5$，蓝色/红色），曲线则保持凹形，直到阈值严格到 $\tau > 0.70$时仍处于较高水平，反映出置信度分布集中在接近确定性的区域。这种由凸转凹的变化标志着对抗性推断能力从弱转强。

实际上，τ 的选择应由属性的敏感程度决定，而非标记记录的比例：对于高度敏感的属性，只有近乎确定的推断才构成不可接受的风险，因此适合采用更严格的阈值（$\tau \geq 0.70$）；而较低的阈值（$\tau \approx 0.30$–0.40）则能提供更广泛的保护，因为中等置信度的推断也可能引发披露方面的顾虑。阈值与依赖关系的交互作用表明，相同的 τ 会因数据结构不同而对应不同程度的对手确定性，因此需要针对具体应用场景进行校准。

5.5 准标识符归因

图 5 展示了不同人口统计特征组合对应的预测对数优势。分析揭示，教育程度与属性推断风险之间存在很强的关联，并且存在显著的交互效应——尤其是年龄与教育程度之间的交互作用——这反映了模拟设计所产生的强准标识符依赖关系（$\kappa = 10$）。

除这些具体模式外，该示例还展示了 RAPID 的方法学优势：逐行风险标记可通过标准回归建模，将推断脆弱性归因于特定的准标识符组合。在实际应用中，此类分析有助于识别需要定向披露控制的高风险人口统计群体，检验有关脆弱性驱动因素的假设（例如，受教育程度是否会调节年龄相关风险？），并量化效应量，以便确定缓解措施的优先次序。

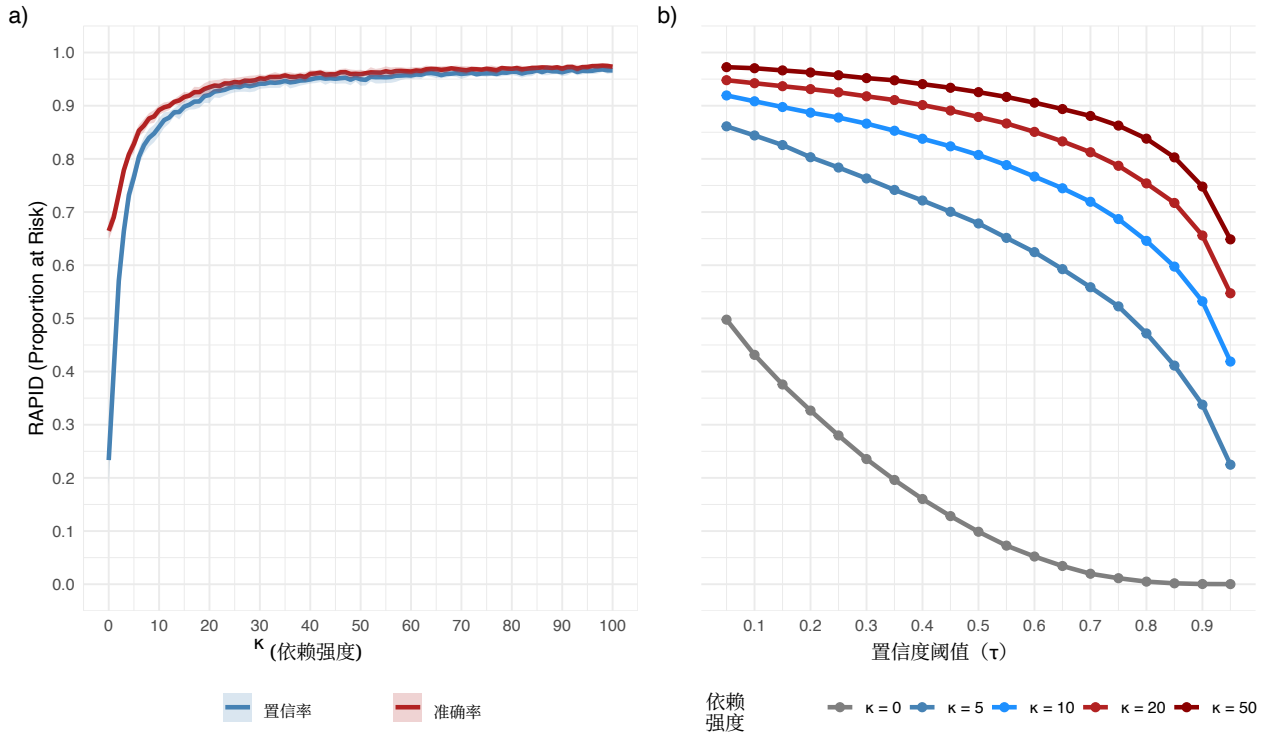

图 4：依赖强度和归一化增益阈值对 RAPID 的影响：(a) RAPID 和攻击者准确率随依赖强度单调上升 κ。RAPID 从 $\kappa = 0$ 时的 0.25 上升至 $\kappa = 100$时的 0.97，在 κ 值较低时增幅最大（$\tau = 0.3$）。这种 S 形增长表明，当准标识符与敏感属性关系从弱转为中等时，属性推断风险会迅速上升；随着依赖关系趋近于确定性水平，风险随后趋于饱和。(b) 不同 κ下，RAPID 随归一化增益阈值 τ 的变化。在依赖较低时（$\kappa = 0$，灰色），曲线呈凸形，反映出攻击者置信度较为分散，在中等阈值下大多数记录都会被过滤掉。在依赖较高时（$\kappa \geq 5$，蓝色/红色），曲线变为凹形，在应用严格阈值（$\tau > 0.7$）之前，数值始终维持在较高水平。这一转变反映了随着依赖关系增强，攻击者置信度分布发生了定性变化。两个面板均显示：10 次模拟的均值 $\pm$ 1 标准差；$n = 1000$ 条记录，分类与回归树合成器，随机森林攻击者。

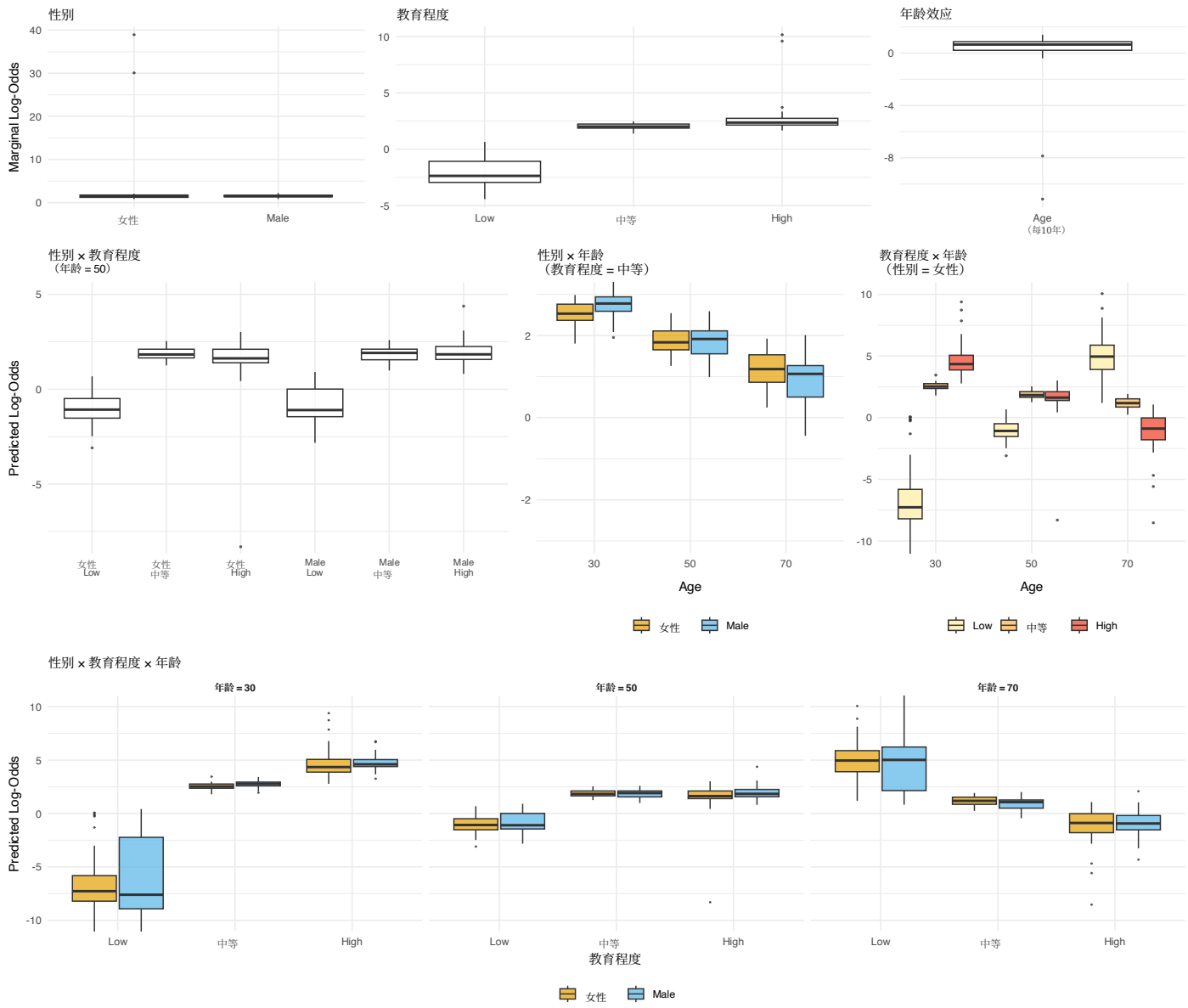

图 5：准标识符归因分析。基于逻辑回归，在 50 次模拟中（$\kappa = 10$、 $\tau = 0.3$、 $n = 1000$）属性推断风险的预测对数优势。**顶部：** 边际效应。**中部：** 在指定条件值下的二阶交互作用。**底部：** 三个年龄水平（30、50、70）下的三阶交互作用。各 Y 轴均进行了单独截取，以增强可见性；未显示部分极端离群值。

6 讨论

RAPID 提供了一种基于模型且符合现实攻击者情境的方法，用于评估发布数据中的属性推断风险。本节将讨论 RAPID 与现有披露风险度量的关系，指出其局限性，并提供实践指南。

6.1 与现有披露度量的关系

- **基于匹配的评估**。RAPID 与 DCAP [Taub 等人 2018] 和 DiSCO [Raab 等人 2025], 等度量一样，旨在量化属性推断风险，但其方法有着根本区别。DiSCO 通过检查列联表来评估风险：当准标识符单元格中的合成数据呈现一致的敏感属性值时，相关记录便会被标记为存在披露风险，这表明知道准标识符的攻击者可能会以很高的置信度推断出敏感属性。DCAP 同样评估攻击者预测敏感属性的准确程度，但它通过攻击者的预期预测准确率来量化风险；该准确率根据合成数据中的条件目标分布计算，而非通过显式分类器得出。

  RAPID 在三个关键方面不同于这些方法。首先，它采用预测建模，而非列联表分析，因此能够自然地处理混合类型数据，无需对连续变量进行离散化，也无需显式枚举准标识符单元格。其次，RAPID 基于阈值的风险准则可进行调整：对于分类敏感属性，数据管理者可校准归一化增益阈值 τ，以反映属性敏感程度和可接受的披露风险；对于连续属性，阈值 ε 则规定了可接受的相对预测误差百分比。第三，通过将基线可预测性直接纳入记录级归一化增益准则，RAPID 确保被标记的记录体现出超出仅凭边际分布所能实现的真实推断优势，从而解决类别不平衡问题：在小单元格中，稀有结果可能仅仅因为基线发生率低，就显得“预测得毫无差错”。

  这些设计选择体现了 RAPID 对对抗性现实性的关注：它模拟具备机器学习能力的复杂攻击者可能达成的目标，而非依赖组合唯一性准则。然而，这种方法也带来了显著局限。由于依赖机器学习模型，RAPID 的风险评估不透明且难以审计：与基于表格的方法不同，数据主体可以通过后者核实其准标识符单元格中的属性值是否一致，而 RAPID 的预测则来自复杂的集成模型。这种不透明性使利益相关方沟通和法规合规变得复杂，尤其是在数据主体依据隐私法规行使解释权时。此外，RAPID 的风险估计取决于建模选择——攻击者模型类型、超参数和阈值参数（τ 分类变量，ε 连续属性）——这些选择带来了需要论证、却缺乏既定指导方针的自由度。阈值 τ 和 ϵ虽具灵活性，却也具有任意性：除了领域专业知识和敏感性判断之外，没有原则性的方法来选择 τ 或 ε；相同的阈值在不同数据结构下对应的对抗确定性水平可能大相径庭。为探究风险估计如何随阈值选择而变化，将`rapid::plot()` 方法应用于 `rapid::rapid()` 结果，可生成敏感性曲线，展示在不同 τ（分类变量）或 ε（连续变量）取值下被标记为有风险的记录比例。使用示例见 B。

  RAPID 适用于准标识符空间复杂、数据为混合类型，或以模拟真实攻击者能力为重的场景。当透明度以及与经典统计披露控制框架的一致性是首要考虑因素时，或者法规或机构要求需要可解释的风险评估时，则更适合采用其他基于表格的度量方法。对于关键应用，同时使用这两种方法可能有助于从互补角度评估披露风险。

- **基于留出的评估**。基于留出的框架通过最近记录距离（DCR）评估隐私，衡量合成记录与训练记录的距离是否比与留出记录的距离更近。DCR 有助于检测记忆化，但无法量化属性推断风险（例如，Yao 等人，2026）：如果合成器保留了预测关系，却没有复制记录，那么即使 DCR 值看起来“安全”，属性推断风险仍可能很高。

  RAPID通过在发布数据上训练并在原始协变量上评分，直接应对属性推断。在合成数据上训练模型并在原始数据上进行评估的范式，已广泛用于

效用 评估——衡量合成数据是否保留分析关系。RAPID将这一范式重新用于风险 评估：当预测目标为敏感属性时，能够表明高效用的相同预测准确率，也表明高披露风险。与DCR不同，RAPID提供了一种可解释且有界的度量，并根据类别流行率进行校准。需要特别指出的是，此处提供的RAPID R实现同时支持两种评估模式：默认的基于匹配的方法（在合成数据上训练，在全部原始记录上评分），以及可选的基于留出的方法（在合成数据上训练，仅在未用于训练合成器的原始记录留出子集上评分）。留出模式使实践者能够在统一框架内同时评估记忆风险和属性推断风险。有关实现细节，请参见B节。

- **贝叶斯方法** 贝叶斯披露风险度量 [Reiter等人，2014年，Hu等人， 2021] 通过对数据生成过程中的不确定性进行积分，给出属性披露的后验概率。这些方法可提供有原则的最坏情况界限，但需要分布假设，且计算开销可能很大。

  RAPID采用频率学派的基于模拟的方法，计算效率高且所需假设较少。RAPID不提供最坏情况界限，而是在指定的攻击者模型下估计实际风险。为进行保守评估，实践者可以评估多个攻击者模型，并报告 $\text{RAPID}_{\max}$ （第2节）。

6.2 局限性

需要承认以下几项局限性：

- **阈值选择：** τ （分类变量）和 ε （连续变量）的选择会显著影响风险估计，且需要领域专业知识。阈值选择并非RAPID独有——k-匿名、l-多样性、DiSCO的一致性标准以及差分隐私的 ε 中也有类似参数——但目前没有普遍适用的指导。 `rapid::plot()` 方法会生成敏感性曲线，以探究被标记记录的比例如何随阈值选择而变化（见附录B）。对于难以论证阈值设置的应用，置换检验等数据驱动方法可作为替代方案。

- **模型依赖性：** RAPID 的风险估计取决于攻击者模型的选择。不同模型可能会对同一条记录得出不同的风险评估。这既是一项优势——能够反映不同攻击者带来的不同威胁——也是一项局限，因为它会引入歧义。为进行保守评估，实践者可以评估多个攻击者模型，并报告 $\text{RAPID}_{\max}$，即所有模型中的最高风险。机器学习模型固有的不透明性，使 RAPID 的评估比依赖组合唯一性准则（后者可能会低估风险）更难审计。这也进一步增加了法规合规和利益相关方沟通的难度。

- **单一敏感属性：** 当前的表述一次评估一个敏感属性的风险。在实践中，数据集往往包含多个敏感变量。将 RAPID 扩展至对多个属性进行联合评估或序贯评估，是未来的一个研究方向。

6.3 实践指南

根据我们的实证结果，我们为实践者提供以下指南：

- 将 $\tau = 0.3$ 作为分类属性的起始值，连续属性则使用 $\varepsilon = 10\%$。这一选择高度依赖变量；对于敏感性较高的变量，应考虑采用更严格的阈值。
- 使用多个攻击者模型（例如随机森林、梯度提升）评估 RAPID，并报告观测到的最高风险，以进行保守评估。
- 使用 `rapid::plot()` 方法生成阈值曲线（RAPID 与 τ 的关系），以了解敏感性，并根据具体政策要求指导阈值选择。
- 使用逐记录风险标记进行归因分析：按子群体对 RAPID 进行分层，交叉制表比较不同准标识符组合下的风险率，或拟合回归模型，以识别哪些准标识符模式会驱动披露风险。

- 在比较合成器或匿名化策略时，将 RAPID 与效用指标结合起来考量，以权衡隐私与效用。

- 针对高风险记录采取缓解策略：当 RAPID 识别出不可接受的风险等级时，有多种干预措施可供选择。不要简单地删除高风险记录——这会改变风险分布，并在后续评估中产生新的高风险记录。相反，可以考虑：1. 通过粗化（对连续变量进行分箱、合并类别）或采用形式化匿名化方法，降低准标识符粒度。使用归因分析（第 5.5 节）确定哪些准标识符会驱动预测准确率；2. 如果高度预测性的准标识符对于分析效用并非必需，则将其从发布数据中排除；3. 向合成数据中的敏感属性添加后处理噪声；4. 使用更强的隐私约束（例如差分隐私、受限条件采样）重新训练合成器；5. 通过受控环境或数据使用协议限制对合成数据的访问；6. 对于敏感属性并非高度机密且利益相关者理解其中权衡的应用场景，接受该风险。

7 结论

传统的统计披露控制侧重于身份披露和记录唯一性，但现代合成数据生成器带来了不同的威胁：高保真合成能够如此有效地保留预测结构，以至于仅凭准标识符就可能推断出敏感属性。RAPID 通过在现实的攻击者假设下直接测量属性推断脆弱性，应对这一挑战。以往的度量关注的是“合成数据生成器是否复制了可识别信息和敏感信息？”，而 RAPID 关注的是“攻击者能否有把握地预测其本不应知道的内容？”

这种重新界定使以往难以做出的实际决策成为可能。数据托管方如今不仅可以根据效用比较合成器，还可以比较其推断风险，从而识别哪些方法会泄露预测信号，哪些方法能够保护隐私。逐记录风险评分揭示了高风险子群体——即推断能够成功的准标识符组合——从而可以进行针对性修复，而非全面抑制。或许最重要的是，RAPID 有助于回答每次高效用合成数据发布都隐含的问题：效用是否过高，以至于反映出对原始数据预测结构的危险性忠实保留？

我们强调 RAPID 有意不作出哪些主张。它并非差分隐私意义上的隐私保证，也不提供最坏情况界限，或消除推断性披露的可能性。相反，RAPID 是一种基于场景的经验诊断：它量化在指定攻击者模型下、以发布数据所保留的预测结构为条件时实际发生的脆弱性。因此，RAPID 是形式化隐私框架的补充，而非替代。差分隐私以效用和可解释性为代价提供可证明保证；RAPID 则提供透明且可操作的风险评估，支持知情的发布决策。

这项研究的时机反映了数据共享领域更广泛的发展趋势。开放科学要求日益要求数据具备可用性，但微观数据发布仍受到法律和伦理限制。合成数据已成为一种务实的解决方案，在官方统计、健康研究和社会科学领域的应用不断加速。与此同时，机器学习的进步以及关于个人的公开信息日益丰富，大幅提升了隐私攻击的可行性。RAPID 提供了缺失的诊断手段：评估那些看起来安全的合成数据，是否确实能抵御现代机器学习让攻击者轻而易举就能发起的推断攻击。

未来的工作包括扩展至多个敏感属性，将效用指标整合到统一的风险—效用仪表板中，以及开发迭代式工作流程：在重新评估之前，有选择地保护 RAPID 识别出的高风险记录。原则上，RAPID 框架也适用于纵向数据，其中时间特征可提供额外的准标识符，并可按观测值评估风险，或按个体汇总评估；调整方法以考虑个体内相关性，是未来研究的自然方向。

代码可用性

核心 RAPID Thees 等人 [2026] 实现可在 https://github.com/qwertzlbry/RAPID获取。完整的可复现代码，包括所有模拟脚本（第 5 节）和真实数据分析（第 4 节），将在期刊发表后发布。

致谢

本研究得到了瑞士国家科学基金会（SNSF）Bridge Discovery 项目（编号 211751）的资助。我们感谢 Jiří Novák 在 RAPID 早期实现工作中所作的贡献以及参与的讨论。

参考文献

Andreas Alfons、Stefan Kraft、Matthias Templ 和 Peter Filzmoser。用于住户调查的近似真实总体数据模拟及其在 EU-SILC 中的应用。《统计方法与应用》，20(3)：383–407，2011年。ISSN 1618-2510，1613-981X。URL https://doi.org/10.1007/s10260-011-0163-2。

Andrés F. Barrientos、Alexander Bolton、Tom Balmat、Jerome P. Reiter、John M. de Figueiredo、Ashwin Machanavajjhala、Yan Chen、Charley Kneifel 和 Mark DeLong。通过合成与验证开放机密研究数据：一项针对美国联邦政府雇员数据的应用。《应用统计年鉴》，12(2)：1124–1156，2018年。URL https://doi.org/10.1214/18-AOAS1194。

Barry Becker 和 Ronny Kohavi。Adult [数据集]。UCI 机器学习知识库，1996年。网址 https://doi.org/10.24432/C5XW20。

Alberto Blanco-Justicia、David Sánchez、Josep Domingo-Ferrer 和 Krishnamurty Muralidhar。关于在机器学习中使用（及误用）差分隐私的批判性综述。《ACM计算调查》，55(8)：第160号文章，2022年。网址 https://doi.org/10.1145/3547139。

Tianqi Chen 和 Carlos Guestrin。XGBoost：一种可扩展的树提升系统。载于《第22届 ACM SIGKDD知识发现与数据挖掘国际会议论文集》，第785–794页，美国加利福尼亚州旧金山，2016年8月。ACM。ISBN 978-1-4503-4232-2。网址 https://doi.org/10.1145/2939672.2939785。

Charles J. Clopper 和 Egon S. Pearson。关于置信限或枢轴限的使用：以二项分布为例。《生物统计学》，26(4):404–413，1934年。URL https://doi.org/10.2307/2331986。

Tore Dalenius。迈向统计披露控制的方法论。《统计学期刊》，15:429–444，1977年。ISSN 0039-7261。

Josep Domingo-Ferrer、David Sánchez 和 Alberto Blanco-Justicia。差分隐私的局限性（及其在数据发布和机器学习中的误用）。*ACM* 通讯，64(7):33–35，2021年。ISSN 0001-0782。URL https://doi.org/10.1145/3433638。

Josep Domingo-Ferrer、David Sánchez 和 Krishnamurty Muralidhar。统计披露控制：向前迈进。*Journal of Official Statistics*，41(3):820–826，2025年。ISSN 0282-423X。URL https://doi.org/10.1177/0282423X241312023。

George Duncan 和 Diane Lambert。微观数据的披露风险。《商业与经济统计期刊》，7(2):207–217，1989年。URL https://doi.org/10.2307/1391438。

George Duncan、Sallie Keller-McNulty 和 Lynne Stokes。披露风险与数据效用：风险—效用保密性图。技术报告 LA-UR-01-6428，洛斯阿拉莫斯国家实验室，2001年。URL https://www.niss.org/sites/default/files/technicalreports/tr121.pdf。

Cynthia Dwork 和 Moni Naor。《统计数据库中防止披露的难点，或差分隐私的必要性》。《隐私与保密期刊》，2(1):93–107，2010年。ISSN 2575-8527。URLhttps://doi.org/10.29012/jpc.v2i1.585。

Cynthia Dwork、Frank McSherry、Kobbi Nissim 和 Adam Smith。《隐私数据分析中根据敏感度校准噪声》。载于《第三届密码学理论会议论文集》，第265–284页，柏林、海德堡，2006年。施普林格出版社。URL https://doi.org/10.1007/11681878_14。

Khaled El Emam、Laura Mosquera 和 Jason Bass。评估完全合成医疗 th
数据中的身份披露风险：模型开发与验证。《医学互联网研究杂志》，22(11)：文章 e23139，202 0.
URL https://doi.org/10.2196/23139。Paul Francis 和 David Wagner。迈向更完善的属性推断脆弱性度量，2025年。UR L
https://doi.org/10.48550/arXiv.2507.01710。

Markus Hittmeir、Rudolf Mayer 和 Andreas Ekelhart。合成数据中的属性披露风险基线。载于第十届 *ACM* 数据与应用安全及隐私会议论文集，CODASPY '20，第 133–143 页，纽约，美国纽约州，2020年。ACM。ISBN 9781450371070。URL https://doi.org/10.1145/3374664.3375722。

Jingchen Hu、Terrance D. Savitsky 和 Matthew R. Williams。基于随机样本对总体规模和重叠进行贝叶斯估计。《隐私与保密期刊》，11(1)，2021年。URL https://doi.org/10.29012/jpc.748。

Anco Hundepool、Josep Domingo-Ferrer、Luisa Franconi、Sarah Giessing、Eric Schulte Nordholt、Keith Spicer 和 Peter-Paul de Wolf。《统计披露控制》。威利调查方法学丛书。约翰·威利父子有限公司，第1版，2012年。ISBN 978-1-119-97815-2。URL https://doi.org/10.1002/9781118348239。

Saloni Kwatra 和 Vicenç Torra。通过属性推断攻击对生成模型创建的合成数据进行实证评估。载于 Felix Bieker、Silvia de Conca、Nils Gruschka、Meiko Jensen 和 Ina Schiering 编，《隐私与身份管理：数字世界中的共享》，第282–291页，Cham，2024年。施普林格自然瑞士。ISBN 978-3-031-57978-3。URL https://doi.org/10.1007/978-3-031-57978-3_18。

Kidist Amde Mekonnen。无需训练数据集的条件生成对抗网络，2024年。URL https://doi.org/10.48550/arXiv.2405.20687。

Marko Miletic 和 Murat Sariyar。将合成数据生成方法用于表格微观数据所面临的挑战。*Applied Sciences*，14(14)：文章 5975，2024年。ISSN 2076-3417。URL https://doi.org/10.3390/app14145975。

Rainer Mühlhoff。预测性隐私：迈向数据分析的应用伦理学。伦理与信息技术，23(4)：675–690，2021年。ISSN 1388-1957，1572-8439。URL https://doi.org/10.1007/s10676-021-09606-x。

Krishnamurty Muralidhar 和 Josep Domingo-Ferrer。数据库重构并非易事，且不同于重新识别。*Journal of Official Statistics*，39(3)：381–398，2023年。URL https://doi.org/10.2478/jos-2023-0017。

Krishnamurty Muralidhar 和 Steven Ruggles。承诺升级：美国人口普查局为在2020年十年一次的人口普查中实施差分隐私所做努力的案例研究，2024年。网址 https://doi.org/10.48550/arXiv.2407.15957。

B. A. Nosek、G. Alter、G. C. Banks、D. Borsboom、S. D. Bowman、S. J. Breckler、S. Buck、C. D. Chambers、G. Chin、G. Christensen、M. Contestabile、A. Dafoe、E. Eich、J. Freese、R. Glennerster、D. Goroff、D. P. Green、B. Hesse、M. Humphreys、J. Ishiyama、D. Karlan、A. Kraut、A. Lupia、P. Mabry、T. Madon、N. Malhotra、E. Mayo- Wilson、M. McNutt、E. Miguel、E. Levy Paluck、U. Simonsohn、C. Soderberg、B. A. Spellman、J. Turitto、G. VandenBos、S. Vazire、E. J. Wagenmakers、R. Wilson 和 T. Yarkoni。促进开放研究文化。*Science*，348(6242):1422–1425，2015年。ISS N 0036-8075。网址 https://doi.org/10.1126/science.aab2374。

Beata Nowok、Gillian M. Raab 和 Chris Dibben。synthpop：在 R 中定制创建合成数据。统计软件期刊 *al*
统计软件期刊，74(11):1–26，2016。ISSN 1548-7660。URL https://doi.org/10.18637/jss.v074.i1 1.

Michael Platzer 和 Thomas Reutterer。《混合类型合成数据的基于留出法的实证评估》。*Frontiers in Big Data*，4：文章编号 679939，2021年。网址 https://doi.org/10.3389/fdata.2021.679939。

Gilli an M. Raab, Beata Nowok, Chris Dibb^en、an d Joshua Snok^e.合成数据 *thpop User Guide*. University of 爱丁堡，2024年。URL https://cran.r-project.org/package=synthpop。R 包手册/使用指南；属性披露工具，包括 DiSCO。Gillian M. Raab、Beata Nowok 和 Chris Dibben。合成数据的实用隐私度量指标，2025年。URLhttps://doi.org/10.48550/arXiv.2406.16826。Jerome P. Reiter、Quanli Wang 和 Biyuan Zhang。多重插补合成数据披露风险的贝叶斯估计。《隐私与保密期刊》，6(1)，2014年。ISSN 2575-8527。URL https://doi.org/10.29012/jpc.v6i1.635。Donald B. Rubin。统计披露限制讨论。*Journal of Official Statistics*，9(2):461–468，1993年。Manel Slokom、Peter-Paul de Wolf 和 Martha Larson。机器学习模型何时泄露：对合成训练数据的探索。载于 Josep Domingo-Ferrer 和 Maryline Laurent 编，统计数据库隐私，《计算机科学讲义》第 13463 卷，第 283–296 页，瑞士 Cham，2022年。施普林格。ISBN 978-3-031-13945-1。URL https://doi.org/10.1007/978-3-031-13945-1_20。瑞士国家科学基金（SNSF）。开放研究数据：当前实践初探。 https://data.snf.ch/stories/open-research-data-2023-de.html，2024年。访问日期：2025年10月15日。Jennifer Taub、Mark Elliot、Maria Pampaka 和 Duncan Smith。合成数据的差分正确归因概率：一项探索。载于 Josep Domingo-Ferrer 和 Francisco Montes 编，统计数据库隐私，第 122–137 页，Cham，2018年。施普林格国际出版社。ISBN 978-3-319-99771-1。URL https://doi.org/10.1007/978-3-319-99771-1_9。Matthias Templ。为公众和研究人员提供效用高且无披露风险的数据：高级统计披露风险评估。《奥地利统计学期刊》，43(4):247–254，2014年。URL https://doi.org/10.17713/ajs.v43i4.43。Matthias Templ、Bernhard Meindl、Alexander Kowarik 和 Olivier Dupriez。合成复杂数据的模拟：R 包 simPop。*Journal of Statistical Software*，79(10):1–38，2017年。URL https://doi.org/10.18637/jss.v079.i10。Oscar Thees、Jiří Novák 和 Matthias Templ。复杂表格数据上合成数据生成器的评估。载于 Josep Domingo-Ferrer 和 Melek Önen 编，统计数据库隐私，第 194–209 页，Cham，2024年。施普林格自然瑞士。URL https://doi.org/10.1007/978-3-031-69651-0_13。Oscar Thees、Roman Müller 和 Matthias Templ。超越权衡曲线：用于评估匿名化数据和合成数据的多变量高级风险—效用图，2025年。URL https://doi.org/10.48550/arXiv.2510.23500。Oscar Thees、Matthias Templ 和 Roman Müller。*RAPID*：通过预测评估推断型披露风险，2026。URL https://github.com/qwertzlbry/RAPID。R 包版本 0.1.0。Joshua Ward、Yuxuan Yang、Chi-Hua Wang 和 Guang Cheng。针对表格生成模型的成员推断攻击集成方法。载于第 *18* 届 *ACM* 人工智能与安全研讨会论文集，AISec '25，第 182–193 页，纽约，NY，美国，2025年。ACM。ISBN 979-8-4007-1895-3。URLhttps://doi.org/10.1145/3733799.3762977。Leon Willenborg 和 Ton de Waal。统计披露控制要素，《统计学讲义》第 155 卷。施普林格，第 1 版，2001年。ISBN 978-1-4613-0121-9。URL https://doi.org/10.1007/978-1-4613-0121-9。Edwin B. Wilson。概率推断、继承法则与统计推断。《美国统计学会志》，22(158):209–212，1927年。URL https://doi.org/10.1080/01621459.1927.10502953。Marvin N. Wright 和 Andreas Ziegler。ranger：一种在 C++ 和 R 中快速实现高维数据随机森林的方法。*Journal of Statistical Software*，77(1):1–17，2017年。URL https://doi.org/10.18637/jss.v077.i01。

Zexi Yao、Nataša Krčo、Georgi Ganev 和 Yves-Alexandre de Montjoye。dcr谬误：衡量合成数据的隐私风险。载于 Vincent Nicomette、Abdelmalek Benzekri、Nora Boulahia-Cuppens 和 Jaideep Vaidya 编，计算机安全——*ESORICS 2025*，第 469–487 页，Cham，2026。施普林格自然瑞士。URL https://doi.org/10.1007/978-3-032-07884-1_24。

A 模拟数据生成

我们模拟生成 n 条相互独立的微观数据记录，包含六个变量：性别（G）、年龄（A）、教育程度（E）、收入（I）、健康评分（H）和疾病状态（D）。该设计通过潜在社会经济地位（SES）变量，编码了符合现实且与政策相关的依赖关系，同时通过全局依赖参数 $\kappa \geq 0$ 保持透明且可调。全文中，$\mathsf{TN}(\mu,\ \sigma^2; [a,b])$ 表示截断至 $[a,b]$ 的正态分布；在指定情况下，我们使用˜标准化预测变量 $X = (X - \mu_X)/\sigma_X$。

依赖机制。
所有变量间的依赖关系都通过一个共享的潜变量传递：

$$\mathrm{SES}_i \sim \mathcal{N}(0,1),$$

该潜变量代表未观测到的社会经济地位。依赖关系的强度由根据 κ得出的信号与噪声权重控制：

$$w_{\text{signal}} = \sqrt{\frac{\kappa}{1+\kappa}}, \qquad w_{\text{noise}} = \sqrt{\frac{1}{1+\kappa}}.$$

当 $\kappa = 0$时，关系完全由噪声驱动（$w_{\text{signal}} = 0$、 $w_{\text{noise}} = 1$）。随着 $\kappa \to \infty$，依赖关系变为确定性的（$w_{\text{signal}} \to 1$、 $w_{\text{noise}} \to 0$）。在默认值_ $\kappa = 1$ 下，信号与噪声的贡献相等（$w_{\text{signal}} = w_{\text{noise}} = 1/\sqrt{2}$），解释方差约为 50%。

Age.
我们从截断正态分布中抽取年龄，以反映成年人口：

$$A_i \sim \mathsf{TN}(\mu_A, \sigma_A^2; [a_{\min}, a_{\max}]), \quad \text{defaults: } \mu_A = 45,\ \sigma_A = 12,\ a_{\min} = 18,\ a_{\max} = 85.$$

教育程度。
教育程度是一个有序分类变量，包含三个水平{0,1,2}，分别对应低、中等和高受教育程度。我们根据同时取决于社会经济地位和年龄的潜在连续变量生成 E_i：

$$L_i = w_{\text{signal}} \cdot (0.8 \cdot \mathrm{SES}_i - 0.4 \cdot \tilde{A}_i) + w_{\text{noise}} \cdot \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0,1),$$

$$E_i = \begin{cases} 0 & \text{if } L_i < c_1, \\ 1 & \text{if } c_1 \leq L_i < c_2, \\ 2 & \text{if } L_i \geq c_2, \end{cases} \quad c_1 = -0.3,\ c_2 = 0.7.$$

负的年龄系数 反映了较年轻队列受教育程度更高的实证模式 .

收入。
我们使用线性模型生成对数收入：

$$\log I_i^\star = w_{\text{signal}} \cdot (0.5 \cdot \mathrm{SES}_i + 0.3 \cdot \tilde{A}_i + 0.25 \cdot E_i) + w_{\text{noise}} \cdot \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0,1),$$

$$I_i = \exp(10 + \log I_i^\star).$$

T常数 10 将收入居中到合理范围（约为 20,000–40,000 美元） .

健康评分。
H 是 $[0,100]$ 上的连续变量健康度量，通过对线性预测量进行 Sigmoid变换创建：

$$H_i^\star = w_{\text{signal}} \cdot \left(0.6 \cdot \mathrm{SES}_i - 0.5 \cdot \tilde{A}_i + 0.2 \cdot E_i + 0.2 \cdot \widetilde{\log I_i}\right) + w_{\text{noise}} \cdot \zeta_i, \quad \zeta_i \sim \mathcal{N}(0,1),$$

$$H_i = \frac{100}{1 + \exp(-H_i^\star)}.$$

年龄系数为负，反映了健康状况随年龄增长而下降；社会经济地位、教育程度和收入系数为正，则代表其具有保护作用。

疾病状态。

$D_i \in \{\texttt{healthy}, \texttt{diabetic}, \texttt{hypertensive}\}$ 通过以 `healthy` 为基线的多项逻辑回归进行抽样。与其他变量不同，疾病依赖性随 κ 线性变化，而非通过信号/噪声权重来调节：

$$\log \frac{\Pr(D_i = \texttt{diabetic})}{\Pr(D_i = \texttt{healthy})} = -1.5 + \kappa \cdot \big(0.8 \cdot \tilde{A}_i - 0.3 \cdot \widetilde{\log I_i} - 0.2 \cdot E_i\big),$$
$$\log \frac{\Pr(D_i = \texttt{hypertensive})}{\Pr(D_i = \texttt{healthy})} = -1.3 + \kappa \cdot \big(1.0 \cdot \tilde{A}_i - 0.2 \cdot \widetilde{\log I_i} - 0.1 \cdot E_i\big).$$

这种线性缩放会在高 κ时产生更强的依赖效应：年龄越大，风险越高；而收入和教育程度越高，则有轻微的保护作用。固定截距（-1.5 和 -1.3）会造成基线类别不平衡，使`healthy` 结果更常见。

性别。

性别为二元变量，与社会经济地位、年龄和教育程度存在轻微依赖：

$$\eta_i = w_{\text{signal}} \cdot (0.3 \cdot \text{SES}_i - 0.2 \cdot \tilde{A}_i + 0.2 \cdot E_i),$$
$$G_i \sim \text{Bernoulli}\big(\text{logit}^{-1}(\eta_i)\big), \quad G_i \in \{\texttt{female}, \texttt{male}\}.$$

控制依赖强度。

全局参数 $\kappa \geq 0$ 控制所有依赖关系的强度：

- $\kappa = 0$：由于疾病模型中的固定截距，变量之间仍存在微弱依赖，但信号贡献趋于消失（$w_{\text{信号}} = 0$）。
- $\kappa = 1$ （默认）：信噪比均衡，因此变量之间存在中等程度的依赖。
- $\kappa \gg 1$：随着$w_{\text{信号}} \to 1$ 和疾病系数增大，关系趋于近乎确定。

由于疾病 logits 随 κ 线性变化，而其他变量采用信噪比变换，因此疾病依赖在高 κ时增强得更快。

默认设置与现实性。

采用默认 $\kappa = 1$时，模拟会生成逼真的边际分布和中等程度的依赖：收入随社会经济地位、年龄和教育程度上升；健康状况随年龄增长而下降，但随社会经济地位提高而改善； `diabetic` 和 `hypertensive` 的概率则随年龄增长而上升，随收入和教育程度提高而下降。通过调整变量特定参数，可将这些默认设置适配为特定领域基线，而无需改变κ。

B 使用示例

RAPID [Thees 等人， 2026] 以开源 R 包的形式实现，并提供 S3 方法以简化分析。典型工作流程分为两个阶段：首先，评估已生成的合成数据集的逐记录风险；其次，若需评估多种合成方法或寻求带有不确定性量化的稳健估计，则使用交叉验证进行方法比较。主要应用场景是在发布前评估特定合成数据集的披露风险：

```
# Install and load required packages
devtools::install_github("qwertzlbry/RAPID")
install.packages("synthpop")
library(RAPID)
library(synthpop)

set.seed(2025)
n <- 1000

# Generate independent variables
```

```
age <- sample(20:70, n, replace = TRUE)
education <- factor(sample(c("low", "medium", "high"), n, replace = TRUE))
gender <- factor(sample(c("M", "F"), n, replace = TRUE))
# Generate disease_status with dependencies
disease_status <- sapply(1:n, function(i) {
probs <- c(0.6, 0.2, 0.2)  # Base: healthy, diabetic, hypertensive
# Older people more likely diabeticif (age[i] > 55) {
probs <- c(0.3, 0.5, 0.2)}# Low education + older → more diabetic
if (age[i] > 55 && education[i] == "low") {probs <- c(0.2, 0.6, 0.2)
}sample(c("healthy", "diabetic", "hypertensive"), 1, prob = probs)})
disease_status <- factor(disease_status)
data_orig <- data.frame(age, education, gender, disease_status)
data_syn <- syn(data_orig, method = "cart", seed = 2025)$syn
# Assess attribute inference riskresult <- rapid(
original_data = data_orig,synthetic_data = data_syn,
quasi_identifiers = c("age", "education", "gender"),
sensitive_attribute = "disease_status",model_type = "rf",
cat_tau = 0.3,return all records = TRUE_    _)# View result
print(result)# RAPID Assessment# ================
# Method: RCS_marginal# Risk level: 15.5 %
# Records at risk: 155 / 1000# Threshold (tau): 0.3
# Detailed summary with model metricssummary(result)
# RAPID Risk Assessment Summary# ==============================
# Evaluation method: RCS_marginal# Attacker model: Random Forest
# Risk Metrics:#   Confidence rate: 0.155
#   Records at risk: 155 ( 15.5 %)# Threshold (tau): 0.3
# Model Performance:#   Accuracy: 0.599
```

```
# Visualize threshold sensitivity
plot(result)

# Identify high-risk records for disclosure control
print(result, type = "high_risk")

disease_status age edu  gender pred_class true_prob baseline normalized_gain cat_tau at_risk
9    diabetic  68  low    M   diabetic    0.6036335    0.292    0.4401603     0.3    TRUE
14   healthy   42  high   F   healthy     0.6587221    0.495    0.3242022     0.3    TRUE
26   diabetic  63  low    M   diabetic    0.5541064    0.292    0.3702068     0.3    TRUE
... and 152 more
```

该程序包提供三种用于结果检查的 S3 方法： `print()` 显示风险等级和阈值的简要摘要； `summary()` 提供详细的评估指标，包括模型性能统计； `plot()` 生成阈值敏感性曲线，展示风险估计如何随 τ （分类变量）或 ε（连续属性）变化。调用 `print(result, type = "high_risk")` 可显示被标记为存在风险的记录，便于实施定向披露控制策略。

评估多种合成方法，或需要获取带置信区间的稳健风险估计时，交叉验证可用于不确定性量化。 `rapid_synthesizer_cv()` 函数实现 k折交叉验证，结果可通过 `print()`进行检查：

```
# Define synthesizer function
cart_synthesizer <- function(data, seed = NULL) {
synthpop::syn(data, method = "cart", m = 1, seed = seed,
 print.flag = FALSE)$syn}
# Cross-validation across 5 folds
cv_result <- rapid_synthesizer_cv(
original_data = data_orig,
synthesizer = cart_synthesizer,
quasi_identifiers = c("age", "education", "gender"),
sensitive_attribute = "disease_status", k = 5,
model_type = "rf", cat tau = 0.3)print(cv_result)
# RAPID Cross-Validation Results
# ==============================
# Evaluation method: RCS_marginal# Attacker model: rf
# K-folds: 5# Threshold (tau): 0.3
# Threshold (epsilon): NA## Risk Estimate:
#   Mean: 0.118#   SD: 0.03#   95% CI: [0.092, 0.144]
```

当合成方法涉及随机性要素时，交叉验证尤为重要，因为它能够量化不同训练—测试划分下风险估计的变异性，并为比较不同合成方法提供有依据的依据。