共线性和样本量对线性回归结果的影响：一项模拟研究

Stephanie C. C. van der Lubbe[1,2], Jose M. Valderas[1,2,3,4], Evangelos Kontopantelis[1,5]*

[1] 家庭医学部，Yong Loo Lin医学院，新加坡国立大学，新加坡[2] 新加坡国立大学综合诊所，新加坡国立大学卫生系统，新加坡，新加坡[3] 家庭医学系，新加坡国立大学卫生系统，新加坡，新加坡[4] 卫生系统绩效研究中心（CRiHSP），新加坡国立大学，新加坡[5] 信息学、影像与数据科学部，生物、医学与健康学院，曼彻斯特大学，英国

* 通讯作者 邮箱：
e.kontopantelis@manchester.ac.uk

摘要

背景：多重共线性会增大 OLS 回归系数的方差，扩大置信区间并降低统计推断的可靠性。应用研究通常采用固定的方差膨胀因子（VIF）阈值（例如 4 或 10）来诊断“有问题的”共线性，并据此决定合并或排除协变量。然而，这种做法往往不分研究样本量差异而一概采用；但共线性的实际影响本质上属于有限样本现象。应用研究的关键问题是：方差膨胀何时会大到足以实质性损害估计精度、覆盖率和推断结论，尤其是在存在模型设定错误的情况下。我们量化了在广泛条件下，共线性与样本量如何共同影响线性回归的表现，并提供了结合具体情境解读 VIF 的实用指南。我们评估了对推断和决策至关重要的结果。

方法： 我们开展了一项模拟研究，将样本量设为 100 至 100,000，并将共线性设为从 VIF = 1（无）到 VIF = 50 （极端）。在每种情景下，我们生成 1,000 个数据集，并拟合 OLS 模型，以估计预先指定的“主要”系数。我们评估了：(i) 覆盖率（95%置信区间包含真实值的比例）；(ii) 平均绝对误差和偏差；(iii) 传统检验效能（95%置信区间不包含 0 的比例）；以及 (iv) 精度保证（95%置信区间完全落在真实效应周围具有实际意义的范围内的概率）。此外，我们还通过省略一个相关预测变量，评估了模型设定错误（有偏）的情景，以量化共线性如何放大偏差。

结果： 在模型设定正确时，共线性并未实质性影响名义覆盖率，也未引入系统性偏差，但在小样本中显著降低了精度：当样本量为 $N \approx 100$ 时，即使是轻度共线性（VIF < 2）也会增大误差，并显著降低两项功效指标；而当样本量为 $N \geq 50{,}000$ 时，即使存在极端共线性（VIF = 50），估计值仍然稳健。在模型设定错误时，共线性会显著放大偏差，导致误差增大、覆盖率降低，并使精度保证和传统检验效能大幅下降，即使 VIF 值较低时也是如此。

结论： 不应机械地应用固定的VIF阈值。必须结合样本量和潜在偏差来源来解读共线性：在小型研究中，适度的VIF也可能产生重大影响；而在大型研究中，极高的VIF在实际中可能微不足道。重要的是，仅为降低VIF而排除预测变量，可能会引入或加剧遗漏变量偏倚，从而削弱推断。所提供的热图可作为实用参考，帮助预判不同研究规模和建模假设下共线性的影响。

关键词：共线性、多重共线性、方差膨胀因子、线性回归、有限样本推断；遗漏变量偏差；精度保证、模拟研究

引言

共线性是指自变量之间存在线性关系，因此一个变量可以部分或完全由其他变量解释。(1,2) 在线性回归分析中，共线性会增大估计系数的方差，从而提高发生第一类错误（假阳性）和第二类错误（假阴性）的可能性。(3,4) 共线性的程度通常通过方差膨胀因子（VIF）来评估，该指标量化了由于与其他预测变量相关而导致的系数方差增幅。

尽管人们长期以来一直担忧过于简单化的经验法则，(5–7)许多研究仍继续采用固定的VIF阈值（通常为4或10）来判断共线性是否构成问题。(8–12)这些界值通常被不加区别地使用，而不考虑样本量，导致同样的阈值既被应用于少于50个观测值的分析(13)，也被应用于包含数百万个数据点的大型数据集。(14)当超过VIF阈值时，研究人员通常通过将相关变量合并为一个预测变量，或将其完全排除，来处理共线性。(5)尽管这些策略能够消除回归模型中的共线性，但僵化的阈值也伴随着权衡：界值设置过高，可能无法解决有问题的共线性；而设置过于严格，则可能导致不必要地排除重要的预测变量。

此前已有少数研究考察了共线性和样本量对线性回归模型的影响。(3,7,15,16)在这一背景下，模拟研究至关重要，因为VIF与性能指标（例如统计功效）之间的关系无法用代数方法表示。然而，这些研究的样本量均不超过1,000个观测值，并且在模拟中仅使用了有限的一组VIF值来探讨共线性。此外，大多数研究未考察共线性对统计功效的影响，而统计功效是判断回归结果可靠性的关键结果。因此，这些研究无法揭示共线性如何与规模大得多的数据集相互作用，也无法说明在不同情景下统计功效会受到怎样的影响。尤其是在流行病学、卫生服务研究、生物医学数据科学及其他依赖回归模型支持循证结论的学科中，大型数据库日益广泛可用且得到应用，这些空白尤为重要。过度简化的VIF阈值——尤其是在不考虑样本量的情况下使用时——可能导致对共线性影响回归结果的错误评估，进而可能对暴露-结局关系得出有偏差或误导性的结论。

在本研究中，我们开展了一项全面的模拟研究，系统考察不同程度的共线性与样本量如何共同影响各种条件下的线性回归结果

。我们模拟了样本量从 100 到 100,000、方差膨胀因子（VIF）值从 1 到 50 的数据集，并评估其对覆盖率、误差、偏差、精度保证（这一基于精度概率的指标已广泛应用于统计软件）以及传统假设检验功效的影响。共线性的定性影响（例如方差膨胀）虽已广为人知，但现有指南对于这些影响在有限样本区间内如何变化，或共线性达到何种阈值后会对推断产生实际影响，鲜有论述。通过绘制数千种情景下这种共同关系的图谱（包括设定正确的模型和设定错误的模型，即遗漏变量模型），我们的研究结果为共线性在何种情况下会实质性影响估计准确性、何时影响可以忽略，以及这些模式在小规模与大规模数据集之间有何差异，提供了新的定量见解。由此生成的热图和性能曲面为不同科学领域的应用研究人员提供了实用且结合具体情境的指导，突破僵化的方差膨胀因子阈值，转而采用基于实证的评估方法，并明确纳入样本量和潜在模型结构。

方法

为考察共线性和样本量对覆盖率、误差、偏差、精度保证和统计功效的影响，我们在方差膨胀因子（VIF）和样本量（N）的各种组合下生成了数据集。这种模拟方法使我们能够系统地改变样本量和共线性，同时保持效应量、误差方差以及预测变量和误差项的分布假设不变，从而确保观察到的回归估计值影响可直接归因于共线性和样本量。我们通过设置范围广泛的 VIF 值来调整共线性，其取值从 1（无共线性）到 50（极高共线性）不等。样本量从 100 变动至 100,000，涵盖从极小样本到极大样本的情况，以体现统计结果在三个数量级范围内受到的影响。对于 VIF 和样本量的每一种组合（称为一个情景），我们生成了 1,000 个数据集，在计算效率与估计统计指标的稳定性之间取得了良好平衡。下面总结了数据集生成过程；更多计算细节请参见 **补充方法**。

数据集生成

数据集生成方式为：从标准正态分布中抽取随机值，组成一个用于 X 的 $N \times X$ 数组，其中 N 为样本量，且 $X = [x_{\text{main}}, x_1, x_2, x_3, x_4, x_5]$。为使预测变量之间达到所需的共线性水平，我们使用由指定相关矩阵导出的乔列斯基因子对 X 进行了变换。对于所有预测变量均存在部分共线性的数据集，非对角

元素 r 在相关矩阵中的计算如下（推导过程参见**补充方法**）：

$$r = \frac{R_i^2(p-2) + \sqrt{R_i^4(p-2)^2 + 4R_i^2(p-1)}}{2(p-1)} \quad \text{Eq. 1}$$

其中，p 表示预测变量的数量，R_i^2 表示将一个预测变量对其他$p-1$ 个预测变量进行回归时的决定系数。误差项ε 的均值设为0，方差设为$\pi^2/3$，与标准逻辑分布假设下逻辑回归中ε 的固定方差相匹配。在假设外生性均值为0且方差有限的情况下，无论具体的误差分布如何，普通最小二乘法始终保持无偏；此处使用逻辑分布仅是为了固定误差方差，以便在不同情景之间进行比较。随后，根据方程2计算结果y：

$$y = \beta_0 + \beta_{\text{main}} x_{\text{main}} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon \quad \text{Eq. 2}$$

其中，真实系数随机设定如下：$\beta_0 = 10$、$\beta_{\text{main}} = 2$、$\beta_1 = 1.3$、$\beta_2 = 1.5$、$\beta_3 = 6$、$\beta_4 = 3$和$\beta_5 = 1$。对于每个情景，该过程重复进行1,000次，生成1,000个维度为$N \times 8$的数据集，其列分别对应x_{main}、x_1、x_2、x_3、x_4、x_5、ε和y。为验证共线性程度，我们计算了方差膨胀因子（VIF）：

$$\text{VIF}(x_i) = \frac{1}{1 - R_i^2} \quad \text{Eq. 3}$$

统计分析

针对每个数据集，使用 Python 的*statsmodels* 软件包进行了普通最小二乘法（OLS）线性回归。此方法可提供系数估计值、标准误、置信区间和统计显著性检验。

为评估主要系数 β_{main} 的估计值准确性，我们针对每个回归评估了覆盖率、误差、偏差、精度保证和统计功效。覆盖率定义为真实系数 β_{main} 落在估计系数 $\hat{\beta}_{\text{main}}$ 的 95%置信区间内的模拟所占比例。覆盖率表明置信区间是否涵盖真实值，但并不反映估计值可能偏离真实值多远（即使误差很大，较宽的置信区间仍可能带来较高的覆盖率）。

误差是估计值与真实值之差，$\hat{\beta}_{\text{main}} - \beta_{\text{main}}$。根据这一差值，我们得出了两个指标：偏差和平均绝对误差（MAE）。偏差反映模拟中系统性高估或低估的情况，因此体现了方向性误差。平均绝对误差表示误差的平均幅度，不考虑误差方向；计算时先取每个单独误差的绝对值，再求平均。

我们采用了两种相关的功效指标：精度保证和传统检验效能。在主要分析方法中，我们根据估计值的精度定义功效：如果 $\hat{\beta}_{\text{main}}$ 的置信区间完全包含在预先设定的范围 $[\beta_{\text{main}} \pm c_{\text{powval}}]$内，则认为该回归具有功效。对 c_{powval} 进行了校准，使得在 *N*=1,000 且不存在共线性（VIF=1）的情景下，功效达到 80%，从而得到 $c_{\text{powval}} = 0.189$ 的取值。我们将这一基于精度的功效指标称为“精度保证”，SAS 和 Stata 等标准统计软件也实现了该指标。这一指标类似于传统检验效能，但独立于 β_{main}的效应量评估估计精度。(17,18) 这一特点使其特别适用于我们的模拟研究，因为研究重点是估计准确性而非假设检验。作为次要指标，我们还采用了传统定义计算功效：如果 $\hat{\beta}_{\text{main}}$ 的置信区间不包含 0，则认为该回归具有功效。除非另有说明，我们将精度保证作为与本研究问题更为相关的指标进行报告。

此外，还使用 Cohen 的 *d* 评估效应量的影响，其定义为 $\beta_{\text{main}}/\sigma_{\varepsilon}$，其中$\sigma_{\varepsilon} = \pi/\sqrt{3}$ 表示误差项 ε 的标准差。通过将β_{main} 从 2 逐步降至 0（每步 0.1）来改变效应量，同时保持 σ_{ε} 不变。除非另有说明，本文结果均使用 $\beta_{\text{main}} = 2$生成，对应的效应量约为 $d \approx 1.10$。所有报告的测量指标，包括覆盖率、统计功效、偏差和绝对误差，均针对主要系数 β_{main} 计算，取每种情景下 1,000 次模拟结果的均值和标准误。所有模拟均使用 Python 3.11.5 完成。

3. 结果

覆盖率、绝对误差和偏差

图1展示了 x_{main} 与 x_1之间的共线性（以方差膨胀因子 VIF 衡量）不断增强时，平均覆盖率、平均绝对误差（MAE）和平均偏差的变化。不同颜色的线条代表不断增大的样本量，范围从 N=100 （深红色）到 100,000（深蓝色）。在所有样本量和共线性水平下，平均覆盖率均保持在约 95%。这些结果

表明，即使存在极端共线性，置信区间仍能在约 95% 的情况下可靠地涵盖真实系数 β_{main}。

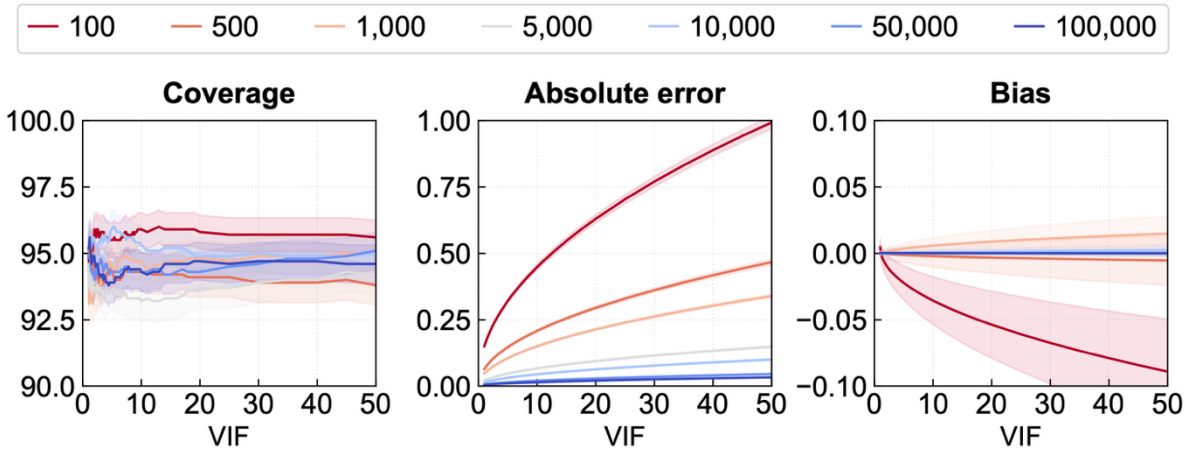

图1。覆盖率、绝对误差和偏差随共线性水平的变化。 $\hat{\beta}_{\text{main}}$ 的平均覆盖率（%）、平均绝对误差和平均偏差随共线性的变化情况如图所示，共线性通过方差膨胀因子（VIF）测量，其中x_{main} 和x_1 存在部分共线性。不同颜色代表不同的样本量，范围从 100（深红色）到 100,000（深蓝色）。阴影区域表示均值估计周围的标准误。

平均绝对误差（MAE）随 VIF 值升高而增加，增幅取决于样本量和共线性水平。对于 N=100（最小样本量），共线性的影响最为显著：MAE 随 VIF 升高而迅速增加，在 VIF=50时达到 0.99。对于更大的样本量，MAE 的增幅则较为平缓；N=100,000（最大样本量）在 VIF=50时的 MAE 仅为 0.03。

对于所有样本量，所有共线性水平下的平均偏差都接近于零，但 N=100 以及在较小程度上 N=1,000 除外。对于这些较小的样本量，随着共线性增加，平均偏差变得更加明显。这种偏差并非源于数据生成过程中的任何系统性偏差，而是有限样本变异性的结果。样本量较小时，模拟预测变量可能会略微偏离其理论上的 $\mathcal{N}(0,1)$ 分布，从而导致各次模拟的平均估计系数出现微小偏移。为证明这种偏差源于有限样本量，我们针对 N=100，在每个方差膨胀因子水平下增加模拟次数，重新计算了偏差。如补充图 S1 所示，增加模拟次数会减小偏差，使其几乎完全消失。这一结果符合预期，因为生成的数据集不存在系统性偏差。因此，若数据中不存在偏差，仅有共线性并不会引入偏差。然而，当存在偏差时，共线性会使其放大。

精度保证“统计功效”

图2 展示了精度保证“统计功效”如何随 x_{main} 与 x_1之间共线性（以方差膨胀因子衡量）的增加而变化，颜色编码与**图1**相同。圆点表示精度保证（PA）达到 0% 时的方差膨胀因子值。如**图2**所示，平均 PA 是对共线性最敏感的指标，而且正如预期，也受样本量影响。对于 N=1,000，在无共线性（方差膨胀因子=1）时，PA 以 80% 为基准，因此 N=100 和 N=500 的模拟在不存在任何共线性的情况下，PA 已低于 80%。对于 N=1,000，当方差膨胀因子为=1时，PA 为 80%；随后迅速下降，最早在方差膨胀因子为=3时降至 0%。

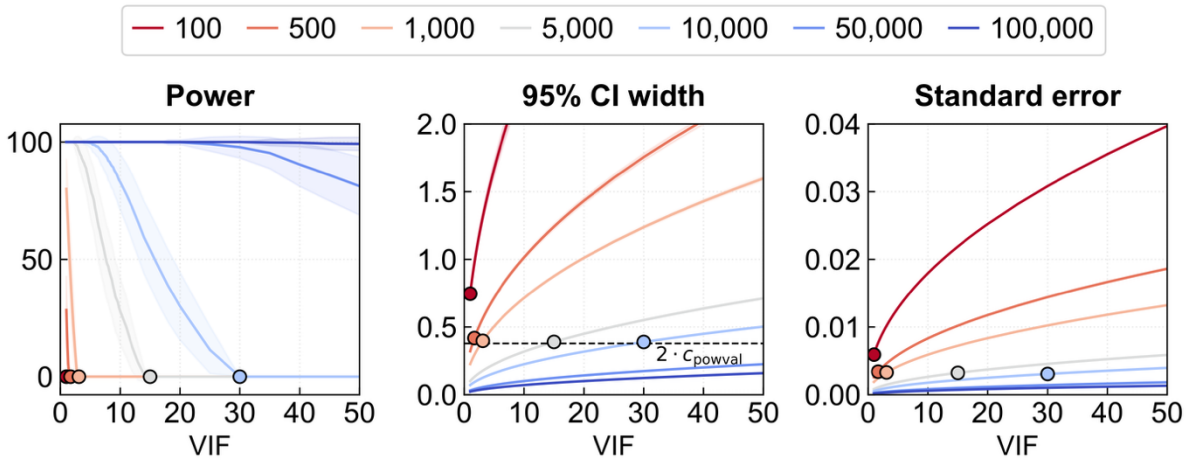

图2。 精度保证 “统计功效”、置信区间宽度和标准误 随共线性水平的变化。图中展示了 $\hat{\beta}_{\text{main}}$ 的平均精度保证（%）、平均95%置信区间宽度和平均标准误（SE）如何随共线性变化；共线性通过方差膨胀因子（VIF）测量，其中 x_{main} 和 x_1 存在部分共线性。不同颜色对应不同的样本量，范围从100（深红色）到100,000（深蓝色）。阴影区域表示均值估计的标准误。

增加样本量可显著提高精度保证对共线性的稳健性。当 N=5,000 和 N=10,000 时，在 VIF=15 和 VIF=30处，精度保证分别降至 0%。对于最大的样本量（N=50,000 和 N=100,000），在整个 VIF 范围内，直至 50，精度保证始终保持较高水平。这些结果凸显了样本量对于在高度共线性条件下维持统计功效的重要性。随着 VIF 增加，精度保证下降的原因在于置信区间变宽（**图2**）。由于共线性会增大 $\hat{\beta}_{\text{main}}$ 的标准误（**图2**），95%置信区间扩展至超出 $2 \cdot c_{\text{powval}}$的范围，从数学上使维持精度保证成为不可能，最终导致其降至 0。这一机制在概念上示于**图3**。

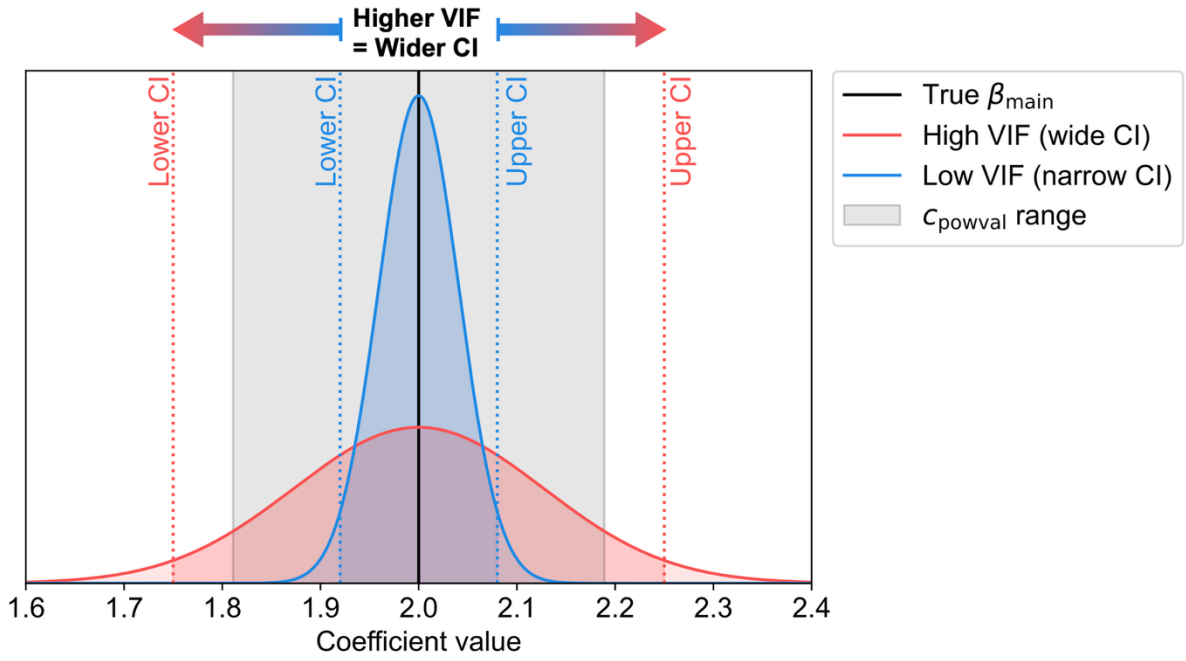

图3。 共线性如何降低精度保证"统计功效"的示意图。 共线性增加会使 $\hat{\beta}_{main}$ 的置信区间（CI）变宽，从而降低 CI 落在 c_{powval} 范围内的可能性。c_{powval} 范围涵盖$\beta_{main} \pm c_{powval}$，采用的方法中规定的相同数值。

效应量的影响

为研究效应量如何影响覆盖率、绝对误差、偏差和精度保证"统计功效"，我们使用取值范围为 0 至 2 的 β_{main} 值（步长为 0.1 ）重复进行了模拟，对应的 Cohen's *d* 值范围为 0 至 ~1.10。

对于所有效应量，覆盖率、平均绝对误差和标准误的结果均保持一致（**补充材料 图 S2**）。由于使用不同 β_{main} 的模拟采用相同的数据生成过程，而且上述指标取决于数据的变异性，而非效应量的绝对大小，因此这一结果符合预期。

由于我们将 "统计功效定义为精度估计（见方法论），因此它在所有效应量 下也保持不变。我们还考察了传统检验效能：如果模拟结果的95%置信区间不包含零，则认为该模拟具有统计功效。如图**图4**所示，传统检验效能会随共线性增强而降低。然而，较大的效应量 表现出更强的抵御共线性的稳健性：距离零更远的估计值，即使标准误更大，也能在其置信区间包含零之前维持这一特性。

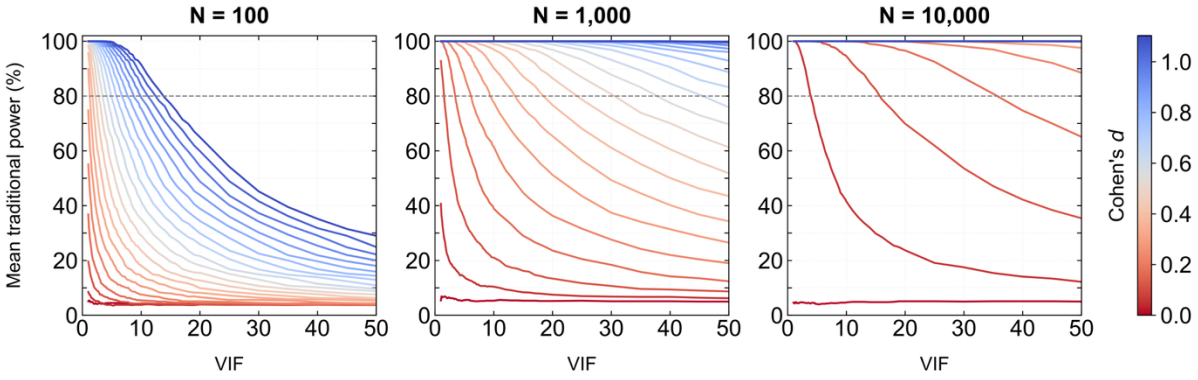

图4。 不同共线性水平和效应量下的传统检验效能。图中显示了 $\hat{\beta}_{\text{main}}$的平均传统检验效能（%）随共线性的变化，共线性通过方差膨胀因子（VIF）衡量，其中 x_{main} 和 x_1 存在部分共线性。每个面板代表不同的样本量：100（左）、1,000（中）和 10,000（右），每种情景进行 1,000 次模拟。颜色表示 Cohen's *d*逐渐增大的取值，从 0（深红色）到 ~1.10（深蓝色），步长为 0.055。

样本量的影响同样显著。当N=100时，对于所有VIF超过=14的效应量，传统检验效能均低于80%。相比之下，当N=10,000时，即使是较小的效应量（$d \geq 0.17$或$\beta_{\text{main}} \geq 0.3$），在整个VIF取值范围内，甚至高达VIF=50 时，仍能保持较高的统计功效（**图4**）。

相关预测变量的数量

我们在所有预测变量均存在部分共线性的情况下重复了模拟。结果表明，与 x_{main} 仅与单个预测变量相关时相比，覆盖率、平均绝对误差、偏差和统计功效基本保持不变（**补充图 S3–S4**）。因此，在保持预测变量总数不变的情况下，增加相关预测变量的数量，对$\hat{\beta}_{\text{main}}$的结果没有显著影响。

此前的所有结果均基于包含六个自变量的数据集（**方程2**）。为评估预测变量总数的影响，我们还使用二十个自变量进行了模拟，设定 $\beta_{\text{main}} = 2$，并为 $[\beta_1, \ldots, \beta_{19}]$ 赋予 0 到 9 之间的数值（见 **补充表S1**），其中所有预测变量均存在部分共线性。与使用六个预测变量的模拟相比，覆盖率、平均绝对误差、偏差和统计功效的结果基本保持不变（**补充图S5–S6**）。因此，与样本量这一具有关键影响的因素不同，预测变量总数和存在部分共线性的预测变量数量均不会对 β_{main} 的结果产生实质性影响。

讨论

与广泛使用的方法相反，共线性并不一定会带来问题。即使存在极端共线性（VIF = 50），只要底层不相关数据不存在偏差，覆盖率仍不受影响，也不会引入偏差。然而，共线性确实会增大估计系数方差，进而导致绝对误差增加、置信区间变宽以及统计功效降低（无论采用传统方法还是基于精度的方法）；这些影响的程度高度取决于样本量。

优势与局限

本研究的一大优势在于，以系统且高分辨率的方式探讨了在广泛的现实情景下，共线性与样本量如何共同影响线性回归推断。渐近理论表明，抽样变异性会随 $1/\sqrt{N}$ 减小，因此随着 N 增大，固定方差膨胀因子的影响会减弱；然而，应用分析基于有限样本开展，因此这种影响减弱的速度以及实际转变的临界界限，仍是本文所探讨的关键实践问题。通过评估多个推断目标（覆盖率、误差、偏差、精度，以及常规假设检验功效和精度保证指标），我们的结果综合呈现了共线性在何种情况下具有实际影响、何时则基本可以忽略。此外，我们还明确考察了正确设定的模型和错误设定（遗漏变量）情景，说明共线性可能与应用研究中常见的偏差机制相互作用。

在解读这些发现时，应考虑几项局限性。首先，我们的主要模拟是在理想化条件下生成的，假设线性模型设定正确，且误差具有独立、均值为零的特性，同时误差项与预测变量相互独立。在这些假设下，普通最小二乘估计量是无偏的，这意味着不存在可被共线性放大的系统性偏差；在最小样本中观察到的任何轻微表观偏差，反映的是有限样本变异性和蒙特卡洛误差，而不是普通最小二乘法无偏性的违反。其次，现实世界数据集经常偏离这些理想条件：测量误差、遗漏变量、模型设定错误（例如非线性）、异方差性或聚类都可能导致偏差，而共线性可能使这些问题更加严重，因为它会增加区分预测变量效应的难度。例如，当遗漏了一个相关变量时，其效应可能被剩余预测变量吸收，尤其是被那些高度共线的预测变量吸收，从而导致有偏的系数和推断质量下降。第三，由于“统计功效”通常被理解为假设检验的统计功效，为避免歧义，我们将主要的精度概率指标称为精度保证（即95%置信区间落在围绕真实效应的、具有实际意义的范围内的概率），并同时报告传统统计功效，以兼顾精度和统计显著性目标。

最后，我们之所以聚焦于普通最小二乘法，是因为在该情境下，人们经常使用固定的方差膨胀因子阈值来证明剔除变量的合理性；将方法扩展到惩罚回归和其他建模框架，可能会产生不同的权衡，也为未来研究提供了自然方向。

启示

在大样本量下（我们的模拟中 N≥50,000），两项相关指标均显示统计功效得以保持；绝对误差仍然较小，且在所有共线性水平下，置信区间都相对较窄。尽管如此，基于僵化的阈值标准，具有较高方差膨胀因子（VIF）的预测变量往往会被排除，这可能导致遗漏对结果有显著贡献的变量。相比之下，在仅有几百个观测值的小样本中，即使是轻微的共线性（例如 VIF=2），也会大幅降低传统检验效能和精度保证。在这种情况下，有意义的预测变量可能被错误地判定为统计学上不显著，从而增加发生第二类错误的风险。这种对样本量的强烈依赖，质疑了常用的 VIF 阈值（如 5 或 10）的有效性，并凸显了在样本量背景下评估共线性的重要性，而非依赖任意设定的界限。

为提供实用指导，我们生成了一张热图，展示样本量、方差膨胀因子与 精度保证“统计功效”之间的相互作用（**图5**）。这张图有助于评估在 特定研究设计中，共线性是否可能影响统计功效。由于此处的统计功效定义为精度估计（见方法），并以 N=1000 和 VIF=1时的 80% 为基准，因此该热图不受效应量影响。针对 $\beta_{\text{main}} \in \{0.5, 1, 1.5, 2\}$ 的传统检验效能的其他热图见补充图 **S7–S10**，可为应用研究提供更多参考。

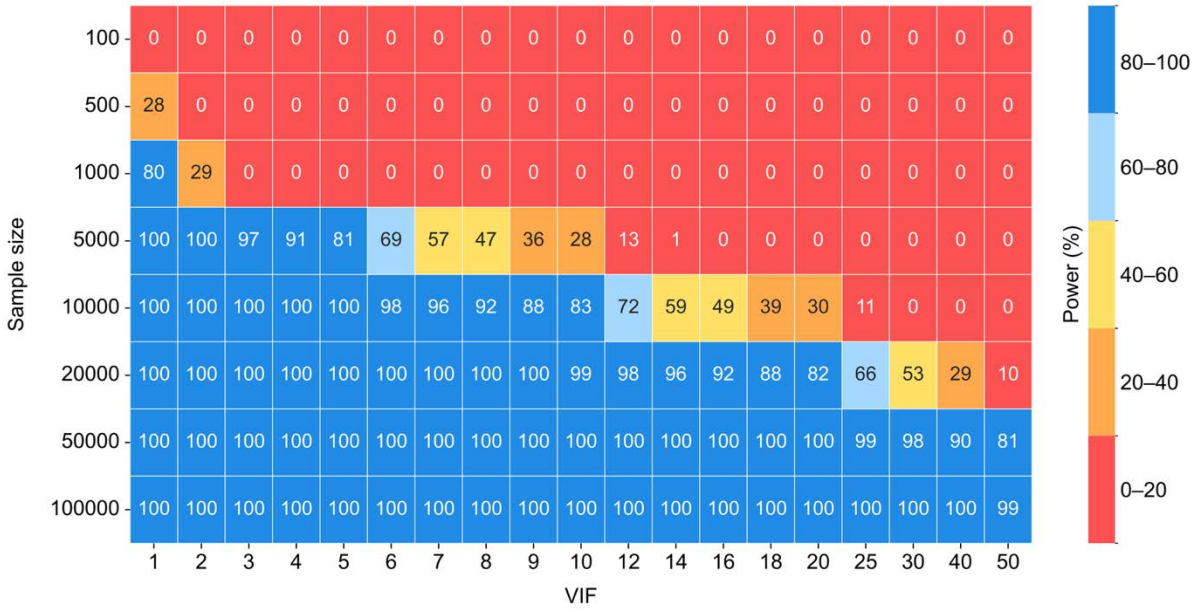

图5。显示 精度保证"统计功效" 在不同共线性水平和样本量下的热图。图中显示了 $\hat{\beta}_{\text{main}}$ 的平均精度保证"统计功效"（%），其取决于共线性（通过方差膨胀因子测量，横轴），其中 x_{main} 和x_1 存在部分共线性，以及样本量（纵轴）。

存在偏差时，共线性的影响会严重得多。为展示这种交互作用，我们对缺少 x_4（一个具有 $\beta_4 = 3$ 的显著预测变量）这一变量的数据集重复进行了线性回归分析，从而引入偏差。如**图S11–S13**所示，共线性会显著放大这种偏差，导致估计误差增大、覆盖率降低，功效指标大幅下降。无论是传统检验效能还是基于精度的统计功效，都会受到显著影响。例如，即便只有极轻微的共线性（VIF ≈ 1.1），也足以使存在偏差的数据集中的精度保证"统计功效"降至零（**图S12**）。这是因为偏差会同时使系数估计值偏离真实值，并扩大置信区间，降低 95%置信区间落在 $\beta_{\text{main}} \pm c_{\text{powval}}$ 范围内的可能性（**图 S13**）。

另一个值得注意的发现来自**图S11–S12**：省略相关预测变量以消除共线性，会导致系数估计值明显更差，而将相关预测变量纳入模型则不会如此。这些发现与Tay(19)的早期研究一致，进一步印证了这样一种观点：尽管共线性常被视为问题，但它的危害可能小于省略变量所引入的偏差。鉴于这些因素，研究人员在解读**图5** 时应保持谨慎，因为基于热图的指导是在假定不存在偏差的理想条件下得出的。在应用研究中，收集的数据可能存在偏差，共线性可能会放大这些偏差，导致统计失真程度超过我们基于无偏数据集进行模拟所揭示的情况。因此，固定

不应不考虑样本量和数据中潜在偏差来源，就应用方差膨胀因子阈值。

结论

我们开展了一项系统模拟研究，评估共线性对线性回归估计值的影响，样本量范围为 100 至 100,000，方差膨胀因子（VIF）范围为 1 至 50。共线性虽不会影响覆盖率或引入偏差，却会增大估计系数的方差，导致绝对误差增加、置信区间变宽，以及统计功效（传统方法或基于精度的方法）下降。重要的是，这些影响的程度高度取决于样本量。在极小的数据集中（N≈100），即使是轻度共线性（VIF<2），其水平远低于常用阈值，也已构成问题。相较之下，在大型数据集中（N≥50,000），即使存在极端共线性，系数估计值仍然可靠。

这些考量带来了几项实用建议。首先，研究人员应避免机械地套用固定的方差膨胀因子（VIF）阈值，因为其实际影响很大程度上取决于样本量。在小样本中，即使是轻度共线性（VIF < 2），也会显著降低估计精度；而在超大样本中，即使是极端共线性（VIF = 50），实际影响也微乎其微。其次，不应孤立地看待共线性：当存在偏差时（无论其源于遗漏变量、测量误差还是模型设定错误），共线性都可能显著放大其影响。第三，单纯为了降低方差膨胀因子（VIF）而剔除预测变量通常并不可取；这种做法往往会引入或加剧遗漏变量偏倚，从而使推断恶化。如果共线性仍令人担忧，相较于剔除变量，惩罚回归或预先设定变量集等替代建模策略可能更可取。最后，热图和指南的使用者应认识到，它们反映的是理想化的无偏条件；在实际应用中，除了样本量和方差膨胀因子（VIF），还必须权衡潜在的偏差来源。

致谢

本研究部分经费由DFM研究能力建设预算提供，项目名称为“技术与关怀：通过数据分析和患者在基层医疗中的声音改善患者预后” [NUHSRO/2022/049/NUSMed/DFM]。EK的部分经费由曼彻斯特英国心脏基金会卓越研究中心（CRE）提供。

参考文献

1. Harrell, Jr. FE.《回归建模策略：线性模型、逻辑回归、序数回归和生存分析的应用》。第2版。瑞士沙姆：施普林格国际出版公司；2015年。598页。（施普林格统计学系列）。 2. Belsley DA、Kuh E、Welsch RE。《回归诊断：识别影响性数据和共线性来源》。新泽西州霍博肯：John Wiley & Sons, Inc.；2004年。292页。 3. Dormann CF、Elith J、Bacher S、Buchmann C、Carl G、Carré G 等。共线性：处理共线性的方法综述及其性能评估模拟研究。《生态地理学》。2013年1月；36(1)：27–46。 4. Tu Y、Clerehugh V、Gilthorpe MS。线性回归中的共线性是口腔健康研究中的严重问题。《欧洲口腔科学杂志》。2004年10月；112(5)：389–97。 5. O’Brien RM。关于方差膨胀因子经验法则的注意事项。《质量与数量》。2007年9月11日；41(5)：673–90。 6. Kalnins A、Praitis Hill K。VIF得分：它有何用处？毫无用处。《组织研究方法》。2025年；28(1)：58–75。 7. Vatcheva KP、Lee M、McCormick JB、Rahbar MH。在流行病学研究中开展的回归分析中的多重共线性。《流行病学》。2016年；6(2)：227。 8. Shen L、Huang X、Liu Y、Bai S、Wang F、Yang Q。多序列磁共振成像影像组学结合影像特征预测高强度聚焦超声治疗子宫肌瘤的难度。《科学报告》。2025年1月25日；15(1)：3259。 9. Pinho CPS、Diniz ADS、De Arruda IKG、Leite APDL、Petribú MDMV、Rodrigues IG。用于估算内脏脂肪的预测模型：人体测量参数的贡献。Yang XF，编辑。《公共科学图书馆·综合》。2017年7月24日；12(7)：e0178958。 10. Sikkema L、Ramírez-Suástegui C、Strobl DC、Gillett TE、Zappia L、Madissoon E 等。健康与疾病状态下肺部的整合细胞图谱。《自然医学》。2023年6月；29(6)：1563–77。 11. Malacon K、Touponse G、Yoseph E、Li G、Wei P（Janet）、Kicielinski K 等。外科医生使用电子健康记录的性别差异。《美国医学会杂志·网络开放》。2024年7月23日；7(7)：e2421717。 12. Brog NA、Hegy JK、Berger T、Znoj H。年龄、动机和情绪调节技能可预测互联网自助干预对新冠肺炎相关心理困扰的治疗结果。《公共卫生前沿》。2022年6月9日；10：835356。 13. van der Wal JM、van Borkulo CD、Haslbeck JMB、Slofstra C、Klein NS、Blanken TF 等。减量停用抗抑郁药期间进行预防性认知疗法与维持性抗抑郁治疗对情感波动和个体情感网络的差异性影响，以及

对复发的影响：一项随机对照试验的二次分析。《eClinicalMedicine》。2023年12月；66：102329。 14. Cho JH、Rhee EJ、Park SE、Kwon H、Jung JH、Han KD 等。腰围与21,749,261名韩国成年人的心肌梗死和缺血性卒中风险：一项全国性人群研究。《糖尿病与代谢杂志》。2019年；43(2)：206–21。 15. Adewoye KB、Rafiu AB、Aminu TF、Onikola IO。研究多重共线性对线性回归估计值的影响。《马来西亚计算机杂志》。2021年；6(1)：698–714。 16. Mason CH、Perreault WDJr。多重回归分析中的共线性、统计功效与解释。《营销研究杂志》。1991年8月；28(3)：268–80。 17.《SAS/STAT® 14.1 用户指南》。检验效能与样本量分析简介（第372页）。北卡罗来纳州卡里：SAS Institute Inc.；2015年。 18.《 Stata 19 检验效能、精度与样本量参考手册》（第 703 页）。德克萨斯州大学城：Stata Press；2025年。 19. Tay R。回归模型中的相关性、方差膨胀与多重共线性。《东亚交通研究学会杂志》。2017年；12：2006–15。

1. 补充方法：数据集的生成

S1.1 生成相关数据集的一般流程**S1.2** 相关矩阵的生成
S1.3 方程1的推导（补充信息中的 S5）

Figure S1. $\hat{\beta}_{\text{main}}$ 的平均偏差随共线性的变化情况，共线性通过方差膨胀因子（VIF）测量，其中 x_{main}和 x_1 部分共线。每种颜色代表每个情景（即每种 VIF/样本量组合）的不同模拟次数，每个情景的样本量固定为 100。阴影区域表示均值估计的标准误。

Figure S2. 样本量为 1,000 时，$\hat{\beta}_{\text{main}}$ 的平均覆盖率（%）、平均绝对误差和平均偏差随共线性的变化情况，共线性通过方差膨胀因子（VIF）测量，其中 x_{main} 和 x_1 部分共线。浅珊瑚色实线表示 $\beta_{\text{main}} = 0$，黑色虚线表示 $\beta_{\text{main}} = 2$。

图 S3。 对于 $\hat{\beta}_{\text{main}}$，随着通过 VIF 测量的共线性变化而变化的平均覆盖率（%）、平均绝对误差和平均偏差，其中所有预测变量均存在部分共线性。不同颜色对应不同的样本量，从 100（深红色）到 100,000（深蓝色）不等。阴影区域表示均值估计值周围的标准误。

图 S4。对于 $\hat{\beta}_{\text{main}}$，随着通过 VIF 测量的共线性变化而变化的平均统计功效（%）、平均 95%置信区间宽度和平均标准误（SE），其中所有预测变量均存在部分共线性。不同颜色对应不同的样本量，从 100（深红色）到 100,000（深蓝色）不等。阴影区域表示均值估计值周围的标准误。

图 S5。 对于 $\hat{\beta}_{\text{main}}$，随着通过 VIF 测量的共线性变化而变化的平均覆盖率（%）、平均绝对误差和平均偏差。模拟使用 20 个自变量进行，这些变量均存在部分共线性。不同颜色代表不同的样本量，从 100（深红色）到 100,000（深蓝色）不等。阴影区域表示均值估计值周围的标准误。

图 S6。$\hat{\beta}_{\text{main}}$ 的平均统计功效（%）、平均 95%置信区间宽度和平均标准误（SE）随共线性（通过 VIF 测量）的变化情况。模拟采用了 20 个自变量，且这些变量均存在部分共线性。每种颜色代表一种

不同的样本量，范围从 100（深红色）到 100,000（深蓝色）。阴影区域表示均值估计的标准误。

图7。$\hat{\beta}_{\text{main}}$ 的平均传统检验效能（%），随共线性（以方差膨胀因子（横轴）衡量）和样本量（纵轴）变化；其中x_{main} 和x_1 存在部分共线性，且$\boldsymbol{\beta}_{\mathbf{main}} = \mathbf{2}$。 $\mathbf{0}$。传统检验效能定义为置信区间不包含 0 的模拟所占的比例。

图8。$\hat{\beta}_{\text{main}}$ 的平均传统检验效能（%），随共线性（以方差膨胀因子（横轴）衡量）和样本量（纵轴）变化；其中x_{main} 和x_1 存在部分共线性，且$\boldsymbol{\beta}_{\mathbf{main}} = \mathbf{1}$。 $\mathbf{5}$。传统检验效能定义为置信区间不包含 0 的模拟所占的比例。

图 9。$\hat{\beta}_{\text{main}}$ 的平均传统检验效能（%），随共线性（以方差膨胀因子（横轴）衡量）和样本量（纵轴）变化；其中x_{main} 和x_1 存在部分共线性，且$\boldsymbol{\beta}_{\mathbf{main}} = \mathbf{1}$。 $\mathbf{0}$。传统检验效能定义为置信区间不包含 0 的模拟所占的比例。

图 10。$\hat{\beta}_{\text{main}}$ 的平均传统检验效能（%）与共线性（通过方差膨胀因子测量，横轴）的关系，其中x_{main} 和x_1 存在部分共线性；纵轴为样本量，$\boldsymbol{\beta}_{\mathbf{main}} = \mathbf{0}$。 $\mathbf{5}$。传统检验效能定义为置信区间不包含 0 的模拟所占比例。

图 S11。 $\hat{\beta}_{\text{main}}$ 的平均覆盖率（%）、平均绝对误差、平均95%置信区间宽度和平均偏差与共线性（通过方差膨胀因子测量）的关系。数据集由六个预测变量均存在部分共线性的数据生成，随后省略一个预测变量（x_4），从而在数据集中引入偏差。每种颜色代表不同的样本量，范围从 100（深红色）到 100,000（深蓝色）。

图S12。 $\hat{\beta}_{\text{main}}$ 的平均统计功效（%）与共线性（通过方差膨胀因子测量，横轴）及样本量（纵轴）的关系。数据集由六个预测变量均存在部分共线性的数据生成，随后省略一个预测变量（x_4），从而在数据集中引入偏差。

图 S13。 偏差导致检验效能损失的示意图。对于无偏数据集，$\hat{\beta}_{\text{main}}$ （蓝色曲线）的95%置信区间（CI） 位于 $\beta_{\text{main}} \pm c_{\text{powval}}$范围内，该范围定义为具有统计功效。然而，引入偏差后，$\hat{\beta}_{\text{main}}$ 的估计值会偏离真实值，且置信区间更宽（红色曲线）。这两种效应都会导致$\hat{\beta}_{\text{main}}$ 的 95%置信区间落在 $\beta_{\text{main}} \pm c_{\text{powval}}$ 范围之外，从而造成检验效能损失。

表 S1。使用 20 个自变量进行模拟时所用的真实系数。

补充信息

共线性和样本量对线性回归结果的影响：一项模拟研究

Stephanie C. C. van der Lubbe[1,2], Jose M. Valderas[1,2,3,4], Evangelos Kontopantelis[1,5]

[1] 新加坡国立大学 Yong Loo Lin医学院家庭医学部，新加坡[2] 新加坡国立大学卫生系统新加坡国立大学综合诊所，新加坡[3] 新加坡国立大学卫生系统家庭医学系，新加坡[4] 新加坡国立大学卫生系统绩效研究中心 (CRiHSP)，新加坡国立大学，新加坡[5] 曼彻斯特大学生物、医学与健康学院信息学、影像与数据科学部，英国

S1. 补充方法：数据集的生成

S1.1 生成相关数据集的一般流程 数据集按以下流程生成：

步骤 1：生成相关矩阵
生成相关矩阵时，会针对每个预测变量设定目标方差膨胀因子。有关如何确定非对角线值的方法学细节，请参阅下一页的*S2* 节：相关矩阵的生成。

步骤 2：根据相关矩阵计算乔列斯基因子。乔列斯基因子是基于步骤 1 中的相关矩阵，使用 Python（版本 3.11.5）中的 NumPy 软件包（版本 1.26.4）的 linalg 模块计算得到的：

```
np.linalg.cholesky(相关性_矩阵)
```

步骤 3：生成包含随机值的 $N \times X$ 数组
通过使用 NumPy 包中的随机模块，从标准正态分布中抽取相互独立的值，生成了形状为 $N \times X$ 的数组：

```
np.random.normal(size=(N, 6))
```

每一列对应 $X = [x_{\text{main}}, x_1, x_2, x_3, x_4, x_5]$中的一个预测变量，每一行代表一个观测值，总计有 N 个观测值。对于每种情景，我们使用相同的一组随机种子（从 0 到 999），先生成完全相同的 1,000 $N \times X$ 个数组，再引入所需程度的共线性。这种方法可确保观测到的估计系数变化完全归因于共线性的影响，而非底层数据中的随机差异。

步骤 4：使用乔列斯基因子进行变换
随后，将 $N \times X$ 数组与乔列斯基因子相乘。此操作会将最初相互独立的变量转换为一组具有所需相关结构的新预测变量，从而引入预期的共线性。

第 5 步：生成误差项 ε
误差项 ε 取自一个均值为 0、方差为 $\pi^2/3$ 的正态分布，使用 NumPy 软件包中的随机模块生

```
np.random.normal(0, np.pi / np.sqrt(3), N)
```

第 6 步：计算结果 y使用以下线性回归方程计算结果 y：

$$y = \beta_0 + \beta_{\text{main}} x_{\text{main}} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon \qquad \text{Eq. S1}$$

真实系数设为：

$$\beta_0 = 10, \beta_{\text{main}} = 2, \beta_1 = 1.3, \beta_2 = 1.5, \beta_3 = 6, \beta_4 = 3, \text{and } \beta_5 = 1.$$

步骤 3 至 6 在每种情景（即样本量和方差膨胀因子的每种组合）下重复 1,000 次，最终得到 1,000 个维度为 $N \times 8$ 的数据集，其列分别表示 x_{main}、x_1、x_2、x_3、x_4、x_5、ε 和 y。

S1.2 相关矩阵的生成

为引入预期水平的共线性，我们生成了与目标方差膨胀因子（VIF）值对应的相关矩阵。对于包含六个变量且仅有 x_{main} 和 x_1 部分共线的数据集，6×6 相关矩阵 $\mathbf{R}_{\text{corr}}$ 的形式如下：

$$\mathbf{R}_{\text{corr}} = \begin{bmatrix} 1 & r & 0 & 0 & 0 & 0 \\ r & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

皮尔逊相关系数 r 直接根据方差膨胀因子（VIF）的定义推导得出：

$$\text{VIF}(x_i) = \frac{1}{1 - R_i^2} = \frac{1}{1 - \rho^2_{x_{\text{main}}, x_1}} = \frac{1}{1 - r^2} \qquad \text{Eq. S2}$$

其中，$\rho_{x_{\text{main}},\ x_1}$是 x_{main} 和 x_1之间的皮尔逊相关系数，对应于相关矩阵中的非对角元素 r。公式 S2 可重新排列，得到 r：

$$r = \sqrt{1 - \frac{1}{\text{VIF}(x_i)}} \qquad \text{Eq. S3}$$

F对于所有预测变量均部分相关的数据集，相关矩阵为

$$\mathbf{R}_{\text{corr}} = \begin{bmatrix} 1 & r & r & r & r & r \\ r & 1 & r & r & r & r \\ r & r & 1 & r & r & r \\ r & r & r & 1 & r & r \\ r & r & r & r & 1 & r \\ r & r & r & r & r & 1 \end{bmatrix}$$

对于该等相关矩阵，所有非对角元素的值均为 r，我们使用以下解析表达式求解 r：

$$r = \frac{R_i^2(p-2) + \sqrt{R_i^4(p-2)^2 + 4R_i^2(p-1)}}{2(p-1)} \qquad \text{Eq. S4}$$

其中 p 是预测变量的数量，而 R_i^2 是将一个预测变量对其他 $p-1$ 个预测变量进行回归时的决定系数。下一节将推导该方程（*S3* 节：方程 *S4* 的推导）。

以下列出与各目标 VIF 值对应的 6x6 相关矩阵的非对角线值，格式为 (VIF, r)：

```
(1.0, 0.000), (1.1, 0.176), (1.2, 0.261), (1.3, 0.326), (1.4, 0.380),
(1.5, 0.424), (1.6, 0.462), (1.7, 0.496), (1.8, 0.525), (1.9, 0.551),
(2.0, 0.574), (2.1, 0.595), (2.2, 0.614), (2.3, 0.631), (2.4, 0.647),
(2.5, 0.661), (2.6, 0.674), (2.7, 0.687), (2.8, 0.699), (2.9, 0.709),
(3.0, 0.719), (3.1, 0.728), (3.2, 0.737), (3.3, 0.745), (3.4, 0.752),
(3.5, 0.759), (3.6, 0.766), (3.7, 0.773), (3.8, 0.779), (3.9, 0.785),
(4.0, 0.790), (4.1, 0.795), (4.2, 0.800), (4.3, 0.804), (4.4, 0.809),
(4.5, 0.814), (4.6, 0.818), (4.7, 0.821), (4.8, 0.825), (4.9, 0.829),
(5.0, 0.832), (5.2, 0.839), (5.4, 0.845), (5.6, 0.850), (5.8, 0.856),
(6.0, 0.860), (6.2, 0.865), (6.4, 0.869), (6.6, 0.873), (6.8, 0.877),
(7.0, 0.880), (7.2, 0.884), (7.4, 0.887), (7.6, 0.890), (7.8, 0.893),
(8.0, 0.895), (8.2, 0.898), (8.4, 0.900), (8.6, 0.903), (8.8, 0.905),
(9.0, 0.907), (9.2, 0.909), (9.4, 0.911), (9.6, 0.913), (9.8, 0.915),
(10,  0.916), (11,  0.924), (12,  0.931), (13,  0.936), (14,  0.941),
(15,  0.944), (16,  0.948), (17,  0.951), (18,  0.953), (19,  0.956),
(20,  0.958), (25,  0.967), (30,  0.972), (35,  0.976), (40,  0.979),
(45,  0.982), (50,  0.983)
```

S1.3 方程*1*的推导（补充信息中的 *S5*）

为推导方程 S4 中 r 的解析表达式，我们从复相关系数的平方 R_i^2 的定义出发。该系数用于量化以一个预测变量对其余 $p-1$ 个预测变量进行回归时所解释的方差比例：

$$R_i^2 = \vec{r}^{\mathsf{T}}\mathbf{R}^{-1}\vec{\boldsymbol{r}} \qquad \text{Eq. S5}$$

其中，$\vec{r}$ 是感兴趣的预测变量与其余 $p-1$ 个预测变量之间的相关性向量，$\mathbf{R}$ 是这些 $p-1$ 个预测变量之间的相关矩阵。[S1] 向量 $\boldsymbol{r}$ 可改写为 $r \cdot 1\vec{\boldsymbol{}}$ 的形式，其中 $1\vec{\boldsymbol{}}$ 是长度为 $p-1$ 的全1向量：

$$R_i^2 = r^2 \cdot \vec{1}^{\mathsf{T}}\mathbf{R}^{-1}\vec{1} \qquad \text{Eq. S6}$$

相关矩阵的逆矩阵为

$$\mathbf{R}^{-1} = \frac{1}{1-r}\left(\mathbf{I}_{p-1} - \frac{r}{1+(p-2)r}\mathbf{J}_{p-1}\right) \qquad \text{Eq. S7}$$

其中，$\mathbf{I}_{p-1}$ 是单位矩阵，$\mathbf{J}_{p-1} = 1\vec{\boldsymbol{}}_{p-1} \cdot 1_{p-1}^{\mathsf{T}}$ 是一个全1矩阵。[S2] 将方程 S7 代入方程 S6，可得：

$$R_i^2 = r^2 \cdot \vec{1}^{\mathsf{T}}\left[\frac{1}{1-r}\left(\mathbf{I}_{p-1} - \frac{r}{1+(p-2)r}\mathbf{J}_{p-1}\right)\right]\vec{1} \qquad \text{Eq. S8}$$

通过进行以下代数化简：

- $\vec{1}_{p-1}^{\mathsf{T}}\mathbf{I}_{p-1}\vec{1}_{p-1} = 1_{p-1}^{\mathsf{T}}\vec{1}_{p-1} = p-1$
- $\vec{1}_{p-1}^{\mathsf{T}}\mathbf{J}_{p-1}\vec{1}_{p-1} = \vec{1}_{p-1}^{\mathsf{T}}\vec{1}_{p-1} \cdot 1_{p-1}^{\mathsf{T}}\vec{1}_{p-1} = (p-1)^2$

公式 8 可改写为：

$$R_i^2 = \frac{r^2}{1-r}\left((p-1) - \frac{r(p-1)^2}{1+(p-2)r}\right) \quad \text{Eq. S9}$$

进一步简化后可得：

$$R_i^2 = \frac{(p-1)r^2}{1+(p-2)r} \quad \text{Eq. S10}$$

将方程 S10 改写为二次形式 $ar^2 + br + c = 0$，并使用求根公式求解r，即可得到用于计算等相关矩阵中矩阵元素r 的最终解析方程：

$$r = \frac{R_i^2(p-2) + \sqrt{R_i^4(p-2)^2 + 4R_i^2(p-1)}}{2(p-1)}$$

补充参考文献

S1. Rencher AC。《多变量分析方法》。纽约：John Wiley & Sons, Inc.；1995 年，第 361、394–395 页。
S2. Gruber, MHJ。《线性模型矩阵代数》。新泽西州霍博肯：John Wiley & Sons, Inc.；2014 年。第 49 页

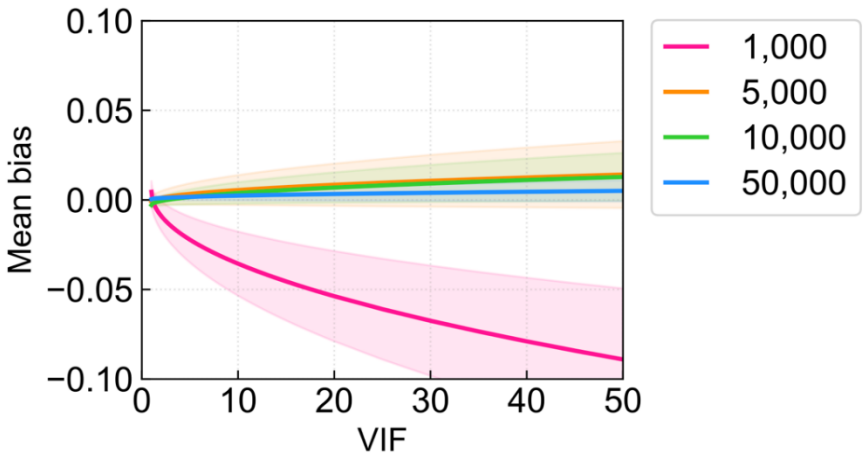

图 S1 | $\hat{\beta}_{\text{main}}$ 的平均偏差随共线性（通过方差膨胀因子 VIF 测量）的变化，其中 x_{main} 和 x_1 存在部分共线性。每种颜色代表每个情景（即每种 VIF/样本量组合）的不同模拟次数，每个情景的样本量固定为 100。阴影区域表示均值估计的标准误。

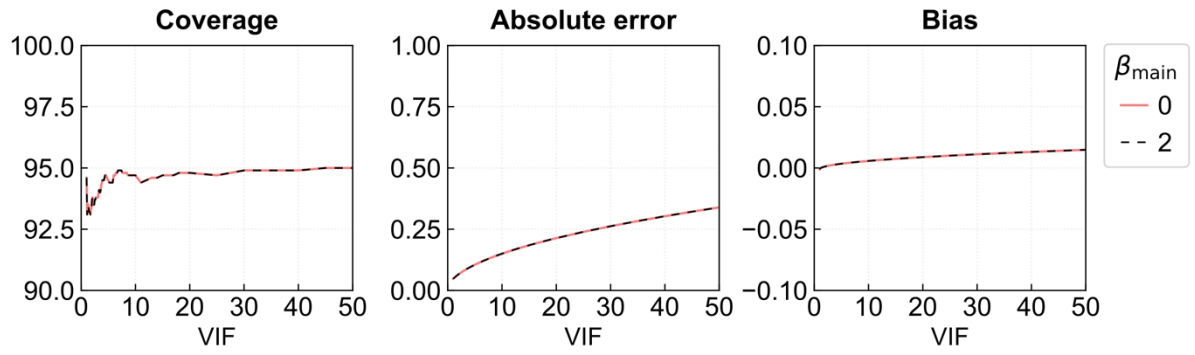

图 S2 | $\hat{\beta}_{\text{main}}$ 的平均覆盖率（%）、平均绝对误差和平均偏差随共线性（通过方差膨胀因子测量）的变化，其中 x_{main} 和 x_1 存在部分共线性，样本量为 1,000。浅珊瑚色实线表示 $\beta_{\text{main}} = 0$，黑色虚线表示 $\beta_{\text{main}} = 2$。

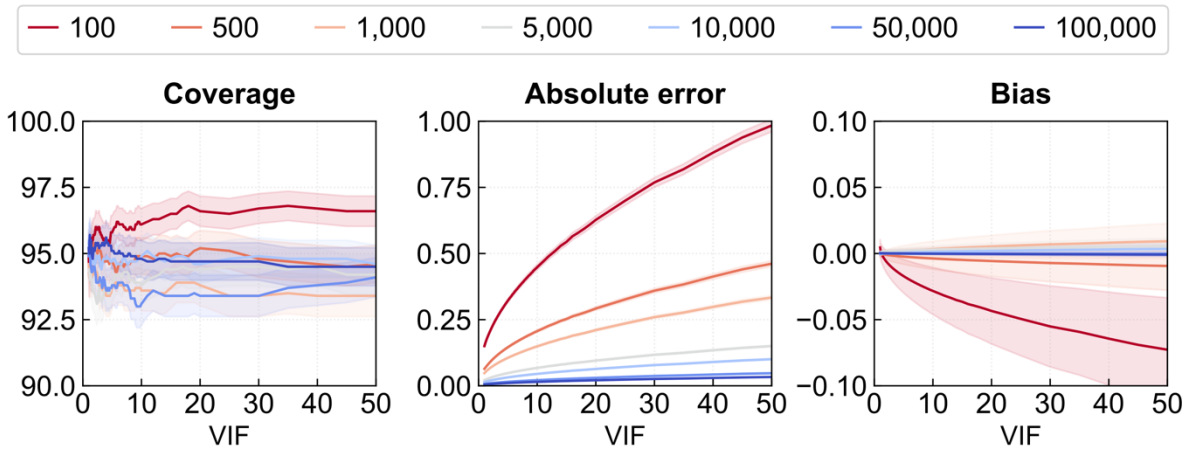

图 S3 | 当所有预测变量均存在部分共线性时，$\hat{\beta}_{\text{main}}$ 的平均覆盖率（以百分比表示）、平均绝对误差和平均偏差随共线性（通过 VIF 测量）的变化。每种颜色对应一个不同的样本量，范围从 100（深红色）到 100,000（深蓝色）。阴影区域表示均值估计周围的标准误。

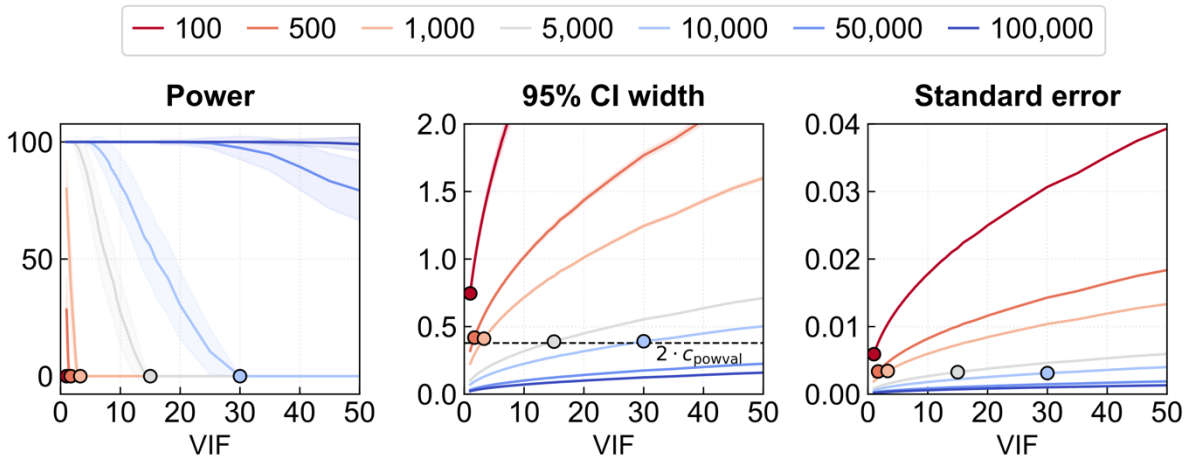

图 S4 | 当所有预测变量均存在部分共线性时，$\hat{\beta}_{\text{main}}$ 的平均统计功效（以百分比表示）、平均 95%置信区间宽度和平均标准误（SE）随共线性（通过 VIF 测量）的变化。每种颜色对应一个不同的样本量，范围从 100（深红色）到 100,000（深蓝色）。阴影区域表示均值估计周围的标准误。

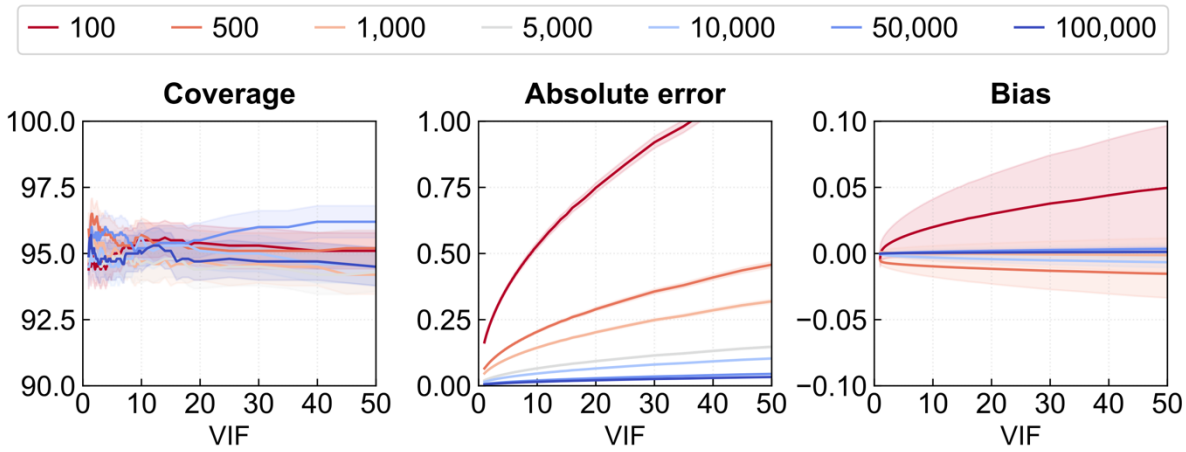

图 S5 | 对于 $\hat{\beta}_{\text{main}}$，平均覆盖率（%）、平均绝对误差和平均偏差随共线性（通过方差膨胀因子测量）的变化。模拟采用了 20 个自变量，且这些变量均存在部分共线性。每种颜色代表一个不同的样本量，范围从 100（深红色）到 100,000（深蓝色）。阴影区域表示均值估计周围的标准误。

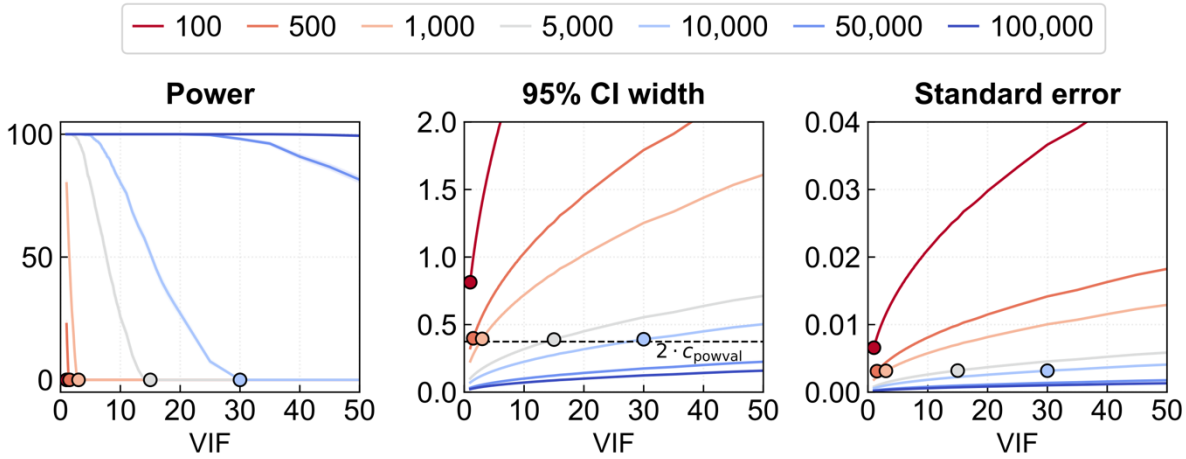

图 S6 | 对于 $\hat{\beta}_{\text{main}}$，平均统计功效（%）、平均 95%置信区间宽度和平均标准误（SE）随共线性（通过方差膨胀因子测量）的变化。模拟采用了 20 个自变量，且这些变量均存在部分共线性。每种颜色代表一个不同的样本量，范围从 100（深红色）到 100,000（深蓝色）。阴影区域表示均值估计周围的标准误。

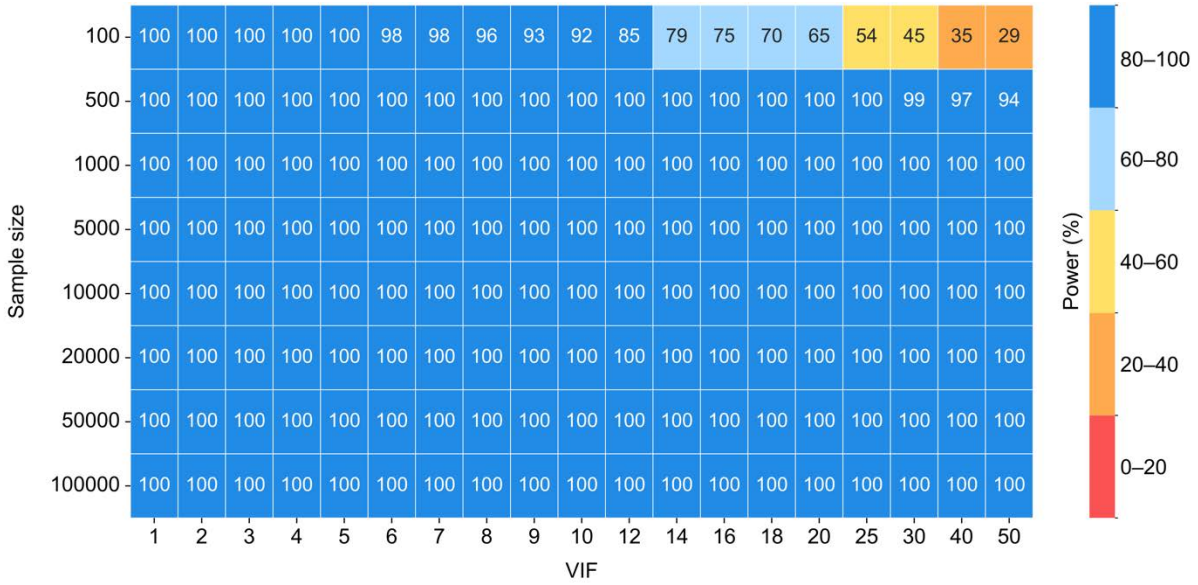

图7 | $\hat{\beta}_{\text{main}}$ 的平均传统检验效能（%），随共线性程度（通过方差膨胀因子测量；横轴）和样本量（纵轴）的变化而变化，其中x_{main}和x_1部分共线，$\boldsymbol{\beta}_{\text{main}} = \mathbf{2}_{\circ}\ \mathbf{0}$。传统检验效能定义为置信区间不包含 0 的模拟所占比例。

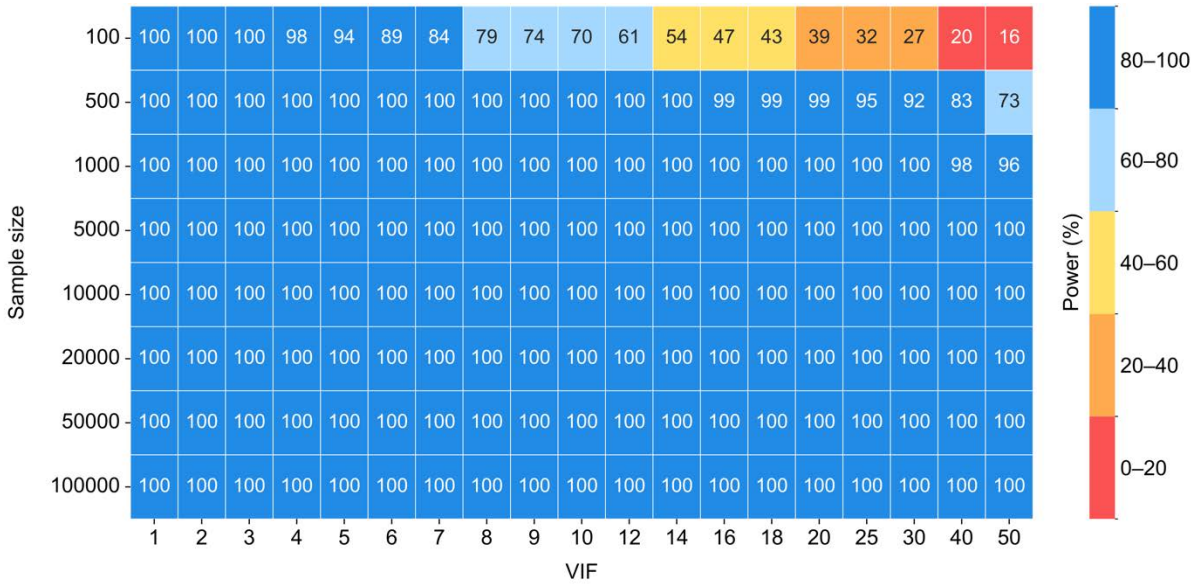

图8 | $\hat{\beta}_{\text{main}}$ 的平均传统检验效能（%），随共线性程度（通过方差膨胀因子测量；横轴）和样本量（纵轴）的变化而变化，其中x_{main}和x_1部分共线，$\boldsymbol{\beta}_{\text{main}} = \mathbf{1}_{\circ}\ \mathbf{5}$。传统检验效能定义为置信区间不包含 0 的模拟所占比例。

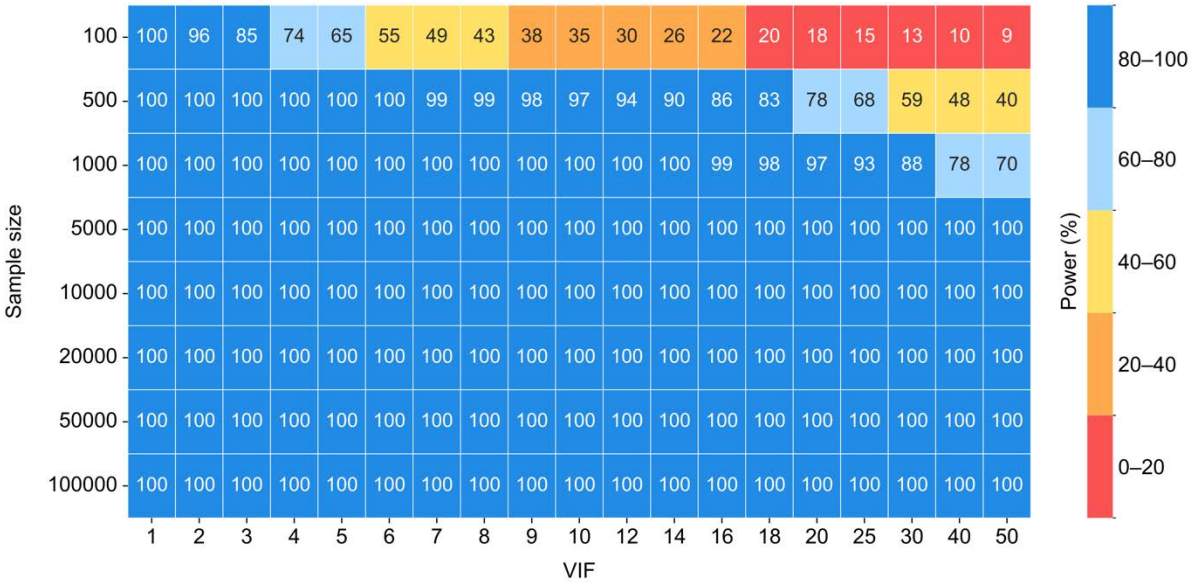

图9 | $\hat{\beta}_{\text{main}}$ 的平均传统检验效能（%），随共线性（通过方差膨胀因子测量，横轴）和样本量（纵轴）的变化情况；其中x_{main}和x_1 存在部分共线性，且$\boldsymbol{\beta}_{\text{main}} = \mathbf{1.0}$。传统检验效能定义为置信区间不包含 0 的模拟所占比例。

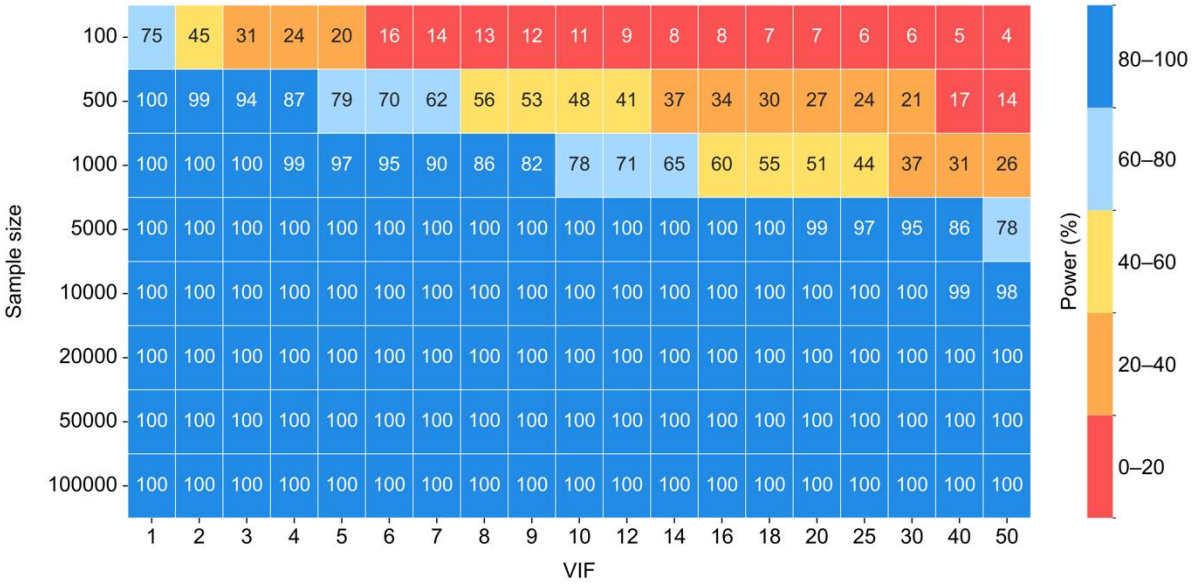

图10 | $\hat{\beta}_{\text{main}}$ 的平均传统检验效能（%），随共线性（通过方差膨胀因子测量，横轴）和样本量（纵轴）的变化情况；其中x_{main}和x_1 存在部分共线性，且$\boldsymbol{\beta}_{\text{main}} = \mathbf{0.5}$。传统检验效能定义为置信区间不包含 0 的模拟所占比例。

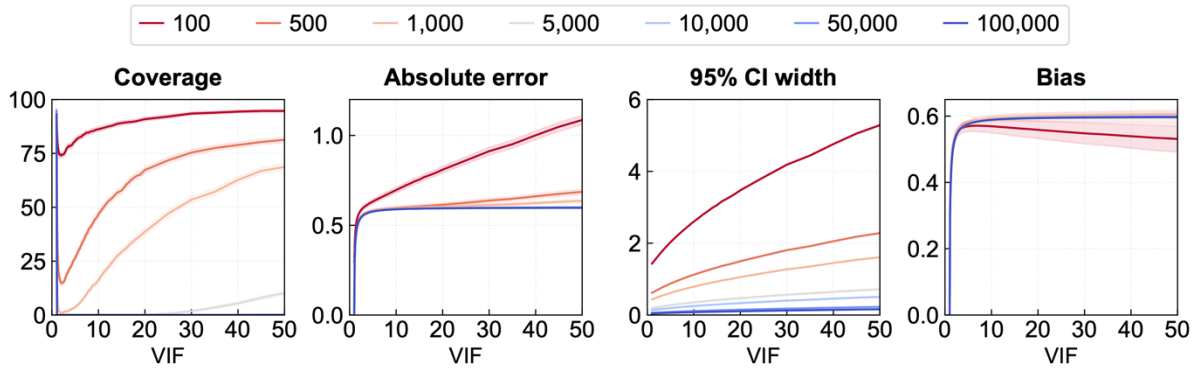

图 S11 | 的平均覆盖率（%）、平均绝对误差、平均 95%置信区间宽度和平均偏差随共线性变化的情况，共线性通过方差膨胀因子测量。数据集生成时，六个预测变量均存在部分共线性，随后省略一个预测变量（x_4），从而在数据集中引入偏差。每种颜色代表不同的样本量，范围从 100（深红色）到 100,000（深蓝色）。

如图所示，**覆盖率** 最初会下降，直到约 VIF=2，随后随着共线性增加而上升。这一趋势的出现是因为偏差导致 $\hat{\beta}_{\text{main}}$ 中出现较大的**绝对误差**（VIF>1.7 时约为 >0.5），这意味着估计系数与真实值 $\beta_{\text{main}} = 2$ 相差甚远。虽然**95% 置信区间宽度** 也会随着共线性增加，但其增幅低于绝对误差，导致在中等 VIF 水平下，真实系数更频繁地落在置信区间之外。当置信区间宽度扩大到足够程度后，真实系数更常落在置信区间内，从而提高覆盖率。共线性条件下，较小样本量的覆盖率较高，因为它们产生的置信区间更宽，使真实系数更有可能落在其中。**偏差**迅速上升，在 VIF=2 附近达到约 0.5，随后收敛至 ~0.57。值得注意的是，除 N=100 的偏差略低外，所有样本量下的偏差都相同。不过，随着每种情景的模拟次数增加，N=100 的偏差会收敛到与较大样本量相同的水平，这一点在图 S1 中也有所体现。

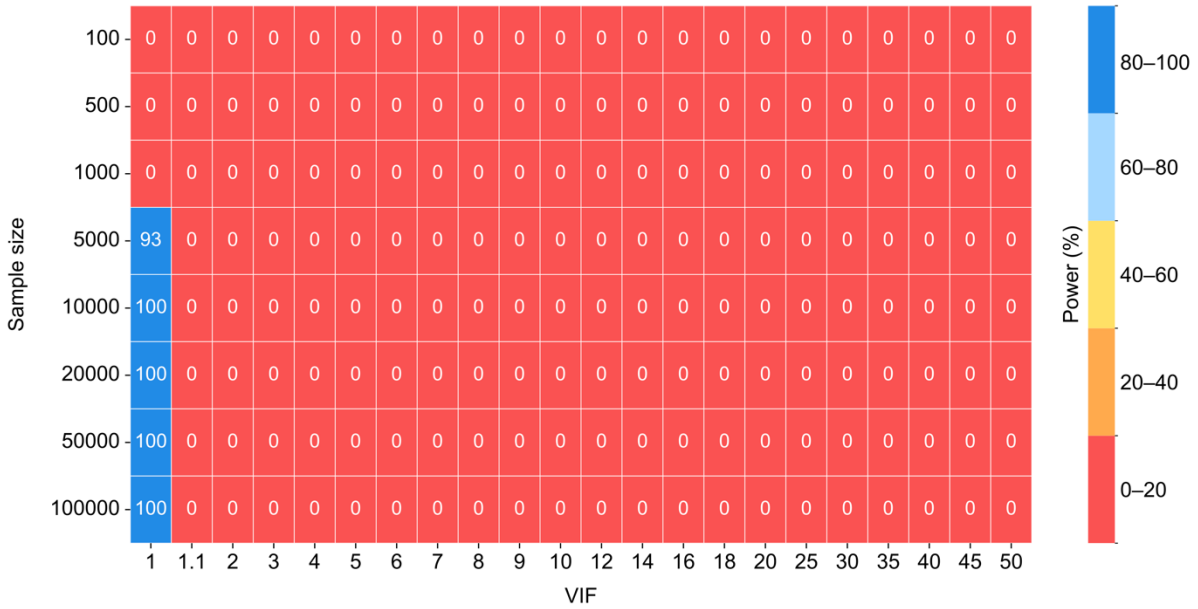

图 S12 | $\hat{\beta}_{\text{main}}$ 的平均功效（%），随通过 VIF 测量的共线性（横轴）和样本量（纵轴）变化。数据集生成时，六个预测变量均存在部分共线性，随后省略了一个预测变量（x_4），从而在数据集中引入偏差。

模拟仅在样本量达到 5000 及以上且不存在共线性时才表现出统计功效。一旦引入共线性（VIF=1.1），统计功效便立即降至零。统计功效下降有两个原因。首先，偏差会使置信区间变宽（图 S13），从而增加 95% 置信区间未落在 $\beta_{\text{main}} \pm c_{\text{powval}}$ 内的可能性。其次，数据集中的偏差会导致估计系数出现偏差 $\hat{\beta}_{\text{main}}$（图 S11），使估计值偏离 $\beta_{\text{main}} \pm c_{\text{powval}}$ 范围。图 S13 以示意图形式展示了这一效应。

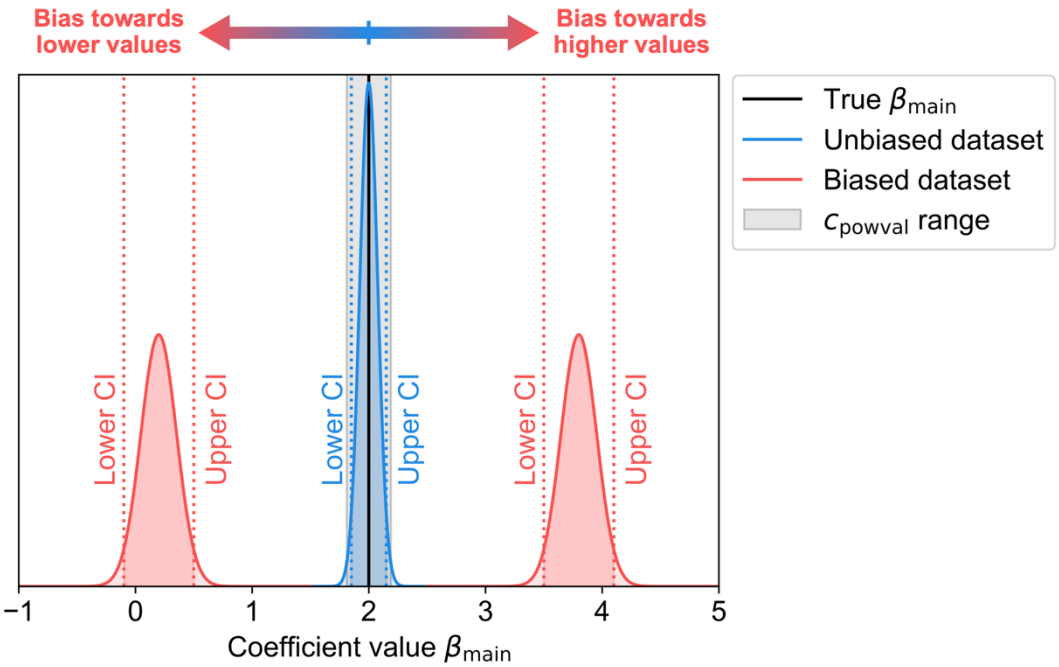

图 S13 | 偏差导致检验效能损失的示意图。对于无偏数据集，$\hat{\beta}_{main}$ 的 95%置信区间（CI）（蓝色曲线）落在 $\beta_{main} \pm c_{powval}$ 范围内，该范围定义为具有统计功效。然而，引入偏差后，$\hat{\beta}_{main}$ 的估计值偏离真实值，且置信区间更宽（红色曲线）。这两种效应都会使$\hat{\beta}_{main}$ 的 95%置信区间落在 $\beta_{main} \pm c_{powval}$ 范围之外，从而导致检验效能损失。

该图还展示了偏差如何导致真实效应被低估或高估，从而可能引发第一类错误和第二类错误。

表 S1 | 真实系数 模拟中使用的系数，包含 20 个相互独立的变量 变量。

系数	值
截距	10.0
β_{main}	2.0
β_1	1.3
β_2	1.5
β_3	6.0
β_4	3.0
β_5	1.0
β_6	6.6
β_7	0.7
β_8	3.1
β_9	2.6
β_{10}	7.5
β_{11}	6.9
β_{12}	9.0
β_{13}	1.3
β_{14}	4.5
β_{15}	0.8
β_{16}	2.6
β_{17}	5.3
β_{18}	0.8
β_{19}	2.4