arXiv:2302.05612v1 [stat.ME] 11 Feb 2023

针对稀疏观测多元函数型数据的基于投影的双样本推断

Salil Koner* 和 Sheng Luo

美国北卡罗来纳州达勒姆杜克大学生物统计与生物信息学系salil.koner@duke.edu,
sheng.luo@duke.edu

Summary

现代纵向研究会收集多个结局作为主要终点，以了解疾病的复杂动态。通常，尤其是在临床试验中，多维响应之间的联合变异对于评估两个或多个组之间的差异特征具有重要作用，而非仅基于单一结局进行推断。我们将纵向设计纳入稀疏观测函数型数据的框架，开发了一种基于投影的双样本显著性检验，用于识别典型多元轮廓之间的差异。该方法建立在广泛采用的多元函数主成分分析基础上，在降低无限维多模态函数维数的同时保留各成分之间的动态相关性。该检验适用于响应的一大类（非平稳）协方差结构，并基于单个p值检测显著的组间差异，从而克服了分别比较各成分均值时需要调整多个p值的问题。有限样本数值研究表明，与最先进的检验程序相比，该检验能够控制第一类错误，并有较强的能力检测显著的组间差异。我们将该检验应用于采用纵向设计的TOMMORROW研究，研究对象为因阿尔茨海默病而处于轻度认知障碍高风险中的个体，以检测吡格列酮组与安慰剂组在认知测验得分方面的差异。

Key words: 阿尔茨海默病；多变量纵向数据；TOMMORROW 试验，稀疏函数型数据；

1. Introduction

双样本推断问题在许多生物统计学应用中十分常见，例如药物发现，以及评估目标药物相较于安慰剂或标准治疗的疗效。为刻画阿尔茨海默病（AD）这一复杂的慢性疾病，通常会收集多变量纵向数据。在随机对照试验（RCT）中，有必要检验治疗对多个纵向结局的效果。例如，

*Salil Koner，电子邮件：salil.koner@duke.edu

© 作者 2023。由牛津大学出版社出版。保留所有权利。如需获得许可，请发送电子邮件至：journals.permissions@oup.com

在 TOMMORROW研究中，研究人员收集了多项认知测评量表分数，以评估吡格列酮能否延缓阿尔茨海默病（AD）所致轻度认知障碍的发生（Burns等，2019，2021）。传统方法会汇总每位受试者的多项测量结果，得到单一综合评分，再基于该评分进行双样本推断问题。这类基于汇总的方法完全忽略了针对每位受试者收集的多维终点之间的动态变异，因此无法量化研究药物对某一终点的影响如何作用于其他终点。部分方法还会进一步简化推断问题，只关注基线与研究最后一次访视之间结局的变化，从而完全忽略结局的纵向特性（Kieburtz 等，2015）。因此，能够同时考虑与疾病进展相关的多个终点之间的这种联合变异以及结局纵向特性的推断程序，对于有力地检测显著组间差异至关重要。

本文聚焦于对重复测量的多维结局进行双样本推断问题。在TOMMORROW研究这一启发性研究背景中，一组处于轻度认知障碍（MCI）高风险的受试者会在个体层面随时间反复填写问卷，自我报告认知表现得分，共十二项。这十二项得分对应五个主要认知领域，即情景记忆、知觉记忆、语义记忆、工作记忆和视空间记忆。研究中的受试者每日口服0.8 mg吡格列酮或安慰剂。研究旨在根据纵向测量的、对应五个主要认知领域的12元认知得分，评估吡格列酮在延缓高风险认知正常受试者发生由阿尔茨海默病（AD）导致的MCI方面的疗效。研究为期五年，预计每六个月随访一次。研究设计的更多细节见第6节及Burns等人（2021）。在现代纵向研究中，受试者通常自行安排填写问卷的时间；因此，以基线为起点测得的响应报告时间点可能因受试者而异。

我们在多元函数型数据（mv-fD）框架下研究双样本推断问题，其中响应是受试者在五年期间观测到的多维认知得分轨迹。传统的混合模型框架通过对时间施加线性或二次等结构假设来量化时间依赖性；与之不同，函数型数据分析（fDA）方法假定响应轨迹的均值和协方差具有平滑性，从而放宽了这些结构条件。然而，现有的多元函数型数据推断程序主要关注函数被密集观测的情形，即轨迹得到完整观测的情形。G´orecki 和 Smaga（2017）首次研究了这一问题，他们将每条个体函数轨迹展开为一组基函数，并根据基表示的随机系数构造了若干类多元方差分析（MANOVA）型检验统计量。Qiu 等人 （2021）采用逐点 Hotelling T平方统计量检验组间无差异的原假设。Jiang 等人 （2017）考察了两组样本经验特征函数之间的基于能量的距离。Krzy´sko 和 Smaga（2021）通过考察基展开系数的特征函数之间的差异，构造了符合性检验。Qu 等人 （2021）基于多元统计深度函数，提出了用于多元函数型数据双样本推断的稳健方差分析（Claeskens等人，2014）。这些程序要求在密集或超密集设计下对个体函数进行完整观测，因此无法应用于纵向设计的TOMMORROW研究；该研究对每位受试者的结果进行稀疏记录，部分受试者可能只有少数几次记录。

本文提出了一种针对稀疏观测多元函数型数据的双样本检验程序。我们的方法将多维函数联合投影到一组数据驱动的多元正交基上，并利用这些投影检验组间差异。通过沿主要变异方向采用多元 Karhunen-Lo´eve 展开来表示数据，可以简约地获得这些投影。我们将这些投影作为整个多维数据的替代，比较投影以检验整体上是否存在显著组间差异，并仅依据一个p值作出判断。因此，只要两组在任一分量函数（例如TOMMORROW研究中的任一项认知得分，共5项）上存在差异，我们基于投影的方法就适合检测显著的组间差异。相比之下，分别进行逐分量组间差异检验，需要处理多个p值，以调整总体第一类错误率（Pomann等人，2016；Wang，2021）。我们还给出了该检验的渐近零分布，并提供了一种基于置换的机制，用于在有限样本情形下近似零分布。数值研究表明，所提出的检验能够控制第一类错误；与备择假设相比，当数据仅轻微偏离原假设时，该检验仍具有较高的检验功效。该程序的计算速度随样本量变化较快，并且能够扩展到多元结果的维数。

本文其余部分安排如下。第 2 节介绍多元函数型数据的框架，并以数学形式表述假设。第 3 节构建基于投影的检验框架，第 4 节给出检验统计量及其渐近零分布。第 5 节的数值研究展示了该检验在检测显著组间差异方面的经验检验水平和检验功效性质。第 6 节讨论将本检验应用于 TOMMORROW研究所得的发现。

2. Statistical Framework

对于每位参与者 $i = 1, \ldots, n$，令 $Y(\ell)_{ij}$ 表示自基线 $\ell = 1$ 起于时间 t_{ij} 测量的第 ℓ 个响应，对应于第 j 次临床访视 $j = 1, \ldots, m_i$。因此，共有 n 名受试者，并在每个 m_i 个观测点上为受试者 i 记录一个 q 维响应向量。在我们的数据应用中，我们将对应于五个认知领域（即情景记忆、知觉、语义、工作记忆和视空）的得分分别建模为 $Y(\ell)_{ij}$，$\ell = 1, \ldots, 5$。用粗体符号 $\mathbf{Y}_{ij} := (Y(1)_{ij}, \ldots, Y(q)_{ij})^\top$ 表示在时间 t_{ij} 为第 i 位受试者联合记录的一组 q 维响应。在全文中，我们假设 q 为固定值，不随 n 增长。数据结构是不平衡的，即为第 i 位受试者记录测量值的观测点 $\{t_{i1}, \ldots, t_{im_i}\}$ 以及重复测量次数 m_i因受试者而异。对于 $\ell = 1, \ldots, q$，我们将 $\{Y(\ell)_{ij} : j = 1, \ldots, m_i\}$ 建模为平滑潜在过程 $X_i^{(\ell)}(\cdot)$ 的含噪实现；该过程定义在紧致域 $\mathcal{T} \subset \mathbb{R}$ 上，并在 $\{t_{i1}, \ldots, t_{im_i}\}$ 处求值，即

$$Y_{ij}^{(\ell)} = X_i^{(\ell)}(t_{ij}) + \epsilon_{ij}^{(\ell)}, \tag{2.1}$$

对于某些测量误差 $\epsilon_{ij}^{(\ell)}$，它们在 i、j 和 ℓ 之间独立分布。我们假设 $\epsilon_{ij}^{(\ell)} \sim (0, \tau_\ell^2)$、$\ell = 1$、... ...、$q$。对于每位参与者 i，我们假设潜在过程 $X_i(\cdot)$ 是稀疏观测的，即 m_i 很小。用粗体符号 $\mathbf{X}_i(\cdot) := (X_i^{(1)}(\cdot) \ldots, X_i^{(q)}(\cdot))$ 表示 q 维潜在过程构成的向量，并假设对每个 i，$\mathbf{X}_i(\cdot)$ 属于 $\mathcal{L}^2[\mathcal{T}]$ —— $\mathcal{T}$ 上所有平方可积的 q 维随机函数所构成的空间，其内积为 $\langle \mathbf{f}_1, \mathbf{f}_2 \rangle = \sum_{\ell=1}^q \langle f(\ell)_1, f_2^{(\ell)} \rangle = \sum_{\ell=1}^q \int_{\mathcal{T}} f(\ell)_1(t) f(\ell)_2(t) dt$，其中 $\mathbf{f}_1$ 和 $\mathbf{f}_2$ 属于 $\mathcal{L}^2(\mathcal{T})$。

令 g_i 为第 i 个受试者的组别指标，即若受试者 i 属于治疗（吡格列酮）组，则 $g_i = 1$；若属于安慰剂组，则为 $g_i = 0$。将 $\mathbf{X}_i(\cdot)$ 的均值记为

$$\mathbb{E}[\mathbf{X}_i(t)] = \boldsymbol{\mu}(t) + g_i \boldsymbol{\eta}(t), \tag{2.2}$$

其中，$\mathbb{E}[X_i^{(\ell)}(t)] = \mu^{(\ell)}(t) + g_i \eta^{(\ell)}(t)$，$\ell = 1, \ldots, q$ 为结局特异性均值。因此，$\mu^{(\ell)}(t)$ 表示在时间 t 第 ℓ 个响应的一般平均响应，而 $\eta^{(\ell)}(t)$ 表示在时间 t 治疗组与安慰剂组之间第 ℓ 个响应的均值差。此外，将该过程的协方差算子定义为 $\boldsymbol{\Xi}$，对于任意的 $f \in \mathcal{L}^2(\mathcal{T})$，算子 $(\boldsymbol{\Xi} f)(t)$ 的元素为 $(\boldsymbol{\Xi} f)^{(\ell)}(t) = \sum_{\ell'=1}^{q} \langle \Sigma_{\ell\ell'}(t, \cdot), f^{(\ell')} \rangle$，其中 $\Sigma_{\ell\ell'}(t, t') = \mathrm{Cov}\{X_i^{(\ell)}(t), X^{(\ell')}{}_i(t')\}$，$1 \leqslant \ell, \ell' \leqslant q$ 为 $\mathbf{X}_i(t)$ 中第 ℓ 个与第 ℓ' 个分量之间的协方差核。对于固定的 t、$t' \in \mathcal{T}$，$\mathbf{X}_i(t)$ 与 $\mathbf{X}_i(t')$ 之间的协方差矩阵可表示为 $\mathrm{Cov}(\mathbf{X}_i(t), \mathbf{X}_i(t')) = \boldsymbol{\Sigma}(t, t') = \{\Sigma_{\ell\ell'}(t, t')\}_{1 \leqslant \ell, \ell' \leqslant q}$。根据上述记号，检验处理效应显著性的零假设在数学上可转化为：

$$H_0 : \boldsymbol{\eta}(t) = (\eta^{(1)}(t), \ldots, \eta^{(q)}(t))^\top = 0 \quad \forall\ t \in \mathcal{T},$$

versus

$$H_1 : \eta^{(\ell)}(t) \neq 0 \quad \text{for some } t \in \mathcal{T}, \text{ and for some } \ell = 1, \ldots, q.$$

检验涉及 q 维函数效应的原假设，一种简单的方法是逐分量检验，即检验对于所有 $\ell = 1$、……、q，每个 $\eta^{(\ell)}(t) = 0\ \forall\, t$ 是否成立。文献中已提出多种检验单变量函数效应显著性的方法，例如 Zhang 和 Chen（2007）提出的基于散度的统计量，该方法适用于密集观测的函数；Crainiceanu 和 Ruppert（2004）提出了基于似然比的检验，以适应稀疏函数型数据。然而，这些分别进行的逐分量检验并未考虑多变量分量之间的联合动态变异。此外，目前没有通用的方法汇总各项检验得出的个体推断，进而对全局零假设作出结论，尤其是在纵向结局数量较多时。我们在第 3 节介绍的基于投影的多变量联合检验程序，旨在解决这些问题。

3. MULTIVARIATE PROJECTION-BASED INFERENCE APPROACH

令 $\{\boldsymbol{\phi}_1(\cdot) \ldots, \boldsymbol{\phi}_k(\cdot) \ \ldots\}$ 为 $\mathcal{L}^2[\mathcal{T}]$ 的一组正交归一基函数，其中对于 $k, k' \geqslant 1$，有 $\langle \boldsymbol{\phi}_k, \boldsymbol{\phi}_{k'} \rangle = \mathbb{I}(k = k')$。这里，$\mathbb{I}(\cdot)$ 表示示性函数。处理效应 $\boldsymbol{\eta}(t)$ 可唯一表示为：对于所有 $t \in \mathcal{T}$，$\boldsymbol{\eta}(t) = \sum_{k=1}^{\infty} c_k \boldsymbol{\phi}_k(t)$，其中，对于 $k \geqslant 1$，$c_k = \langle \boldsymbol{\eta}, \boldsymbol{\phi}_k \rangle = \sum_{\ell=1}^{q} \langle \eta^{(\ell)}, \phi_k^{(\ell)} \rangle$ 是 $\boldsymbol{\eta}(t)$ 在 $\boldsymbol{\phi}_k(t)$ 上的投影。这意味着，$\boldsymbol{\eta}$ 的函数范数的平方为 $\|\boldsymbol{\eta}\|^2 = \sum_{k=1}^{\infty} c_k^2$。注意，在原假设下，对于所有 $k \geqslant 1$，$c_k = 0$ 成立。另一方面，在 H_1 下，至少存在一个 $k \geqslant 1$ 使得 $c_k \neq 0$ 成立。在这种正交基表示下，零处理效应的原假设等价于检验是否

$$H_{0,k} : c_k = 0 \quad \text{vs} \quad H_{1,k} : c_k \neq 0, \tag{3.3}$$

对于所有 $k \geqslant 1$。这促使我们使用投影 $\{c_k\}_{k \geqslant 1}$ 来检验多变量处理效应是否显著。

考虑将中心化过程投影到特征函数 $\boldsymbol{\phi}_k(t)$ 上得到的标量投影 $\zeta_{i,k} := \langle \mathbf{X}_i - \boldsymbol{\mu}, \boldsymbol{\phi}_k \rangle$。根据 (2.2)，$\mathbb{E}(\zeta_{i,k}) = g_i \langle \boldsymbol{\eta}, \boldsymbol{\phi}_k \rangle = g_i c_k$，且 $\mathrm{Var}(\zeta_{i,k}) = \langle \boldsymbol{\Xi} \boldsymbol{\phi}_k, \boldsymbol{\phi}_k \rangle$ 对所有 $i = 1, \ldots, n$ 均成立。在原假设 H_0 下，$\mathbb{E}(\zeta_{i,k}) = 0$ 对所有 i 均成立；而在 H_1 下，至少对于

某些 $k \geqslant 1$， $\mathbb{E}(\zeta_{i,k}) = c_k \neq 0$ 对所有满足 $g_i = 1$ 的 i 均成立，即对于治疗组。因此，随机变量 $\zeta_{i,\ k}$ 可用于检验是否 $c_k = 0$。一种自然的识别方法是测量治疗组与对照组之间 $\zeta_{i,\ k}$ 平均值的差异，即 $D_k := n_1^{-1}\sum_{i:g_i=1}\zeta_{i,k} - n_0^{-1}\sum_{i:g_i=0}\zeta_{i,k}$，其中 $n_1 = \sum_{i=1}^{n}\mathbb{I}(g_i = 1)$，且 $n_0 = \sum_{i=1}^{n}\mathbb{I}(g_i = 0)$ 分别为治疗组和安慰剂组中的受试者人数。注意 $n_1 + n_0 = n$。在受试者之间响应相互独立的情况下，各投影在 i 上也相互独立分布，因此 $\mathrm{Var}(D_k) = (1/n_1 + 1/n_0)\langle \boldsymbol{\Xi}\boldsymbol{\phi}_k, \boldsymbol{\phi}_k\rangle$。因此，对于每个 $k \geqslant 1$， D_k 可用于检验 $H_{0,k}$ 与 $H_{1,k}$。

为了检验原假设 H_0，我们需要对 $H_{0,\ k}$ 与 $H_{1,\ k}$ 进行检验，其中对所有 $k \geqslant 1$均成立。这意味着，为了对H_0作出全局结论，我们必须检验可数无穷多个假设，这是不可行的。在实际应用中，我们必须对$H_{0,\ k}$、 $k = 1$，……， K进行同时检验，其中 K取较大的值。因此，实施基于投影的推断程序不仅需要选择正交基，还需要选择正交基的最优数量K。即使选定了 K，我们仍需结合从每个较简单假设中得到的推断，才能对 H_0作出全局推断。

从理论上讲，任何预先设定的多元正交基函数都可以使用，例如傅里叶基、小波或勒让德基的多元版本。然而，在给定基下选择截断参数 K 会变得困难，原因在于：i）对于一组数据，没有客观的方法来选择最优的基函数数量，而且通常需要检验大量形如 $H_{0,\ k}$的较简单假设；ii）由于投影 $\{\zeta_{i,k}\}_{k=1}^{K}$ 彼此相关，当 K较大时，难以直接处理差向量 $\{D_k\}_{k=1}^{K}$'s 之间的相关性，进而利用 $(D_1, \ldots\ldots, D_K)^\top$构造检验。为避免这一问题，我们从协方差算子 $\boldsymbol{\Xi}$中选择一组数据驱动的多元特征基，具体如下。

3.1 数据驱动投影

假设$\mathbf{X}(t)$的对称非负协方差核是连续的，因此线性协方差算子 $\boldsymbol{\Xi}$ 在 $\mathcal{L}^2(\mathcal{T})$中是紧的、自伴的（Hutson等人，2005年，第7章）。根据希尔伯特–施密特定理，存在一组 q维正交基函数（也称为特征函数） $\{\boldsymbol{\psi}_k(\cdot)\colon = (\psi_k^{(1)}(\cdot)\ldots, \psi_k^{(q)}(\cdot))^\top\}_{k\geqslant 1} \in \mathcal{L}^2(\mathcal{T})$，满足 $\langle \boldsymbol{\psi}_k, \boldsymbol{\psi}_{k'}\rangle = \mathbb{I}(k = k')$，其中 k， $k' \geqslant 1$ 使得

$$(\boldsymbol{\Xi}\boldsymbol{\psi}_k)^{(\ell)}(t) = \sum_{\ell'=1}^{q}\langle \Sigma_{\ell\ell'}(t,\cdot)\,, \psi_k^{(\ell')}\rangle = \lambda_k \boldsymbol{\psi}_k^{(\ell)}(t) \qquad \ell = 1, \ldots, q, \quad k = 1, 2, \ldots,$$

其中， $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant 0$ 是 $\boldsymbol{\Xi}$ 的有序特征值，其迹为 $\mathrm{tr}(\boldsymbol{\Xi}) = \sum_{k=1}^{\infty}\lambda_k < \infty$。应用多变量版 Mercer 定理，可得协方差核的谱分解为 $\boldsymbol{\Sigma}(t, t') = \sum_{k=1}^{\infty}\lambda_k \boldsymbol{\psi}_k(t)\boldsymbol{\psi}_k(t')$，其中该收敛关于 $t \in \mathcal{T}$ 是一致的。协方差算子 $\boldsymbol{\Xi}$ 的特征函数 $\{\boldsymbol{\psi}_k(t)\}_{k\geqslant 1}$ 将作为我们选择的正交基 $\{\phi_k(t)\}$（在第3节中介绍），用于实施数据驱动的投影检验。利用这些特征基，潜在过程具有多变量 Kahrunen-Loeve（KL）表示:

$$\mathbf{X}_i(t) = \boldsymbol{\mu}(t) + \sum_{k=1}^{\infty}\xi_{i,k}\boldsymbol{\psi}_k(t), \tag{3.4}$$

其中 $\xi_{i,\ k} = \langle \mathbf{X}_i - \boldsymbol{\mu},\ \boldsymbol{\psi}_k\rangle$ 是将 $\mathbf{X}_i(t)$ 投影到 q元特征函数 $\boldsymbol{\psi}_k(t)$ 上所得的投影，称为多元函数主成分（mv-fPC）得分。我们进一步利用这些得分$\{\xi_{i,k}\}_{i=1}^{n}$ 构造一个关于 $H_{0,\ k}$ 对所有 $k \geqslant 1$均成立的检验。使用 $\{\boldsymbol{\psi}_k(t)\}_{k\geqslant 1}$ 作为

我们选择正交基有多方面的优势。首先，卡尔南–洛维展开为潜在过程提供了简约表示，因此 $\{\psi_k(t)\}_{k\geqslant 1}$ 是表示数据的自然选择。其次，通过确保数据中的大部分变异得到解释，可以客观地选择特征函数的最优数量 K。换言之，若将上述表示（3.4）截断至前 K 个特征函数，则 $X_i^{(K)}(t) = \mu(t) + \sum_{k=1}^{K} \xi_{i,k}\psi_k(t)$ 是无限维随机过程 $X_i(t)$ 的最佳 K 秩近似。第三，多元函数主成分得分 $\{\xi_{i,k}\}_{k\geqslant 1}$ 不同于 $\{\zeta_{i,k}\}_{k\geqslant 1}$，后者是将 $X_i(t) - \mu(t)$ 投影到任意正交基 $\{\phi_k(t)\}_{k\geqslant 1}$ 上所得的投影，因为 K 元组 $(\xi_{i,1}, \ldots, \xi_{i,K})^\top$ 可作为识别整个函数 $X_i(t)$ 变异的关键替代指标，且各得分之间互不相关。此外，我们得到 $\xi_{i,k}$ 的方差的简化表达式，即 $\mathrm{Var}(\xi_{i,k}) = \langle \Xi\psi_k, \psi_k\rangle = \lambda_k$。在（2.2）条件下，多元函数主成分得分均值之差，

$$\mathcal{D}_k := n_1^{-1} \sum_{i:g_i=1} \xi_{i,k} - n_0^{-1} \sum_{i:g_i=0} \xi_{i,k} \qquad k = 1, 2, \ldots, K$$

其均值为 $c_k^* = \langle \eta, \psi_k\rangle$，处理效应 $\eta(t)$ 在特征函数 $\psi_k(t)$ 上的投影为 $\lambda_k(1/n_1 + 1/n_0)$，方差为 $\mathcal{D}_k(\{\xi_{i,k}\}_{k=1}^K)$。此外，由于投影 k在 K 上彼此不相关， K之间也互不相关。在给定特征函数的有限截断的条件下（ K（的估计将在第 4 节讨论）），维向量 $\boldsymbol{\mathcal{D}} := (\mathcal{D}_1, \ldots, \mathcal{D}_K)^\top$的均值为 $(c_1^*, \ldots, c_K^*)^\top$，协方差为 $(1/n_1 + 1/n_0)\mathrm{diag}(\lambda_1, \ldots, \lambda_K)$。注意，在 H_0下，$(c_1^*, \ldots, c_K^*) \equiv 0$。因此，我们通过多元函数主成分分析（fPCA）获得的两组之间 K维数据驱动得分的差异，建立了一个用于检测两组多元函数型数据均值差异的框架。下一节将正式介绍基于得分的检验。

4. PROJECTION-BASED MULTIVARIATE FPCA SCORE BASED TEST

真实的特征函数及其对应的多元函数主成分得分是未知的，需要根据数据进行估计。这里简要介绍特征分量的估计策略。令 $\widehat{\mu}(t)$ 和 $\Sigma(t,t')$ 分别为多元均值和协方差函数的一致估计量。关于用于估计多元函数型数据协方差函数的不同平滑方法，参见 Chiou 等人（2014年）；Happ 和 Greven（2018年）；Li 等人（2020年）。$\widehat{\phi}_k(t)$ 的一致估计量可通过对估计得到的协方差进行谱分解获得，即 $\widehat{\Sigma}(t,t') = \sum_k \lambda_k\psi_k(t)\psi_k(t')$。最优特征函数的数量 K 可通过设定预先指定的变异解释百分比（PVE），或采用 Akaike 信息准则（AIC）来确定。对于密集观测函数型数据，可以通过计算 $\xi_{ik} = \langle X_i - \widehat{\mu}, \phi_k\rangle$ 获得多元函数主成分得分的一致估计量$\widehat{\ }$。然而，对于稀疏数据，我们无法观测到整个函数 $X_i(t)$$\widehat{\ }$。在这种情况下，我们建议将 $\xi_{ik} = \langle X_i - \widehat{\mu}, \phi_k\rangle, k = 1, \ldots, K,$$\widehat{\ }$ 作为得分的估计量，其中 $X_i(t)$ 是在工作高斯假设下对未观测轨迹 $X_i(t)$ 的最佳线性无偏预测（Li 等人，2020年）。关于协方差函数估计方法的详细描述，以及 ξ_{ik} 的显式公式，见补充材料的 S7.1 节$\widehat{\ }$。

Let $\widehat{\boldsymbol{\xi}}_i = (\widehat{\xi}_{i1}, \ldots, \widehat{\xi}_{iK})^\top$ 为第 i位受试者估计的 K维多元函数主成分得分。定义，$\widehat{\boldsymbol{\xi}}_{1+} = n_1^{-1}\sum_{i:\ g_i=1}\widehat{\boldsymbol{\xi}}_i$，且 $\widehat{\boldsymbol{\xi}}_{0+} = n_0^{-1}\sum_{i:\ g_i=0}\widehat{\boldsymbol{\xi}}_i$ 分别为治疗组和安慰剂组估计得分的平均值。类似地，定义，$\widehat{\boldsymbol{\Lambda}}_1 = (n_1-1)^{-1}\sum_{i:\ g_i=1}(\widehat{\boldsymbol{\xi}}_i - \widehat{\boldsymbol{\xi}}_{1+})(\widehat{\boldsymbol{\xi}}_i - \widehat{\boldsymbol{\xi}}_{1+})^\top$，以及 $\widehat{\boldsymbol{\Lambda}}_0 = (n_0-1)^{-1}\sum_{i:\ g_i=0}(\widehat{\boldsymbol{\xi}}_i - \widehat{\boldsymbol{\xi}}_{0+})(\widehat{\boldsymbol{\xi}}_i - \widehat{\boldsymbol{\xi}}_{0+})^\top$，它们分别是两组最佳线性无偏预测估计得分的样本方差。此外，将合并样本协方差定义为

$\widehat{\boldsymbol{\Lambda}} = \{(n_1-1)\boldsymbol{\Lambda}_1 + (n_0-1)\boldsymbol{\Lambda}_0\}/(n-2)$。为检验 H_0，可利用多元函数主成分得分构造 Hotelling T平方统计量（Muirhead，2009，第 6 章，公式 10），如下：

$$T_n = \frac{n_1 n_0}{n_1+n_0} \widehat{\boldsymbol{\mathcal{D}}}^{\top} \widehat{\boldsymbol{\Lambda}}^{-1} \widehat{\boldsymbol{\mathcal{D}}}, \tag{4.5}$$

其中 $\widehat{\boldsymbol{\mathcal{D}}} = \boldsymbol{\xi}_{1+} - \boldsymbol{\xi}_{0+}$。在原假设下，当样本量较大时，$T_n$ 近似服从自由度为 K 的卡方分布。然而，对于固定样本量 n，为考虑用样本版本估计函数主成分得分的真实协方差所带来的变异，在对多元函数主成分得分作出工作高斯假设的前提下，我们将 Hotelling T 平方分布与 F分布联系起来（定理5.9，H¨ardle 和 Simar，2019）。若满足以下条件，我们的检验规则将在指定显著性水平 $\alpha \in$ （0，1）下拒绝 H_0：

$$T_n > \frac{(n-2)K}{(n-K-1)} F_{\alpha}(K, n-K-1), \tag{4.6}$$

其中，K是根据数据估计得到的多元函数主成分得分的维数，而 $F_{\alpha}(a,\ b)$是自由度为 a 和 b 的 F分布的第100(1 –α)%分位数。下一个命题从理论上确认该检验能够控制第一类错误。

命题4.1。 假设观测响应 $\{\mathbf{Y}^{ij}:\ j = 1, \ldots, m_i\}_{i=1}^{n}$符合模型 *(2.1)*，且 $\sup_i m_i < \infty$。进一步假设均值函数和特征分量$\widehat{\ }$的估计是一致的，即 $\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\| = o_p(1)$，$\|\boldsymbol{\psi}_k - \boldsymbol{\psi}_k\| = o_p(1)$，$\|\lambda_k - \lambda_k\| = o_p(1)$ 对所有$k = 1, \ldots, K$均成立，且 $\|\widehat{\tau}_{\ell}^2 - \tau_{\ell}^2\| = o_p(1)$ 对所有 $\ell = 1, \ldots, q$均成立，并且 $\lim_{n\to\infty} n_1/n \to w \in (0,1)$。那么，给定截断参数 K，在零假设下，对任意 $\alpha \in (0,1)$

$$\mathbb{P}\left(T_n > \frac{(n-2)K}{(n-K-1)} F_{\alpha}(K, n-K-1) \,\middle|\, H_0 \text{ is true} \right) \leqslant \alpha \qquad \text{as } n \to \infty.$$

上述命题所作的假设在函数型数据分析文献中相当标准。Chiou 等人（2014年）建立了多元函数主成分分析中特征分量的一致性。该命题的证明与 Wang（2021）针对单变量情形的定理 1 的证明类似，并利用了多元响应 q 的维数是有限的这一事实；为避免冗余，此处省略证明。

该命题是在给定估计特征函数数量 K的条件下提出的。需要注意的是，估计特征函数的数量 K 会随着 n增加。然而，K 随 n 增长的速率要慢得多。固定 K 不会影响检验的大小，但可能会影响检验功效。具体而言，考虑这样一种假设情形：协方差的真实特征基被截断至 $\{\boldsymbol{\psi}_k(t)\}_{k=1}^{K}$，并且 $\boldsymbol{\eta}(t)$ 沿这 K 个主导方向的投影为零假设，但沿第（$K+1$）个方向显著异于零，即 $\boldsymbol{\psi}_{K+1}(t)$。沿前 K 个特征函数检验原假设将不具有检验功效。然而，设定较大的解释方差比例（例如 99% 或 99.9%）将导致选择更多的特征函数，从而降低陷入这种不利情形的可能性。

基于置换的零分布近似：对于较小的样本量（例如 $n < 50$），也可以通过置换检验来近似 Hotelling T 平方统计量的零分布，而不使用 F 分布。置换检验的步骤如下。

$\widehat{\ }$1. 将所有受试者的多元函数主成分得分样本 $\{\boldsymbol{\xi}_i\}_{i=1}^{n}$ 随机分为两组，大小分别为 n_1 和 n_0。将 $g_i = 1$ 分配给第一组，将 $g_i = 0$ 分配给第二组。

2. 使用新的置换样本，根据公式（4.5）计算检验统计量的值。

3. 重复步骤 1-2 B 次，从检验统计量的零分布中获取样本 $\{T_{n,b}\}_{b=1}^{B}$ ，其中 B 取较大的值，例如 5000。

4. p值计算为 $B^{-1}\sum_{b=1}^{B}\mathbb{I}(T_{n,\,b} > T_n)$，据此对 H_0 得出结论。

方差不等时的检验统计量：式 (4.5) 中构造的 Hotelling T 平方统计量基于以下假设：两组的多元函数主成分得分的方差相同。这一假设主要源于我们所作的假设，即无论受试者 i 属于哪一组，$\mathbf{X}_i(t)$ 的协方差对所有$i = 1, \ldots\ldots, n$ 都相同。然而，这一假设并非总是成立。在这种情况下，我们可以在假设方差不等的前提下，采用 Satterthwaite 近似（Nel 和 Van der Merwe，1986）对统计量进行近似，从而进行检验：

$$T_{n,UV} = \widehat{\boldsymbol{\mathcal{D}}}^{\top}(n_1^{-1}\widehat{\boldsymbol{\Lambda}_1} + n_0^{-1}\widehat{\boldsymbol{\Lambda}_0})^{-1}\widehat{\boldsymbol{\mathcal{D}}}.$$

当显著性水平为 $\alpha \in (0, 1)$ 时，我们拒绝 H_0。

$$T_{n,UV} > \frac{fK}{(f-K-1)}F_{\alpha}(K, f-K-1), \tag{4.7}$$

其中 f 为有效自由度。f 的精确公式见 Nel 和 Van der Merwe（1986，第 12 页）。

结果特异性均值函数的推断：在对零假设作出整体结论之后，若拒绝零假设，自然会关注哪些 q 个结局在治疗组与安慰剂组之间存在显著差异。可通过分别考察估计的处理效应 $\widehat{\eta}^{(\ell)}(t)$、$\ell = 1, \ldots\ldots, q$ 来回答这一问题，并获得 $\widehat{\eta}^{(\ell)}(t)$ 的自助法标准误。有关函数型数据中受试者的自助抽样机制，参见 Park等人（2018）。获得自助法标准误后，可通过 Cui等人（2022）的算法 2，构造真实处理效应 $\eta^{(\ell)}(t)$ 的 100$(1-\alpha)$% 同时置信带。为调整对处理效应的 q 个分量分别进行多重检验时的第一类错误，我们设定 $\alpha^* = \alpha/q$ 来构造个别置信带。

扩展到多于两组：当组数 G 多于两组时，检验 G 组之间多变量均值函数的相等性，可转化为检验 G 组受试者的 K 维多元函数主成分得分 $\{\widehat{\boldsymbol{\xi}}_i : g_i = r\}_{r=1}^{G}$ 是否具有相同的均值。这类似于一个涉及多元函数主成分得分的多元方差分析（MANOVA）问题。定义 $n_r = \sum_{i=1}^{n} I(g_i = r)$ 为组频数，$\widehat{\boldsymbol{\xi}}_{r+} = n_r^{-1}\sum_{i:\ g_i=r}\widehat{\boldsymbol{\xi}}_i$ 为组均值，并将 $\widehat{\boldsymbol{\Lambda}}_r = (n_r-1)^{-1}\sum_{i:\ g_i=r}(\widehat{\boldsymbol{\xi}}_i - \widehat{\boldsymbol{\xi}}_{r+})(\widehat{\boldsymbol{\xi}}_i - \widehat{\boldsymbol{\xi}}_{r+})^{\top}$ 分别定义为第 r 组的协方差，$r = 1, \ldots, G$。此外，记 $\widehat{\boldsymbol{\xi}}_{++} = n^{-1}\sum_{i=1}^{n}\widehat{\boldsymbol{\xi}}_i$ 为多元函数主成分得分的总体均值。为进行多元方差分析，我们构造组间误差矩阵$\mathbf{Q}$

$$\sum_{r=1}^{G} n_r\ \widehat{\boldsymbol{\xi}}_{r+} - \widehat{\boldsymbol{\xi}}_{++}\ \ \widehat{\boldsymbol{\xi}}_{r+} - \widehat{\boldsymbol{\xi}}_{++}\ ^{\top} \qquad {}_{E} = \sum_{r=1}^{G} n_r - 1\ \overset{H}{\widehat{\boldsymbol{\Lambda}}}_r\ (\,)(\,),$$

以及组内误差矩阵$\mathbf{Q}(\,)$。基于$\mathbf{Q}_H$ 和$\mathbf{Q}_E$ 可以构造多个检验统计量，例如 Wilks' Lambda、Lawley–Hotelling 迹和 Pillai 迹等（Johnson 和 Wichern，2002）。我们给出 Lawley–Hotelling 迹统计量的形式，其在 H_0 取较大值时拒绝

$$T_{LH} = \mathrm{tr}(\mathbf{Q}_H\mathbf{Q}_E^{-1}).$$

在原假设下，T_{LH} 近似服从 χ^2 分布，具有 $K(G-1)$ 个自由度。注意，K 是根据数据估计得到的多元函数主成分得分的维数。上述检验统计量中，没有任何一个被证明在检验功效方面优于其他统计量。对于 $G=2$，它们都化为式（4.5）中的 Hotelling T 平方统计量。

5. Numerical studies

本节通过数值模拟，考察所提出的检验在检测偏离原假设时的有限样本一类错误和检验功效。为此，我们通过以下机制生成三维（即 $q=3$）函数型数据，

$$Y_{ij}^{(\ell)} = \mu^{(\ell)}(t_{ij}) + g_i\eta^{(\ell)}(t_{ij}) + \sum_{k=1}^{3}\xi_{ik}\psi_k^{(\ell)}(t_{ij}) + \epsilon_{ij}^{(\ell)},$$

对于 $\ell=1$、2、3、$j=1$、...、m_i，以及 $i=1$、...、n。对于所有 i、j，观测点 t_{ij} 从 $[0,1]$ 中均匀抽取。均值向量设为 $\mu^{(1)}(t) = 5\sin(2\pi t)$、$\mu^{(2)}(t) = 5\cos(2\pi t)$ 和 $\mu^{(3)}(t) = 5(t-1)^2$,；处理效应设为 $\eta^{(\ell)}(t) = 5\delta(t/4-0.5)^3$，其中 $\ell=1$ 为 2、3。参数 δ 控制偏离原假设的程度；当 $\delta=0$ 时，H_0 成立。对于每个 $i=1$，...，n，组别指标 g_i 从 $\{0,1\}$ 中随机抽取。对于随机成分，函数主成分得分 $\xi_{i,k}$ 的均值为零，且当 $k=1$ 为 2、3 时，方差为 λ_k，其中 $\lambda_1=6$、$\lambda_2=3$ 和 $\lambda_3=1.5$。标准正交特征函数设为 $\boldsymbol{\psi}_1(t) = \sqrt{2/3}[\sin(2\pi t)、\cos(4\pi t)、\sin(4\pi t)]^\top$；$\boldsymbol{\psi}_2(t) = \sqrt{2/3}[\sin(\pi t/2)、\sin(3\pi t/2)、\sin(5\pi t/2)]^\top$；以及 $\boldsymbol{\psi}_3(t) = \sqrt{2/3}[\sin(\pi t)、\sin(2\pi t)、\sin(3\pi t)]^\top$。对于所有 i、j 和 ℓ，测量误差 $\epsilon_{ij}^{(\ell)} \sim$ 服从 $(0,\sigma_e^2)$，其中 $\sigma_e=0.2$。

我们考虑一个包含三个重要因素的析因设计：样本量 n、观测函数的稀疏程度，以及函数主成分得分的分布。为展示检验在样本量从小到大时的表现，我们考虑五种不同的样本量 n：$n=50$、70、100、200 和 300。我们选择三个不同的稀疏程度等级，根据函数被观测的次数确定，即高：$m_i \in \{4,5,6,7\}$，中：$m_i \in \{8,9,10,11,12\}$，以及低：$m_i \in \{15,16,17,18,19,20\}$，适用于所有 i。对于给定的 m_i，观测点 t_{ij} 从 $[0,1]$ 中由 51 个等距点组成的网格内随机无放回地选取。对于最后一个因素，得分 $\xi_{i,k}$ 来自高斯 分布或非高斯 分布；若来自非高斯分布，则由两个正态分布的混合生成：以 0.5 的概率取 $N(\sqrt{\lambda_k}/2,\ \lambda_k/2)$，以 0.5 的概率取 $N(-\sqrt{\lambda_k}/2,\ \lambda_k/2)$，$k=1$、2、3。

5.1 计算细节

对于每个成分 $\ell=1$、2、3，我们在工作独立假设下，使用 mgcv 软件包中的 gam()函数（Wood, 2011）估计共同均值函数 $\mu^{(\ell)}(t)$ 和处理效应 $\eta^{(\ell)}(t)$，从而得到所有i、j和 ℓ对应的残差 $\widehat{E}_{ij}^{(\ell)} = Y_{ij}^{(\ell)} - \widehat{\mu}^{(\ell)}(t_{ij}^{(\ell)}) - g_i\widehat{\eta}^{(\ell)}(t_{ij}^{(\ell)})$。节点数设为 10，并均匀放置。在第二步中，我们使用 R软件包mfaces（Li 和 Xiao，2021）中的 mface.sparse()函数，对残差 $\{E_{ij}^{(\ell)}\}_{i、j、\ell}$应用多元函数主成分分析（mv-fPCA），从而得到估计的特征值 λ_k和特征函数 $\boldsymbol{\psi}_k(t)$。特征函数的数量 K 根据预先设定的 99% 解释方差比例选取。在第三步中，对于所有 $i=1$、...、n，我们使用第 4 节所述的方法估计多元函数主成分得分 $\widehat{\boldsymbol{\xi}}_i = (\widehat{\xi}_{i,1},\ldots,\widehat{\xi}_{i,K})^\top$。最后，对两组的估计得分 $\{\boldsymbol{\xi}_i:\ g_i=1\}_{i=1}^n$ 和 $\{\boldsymbol{\xi}_i:\ g_i=0\}_{i=1}^n$ 进行检验

表 1. 单次重复的计算时间（秒） 次

	high	中	*low*
$n = 100$	14.26	15.30	18.64
$n = 200$	27.08	28.33	38.72

使用 R 软件包 Hotelling（Curran 和 Hersh，2021）中的 hotelling.test()函数检验其是否相等，并依据单个 p 值对原假设 H_0 作出结论。

我们检验程序的计算成本主要取决于执行多元函数主成分分析的复杂度，该过程需要 $\mathcal{O}(n \max\{q^3m, q^2m^2\})$ 次运算。这里 $m = \max_i m_i$。因此，当每位受试者的观测次数 m_i 有限时，检验的复杂度会随多元响应的维数呈立方增长。表 1 记录了不同样本量和稀疏度水平下单次重复的中位计算时间（秒）。

5.2 检验性能评估

5.2.1 检验水平性质 表2展示了所提出的检验在名义水平 $\alpha = 0.01$、0.05、0.1、0.15 下的经验第一类错误率，涵盖不同样本量和稀疏度水平，并分别考察得分服从高斯 分布和非高斯 分布的情形。数值由检验规则（4.6）计算得出。估计值及其标准误（括号内）基于 10、000 次模拟得到。除了少数样本量较小（$n = 50$，70）且稀疏度水平较高的情况外，经验检验水平都在名义水平的两个标准误范围内 α，这表明我们的检验能够很好地控制第一类错误。即使在 $n = 50$ 情况下，我们也看到，随着稀疏程度从高 降至低，第一类错误的膨胀得到了控制。这表明，只要观测响应的稀疏程度不过高，即使每组的样本量较小，也可以进行所提出的检验。检验水平略有膨胀，可能是由于样本量较小和/或观测函数的稀疏程度较高，导致特征函数估计不佳。有意思的是，即使得分服从非高斯 分布，检验水平仍能得到控制，这反映了我们的检验对于偏离正态性的情况具有稳健性。

5.2.2 检验功效 将检验的显著性水平固定为 $\alpha = 0.1$。图1(a)展示了所提出的检验在不同样本量和观测函数稀疏程度下，得分为高斯时关于 δ 的功效曲线；图1(b)则展示了得分为非高斯 时的相应结果。正如预期，检验功效随样本量增加而提高。此外，在样本量固定时，稀疏程度越低，检验功效提升越快。比较图1左图和右图中的功效曲线可以看出，所提出的检验具有稳健性，因为当数据分布偏离正态性时，其检验功效仍然相近。

我们进一步将本检验的性能与 Pomann 等人（2016）提出的推断程序进行比较。该程序针对每个 $k = 1, \ldots, K$ 使用函数主成分得分进行多个双样本单变量检验，并结合 Bonferroni校正进行多重比较校正。尽管该程序主要针对单变量函数数据开发，我们仍将其应用于多变量函数数据情形。如图2所示，在所有稀疏度水平下，无论是高斯情形（图2(a)）还是非高斯情形（图2(b)），我们的检验功效均优于其主要竞争方法；当偏离原假设的程度较小且样本量不大时，优势尤为明显。这表明，当

表2. 基于10,000次模拟的所提出的检验的经验第一类错误率。括号内为标准误。

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.15$
得分分布：高斯				
稀疏性：高 $(m_i \sim \{4, \ldots, 7\})$				
$n = 50$	0.010 (0.001)	0.055 (0.002)	0.108 (0.003)	0.163 (0.004)
$n = 70$	0.012 (0.001)	0.051 (0.002)	0.106 (0.003)	0.156 (0.004)
$n = 100$	0.010 (0.001)	0.051 (0.002)	0.104 (0.003)	0.160 (0.004)
$n = 200$	0.009 (0.001)	0.049 (0.002)	0.101 (0.003)	0.153 (0.004)
$n = 300$	0.011 (0.001)	0.049 (0.002)	0.100 (0.003)	0.149 (0.004)
稀疏性：中 $(m_i \sim \{8, \ldots, 12\})$				
$n = 50$	0.009 (0.001)	0.051 (0.002)	0.103 (0.003)	0.159 (0.004)
$n = 70$	0.010 (0.001)	0.053 (0.002)	0.101 (0.003)	0.152 (0.004)
$n = 100$	0.009 (0.001)	0.047 (0.002)	0.097 (0.003)	0.152 (0.004)
$n = 200$	0.009 (0.001)	0.048 (0.002)	0.097 (0.003)	0.143 (0.004)
$n = 300$	0.009 (0.001)	0.047 (0.002)	0.094 (0.003)	0.148 (0.004)
稀疏性：低 $(m_i \sim \{15, \ldots, 20\})$				
$n = 50$	0.009 (0.001)	0.051 (0.002)	0.100 (0.003)	0.151 (0.004)
$n = 70$	0.010 (0.001)	0.050（0.002）	0.098（0.003）	0.144（0.004）
$n = 100$	0.010（0.001）	0.047（0.002）	0.096（0.003）	0.148 (0.004)
$n = 200$	0.010 (0.001)	0.048 (0.002)	0.099 (0.003)	0.146 (0.004)
$n = 300$	0.010 (0.001)	0.050 (0.002)	0.099 (0.003)	0.152 (0.004)
得分分布：高斯混合				
稀疏性：高 $(m_i \sim \{4, \ldots, 7\})$				
$n = 50$	0.010 (0.001)	0.053 (0.002)	0.112（0.003）	0.164（0.004）
$n = 70$	0.012（0.001）	0.059（0.002）	0.112（0.003）	0.163（0.004）
$n = 100$	0.011 (0.001)	0.052 (0.002)	0.105 (0.003)	0.158 (0.004)
$n = 200$	0.008 (0.001)	0.046 (0.002)	0.095 (0.003)	0.148 (0.004)
$n = 300$	0.009 (0.001)	0.046 (0.002)	0.097 (0.003)	0.146 (0.004)
稀疏性：中 $(m_i \sim \{8, \ldots, 12\})$				
$n = 50$	0.011 (0.001)	0.056 (0.002)	0.110 (0.003)	0.161 (0.004)
$n = 70$	0.010 (0.001)	0.052 (0.002)	0.104 (0.003)	0.154 (0.004)
$n = 100$	0.011 (0.001)	0.052 (0.002)	0.101 (0.003)	0.154 (0.004)
$n = 200$	0.010（0.001）	0.050（0.002）	0.104（0.003）	0.155（0.004）
$n = 300$	0.009（0.001）	0.048（0.002）	0.100 (0.003)	0.153 (0.004)
稀疏性：低 $(m_i \sim \{15, \ldots, 20\})$				
$n = 50$	0.011 (0.001)	0.052 (0.002)	0.104 (0.003)	0.152 (0.004)
$n = 70$	0.011 (0.001)	0.050 (0.002)	0.100 (0.003)	0.154 (0.004)
$n = 100$	0.010 (0.001)	0.050 (0.002)	0.104 (0.003)	0.156 (0.004)
$n = 200$	0.011 (0.001)	0.049 (0.002)	0.099 (0.003)	0.150 (0.004)
$n = 300$	0.009 (0.001)	0.046 (0.002)	0.094 (0.003)	0.146 (0.004)

认知域	测试
情景记忆	CVLT短延迟, CVLT长延迟, BVMT延迟回忆
执行功能（工作记忆）	韦氏成人智力量表第三版 数字广度测验 – 倒背广度, 连线测验（B部分）
语言（语义记忆）	多语言命名测验（MiNT）, 语义流畅性（动物）, 词汇/音位流畅性
注意力（知觉速度）	连线测验（A部分）, 韦氏成人智力量表第三版 数字广度测验——顺背广度
视空间能力	画钟测验, 简明视空间记忆测验图形的临摹

表 3. 五个认知领域的测试组合：情景记忆、工作记忆、语义、知觉和视空。

由于检测偏离零假设的情况可能较为困难，因此我们的方法比其他方法更具检验功效。这主要是因为采用 Bonferroni校正调整第一类错误率的多重检验程序较为保守。

6. ANALYSIS OF TOMMORROW STUDY

6.1 研究详情

数据收集自一项三期、多中心、随机、安慰剂对照、双盲、平行组研究。研究对象为美国附属研究诊所和私人研究诊所中年龄在65至83岁之间、认知健康且因阿尔茨海默病而具有较高轻度认知障碍（MCI）发病风险的个体。参与者按1:1随机分配，分别口服外观相同的研究药物吡格列酮片或安慰剂片。参与者在基线时接受评估，此后每隔6个月到诊所进行一次访视；访视期间，参与者及其研究伙伴填写问卷，以评估参与者的临床状况和认知状态。研究伙伴知晓参与者的认知状态。研究总时长约为5年。最终数据包含1947名高风险个体，其中吡格列酮组有985名受试者，安慰剂组有962名受试者。有关研究设计的详细说明，参见Burns等人（2021）。

6.1.1 分析的结局与目标：测试组包含十二项表现测试，涉及五个认知领域：情景记忆、执行功能（工作记忆）、语言（语义记忆）、注意力（知觉速度）和视空间能力。测试在基线时及之后每隔六个月进行一次。与这五个认知领域对应的测试列于表 3。除连线测验（B 部分和 A 部分）外，得分越高表示认知能力越好。因此，我们将连线测验得分乘以 -1，以保留“得分越高越好”这一结局与认知能力之间的关系。十二项测试的原始得分减去其基线均值，再除以基线

标准差，以得到标准化得分。随后，在每个领域内对标准化测试得分取均值，得到对应于五个认知领域的五维综合得分，作为我们分析中的响应变量。每位受试者实际进行测试并获得得分的日期，并不恰好与上次记录测量值的日期相隔六个月。为更好地了解认知能力下降，我们将从基线到第 j 次访视期间、获取测试得分的确切日期之间的天数差记为 t_{ij} ，并据此将综合测试得分建模为随时间平滑变化的函数。这意味着，对于研究中的所有受试者， $t_{i1} = 0$ 。

6.2 统计分析与研究结果

在本节中，我们展示通过实施基于投影的检验所得的结果与发现，验于TOMMORROW研究。

6.2.1 基于投影的检验的实现我们使用对应于五个认知领域的五维标准化综合得分拟合模型(2.1)，以实施基于投影的检验。模型的均值部分通过 mgcv 软件包中的 gam()函数（Wood, 2011）进行拟合，以估计总体平滑均值 $\widehat{\mu}^{(\ell)}(t)$ 以及吡格列酮的平滑效应$\widehat{\eta}^{(\ell)}(t)$。我们未在模型中纳入任何基线协变量，因为补充材料 S7.2 节中的基线分析表明，两组在这些基线协变量方面具有同质性。图 3 展示了对应于五个认知领域的观测表现得分经局部加权回归平滑（LOESS）平滑后的散点图，并叠加了估计的平均响应，即安慰剂组的 $\widehat{\mu}^{(\ell)}(t)$，以及治疗组的 $\widehat{\mu}^{(\ell)}(t) + \widehat{\eta}^{(\ell)}(t)$、 $\ell = 1$、.....、 q。治疗组与安慰剂组之间观测到的 LOESS 平滑曲线差异幅度较小。

估计均值成分后，我们计算残差 $\widetilde{\boldsymbol{\varepsilon}}_{ij} = \widetilde{\boldsymbol{\mu}}(t_{ij}) - \boldsymbol{g}_i \boldsymbol{\eta}(t_{ij})$ 对每个 i， j。利用残差，我们使用 R 软件包 ^mfaces（Li 和 Xiao，2021）估计平滑协方差函数，并通过对估计的协方差进行谱分解，得到主要特征函数 $\{\boldsymbol{\psi}_k(t)\}_k$ 。将解释方差比例设为 99%，即可得到 $K = 8$ 个主方向。对于估计出的八个方向，我们得到每位参与者的 8 维最佳线性无偏预测多元函数主成分得分 $\widehat{\boldsymbol{\xi}}_i = (\widehat{\xi}_{i,1}, \ldots, \widehat{\xi}_{i,8})^\top$， $i = 1, \ldots, 1947$。最后，我们基于治疗组和安慰剂组的得分实施多元 Hotelling T 平方检验，分别采用方差相等（公式 (4.6)）和方差不等（公式 (4.7)）的假设。表 4 给出了基于 F 分位数和置换检验的检验统计量与 p值。所有 p值都相近，表明在所有认知领域中，吡格列酮组与安慰剂组的认知得分没有显著差异。这与基于 TOMMORROW研究发表的其他结果一致：与安慰剂组相比，吡格列酮并未显著改善认知表现（Burns 等人，2021）。

图4展示了治疗在全部五个认知领域中的估计平滑效应，以及95%基于自助法标准误的同时置信带（浅绿色阴影部分）。结果基于10,000个受试者自助法样本得出。由于所有置信带均包含零函数，这表明吡格列酮效应不显著。这进一步增强了基于投影的检验所得出的结论。

方差	T_n	p值	
		F检验	置换
相等	10.64	0.226	0.223
不相等	10.65	0.226	0.226

表 4. 基于F分布的临界值检验和置换检验的检验统计量及 Hotelling T 平方检验的 p值。检验在两种设定下进行：i) 方差相等；ii) 方差不等。置换检验的 p值基于从零分布中抽取的 10000 个样本计算得出。

7. DISCUSSION

本文提出了一种针对稀疏观测多元函数型数据的基于投影的双样本推断程序。该检验的主要优势在于，它只返回一个p值，用于判断任意一个 q组分的均值函数是否存在显著差异。因此，它避免了分别比较各组分均值时需要对多个p值进行校正的问题。我们希望，这种基于投影的检验方法将日益应用于纵向设计的随机临床试验——这类试验会收集多个次要终点，以检验两个或多个组之间的治疗效果；此外，该方法也适用于观察性研究。

我们的检验程序通过应用多元函数主成分分析，降低无限维响应的维数；这种方法在函数型数据分析文献中很受欢迎。主要的计算复杂度来自多元数据函数主成分分析的相关计算，其随响应的维数 q 以立方速率增长。为此，本文始终假设多元响应的维数固定。然而，当响应的维数变得任意大时，就需要新的计算技术来实施函数主成分分析。将该程序扩展到 q 增长至 ∞ 的情形还需进一步研究，因为可能需要施加额外的稀疏性条件才能进行多元函数主成分分析。完成函数主成分分析后，检验统计量和 p值的计算很快。

所提出的检验程序为纵向设计下的多维响应双样本推断问题提供了全新的视角。尽管对于稀疏数据，得分是在高斯假设下估计的，但我们的模拟研究表明，该检验对非高斯数据也具有良好的表现。需要注意的是，检验程序的性能取决于特征分量估计的质量。因此，在稀疏设计下，检验统计量收敛到渐近零假设分布的速率，主要取决于协方差分量的估计速率。

在通过多元函数主成分分析表示响应轨迹时，我们实际上假设所有结果均以相同单位测量，并且具有相近的变异范围。在第6节中，我们对结果进行了标准化，以确保这一点。然而，如果某个特定分量具有相对较大的变异性，我们可以使用 Chiou *and others*（2014年）提出的归一化多元函数型主成分分析来表示这些函数，并使用归一化函数主成分得分完成其余的检验程序。

我们这种基于投影的检验方法还有一个优点：原则上，它可以扩展到结构更复杂的多元函数型数据情形，例如，各分量函数定义在异质定义域上（如数据由一维函数和图像组成）的情形，此时可采用 Happ和Greven（2018）提出的多元函数主成分分析方法。同样，也可以通过在黎曼流形上进行函数主成分分析（Dai和M¨uller，2018），将该检验方法应用于流形值多元函数型数据的双样本推断问题。

Software

整个模拟研究的 R 代码已公开发布于 https: //github.com/SalilKoner/ProjectionTesting。

Funding

本研究得到美国国家老龄研究所的资助（拨款 AG064803、P30AG072958 和 P30AG028716，授予 Sheng Luo），以及 National Philanthropic Trust-比尔及梅琳达·盖茨基金会/阿尔茨海默病药物发现倡议（ADDI）的资助。

Acknowledgements

本研究使用的数据来自 TOMMORROW 试验（https://clinicaltrials.gov/ ct2/show/NCT02932410）。作者谨向 TOMMORROW 试验的志愿参与者、项目合作伙伴，以及临床研究中心的研究者和工作人员致以诚挚谢意。TOMMORROW 试验（NCT01931566）由武田制药公司与 Zinfandel Pharmaceutical Company（北卡罗来纳州教堂山）合作资助。作者还要感谢杜克临床研究所（DCRI）的项目团队，特别是 Rebecca Wilgus MS、Jack Shostak 博士和 Weibing Xing 博士，他们协助获取并处理本分析所用的 TOMMORROW 数据集。

References

Burns, Daniel K, Alexander, Robert C, Welsh-Bohmer, Kathleen A, Culp, Meredith, Chiang, Carl, O'Neil, Janet, Evans, Rebecca M, Harrigan, Patrick, Plass-man, Brenda L, Burke, James R 等. (2021). 吡格列酮延缓阿尔茨海默病高风险人群认知障碍的安全性和有效性（TOMMORROW）：一项预后生物标志物研究及三期随机、双盲、安慰剂对照试验。《柳叶刀·神经病学》**20**(7), 537–547.

Burns, Daniel K, Chiang, Carl, Welsh-Bohmer, Kathleen A, Brannan, Stephen K, Culp, Meredith, O'Neil, Janet, Runyan, Grant, Harrigan, Patrick, Plassman, Brenda L, Lutz, Michael 等. (2019). TOMMORROW研究：阿尔茨海默病发病延迟临床试验的设计。《阿尔茨海默病与痴呆：转化研究与临床干预》**5**,661–670.

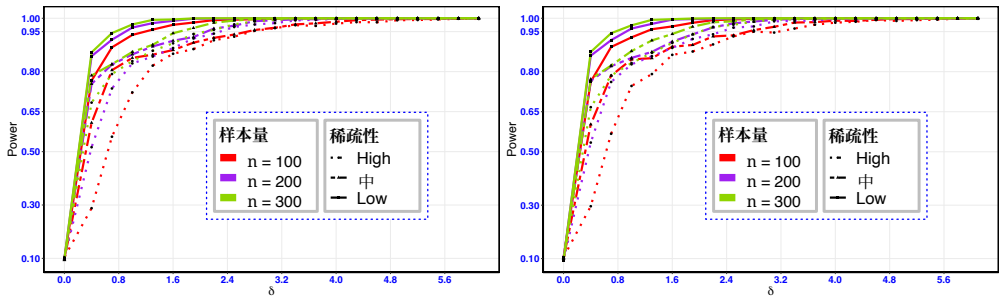
Chiou, Jeng-Min, Chen, Yu-Ting and Yang, Ya-Fang.（2014年）。多元函数主成分分析：一种标准化方法。《中华统计学志》，1571–1596。

Claeskens, Gerda, Hubert, Mia, Slaets, Leen and Vakili, Kaveh.（2014年）。多元函数半空间深度。*Journal of the American Statistical Association* **109**(505)，411–423。

Crainiceanu, Ciprian M and Ruppert, David.（2004年）。含一个方差分量的线性混合模型中的似然比测试。《英国皇家统计学会志：*B*辑》**66**(1)，165–185。

Cui, Erjia, Leroux, Andrew, Smirnova, Ekaterina and Crainiceanu, Ciprian M. (2022年). 纵向函数模型的快速单变量推断. 计算与图形统计学杂志 **31**(1), 219–230.

Curran, James and Hersh, Taylor. (2021). Hotelling：Hotelling 的 T^2 测试及其变体。R 软件包版本 1.0-8。Dai, Xiongtao and Müller, Hans-Georg. (2018). 黎曼流形和球面上函数型数据的主成分分析。The Annals of Statistics 46(6B), 3334–3361。Górecki, T and Smaga, Ł. (2017). 函数型数据的多元方差分析。《应用统计学杂志》44(12), 2172–2189。Happ, Clara and Greven, Sonja. (2018). 定义在不同（维数）域上的数据的多元函数型主成分分析。Journal of the American Statistical Association 113, 649–659。Härdle, Wolfgang Karl and Simar, Léopold. (2019). 应用多元统计分析。Springer Nature。Hutson, Vivian, Pym, J and Cloud, M. (2005年). 泛函分析与算子理论的应用。Elsevier。Jiang, Qing, Meintanis, Simos G and Zhu, Lixing. (2017). 多元函数型数据的双样本检验。载于：《函数型统计及相关领域》。Springer，第145–154页。Johnson, Richard Arnold and Wichern, Dean W. (2002). 应用多元统计分析，第5卷。Prentice Hall，新泽西州 Upper Saddle River。Kieburtz, Karl, Tilley, Barbara C, Elm, Jordan J, Babcock, Debra, Hauser,Robert, Ross, G Webster, Augustine, Alicia H, Augustine, Erika U, Aminoff,Michael J, Bodis-Wollner, Ivan G 等. (2015). 一水肌酸对帕金森病患者临床进展的影响：一项随机临床试验。《美国医学会杂志》313(6), 584–593。Krzyśko, Mirosław and Smaga, Łukasz. (2021). 使用特征函数对函数型数据进行双样本检验。《奥地利统计学杂志》50(4), 53–64。Li, Cai and Xiao, Luo. (2021). mfaces：多元稀疏函数型数据的快速协方差估计。R 软件包版本 0.1-3。Li, Cai, Xiao, Luo and Luo, Sheng. (2020年). 多元稀疏函数型数据的快速协方差估计。《Stat》9(1), e245。Muirhead, Robb J. (2009). 多元统计理论概述。John Wiley & Sons。Nel, DG and Van der Merwe, CA. (1986). 多元 Behrens-Fisher 问题的一种解法。《统计学通讯：理论与方法》15(12), 3719–3735。Park, So Young, Staicu, Ana-Maria, Xiao, Luo and Crainiceanu, Ciprian M. (2018). 复杂函数型模型的简易固定效应推断。《生物统计学》19(2), 137–152。Pomann, Gina-Maria, Staicu, Ana-Maria and Ghosh, Sujit. (2016). 一种针对函数型数据的双样本无分布检验及其在多发性硬化症扩散张量成像研究中的应用。《皇家统计学会杂志：C辑（应用统计）》65(3), 395–414。



（a）高斯数据　　（b）非高斯数据

图1。所提出的检验的经验功效随 δ（横轴）的变化情况。（a）左图对应高斯得分情形，（b）右图对应非高斯情形。δ 度量了偏离 H_0的程度。

Qiu、 Zhiping、 Chen、 Jianwei and Zhang、 Jin-Ting。（2021）。具有应用的多元函数型数据两样本测试。计算统计与数据分析 **157**，107160。

Qu、 Zhuo、 Dai、 Wenlin and Genton、 Marc G。（2021）。具有环境应用的稳健函数型多元方差分析。*Environmetrics* **32**（1），e2641。

Wang, Qiyao. (2021). 稀疏函数型数据的双样本推断。*Electronic Journal ofStatistics* **15**(1), 1395–1423.

Wood, S. N. (2011). 半参数广义线性模型的快速稳定限制最大似然和边际似然估计。*Journal of the Royal Statistical Society(B)* **73**(1), 3–36.

Yao, Fang, Müller, Hans-Georg and Wang, Jane-Ling. (2005年). 稀疏纵向数据的函数型数据分析。*Journal of the American statistical association* **100**(470), 577–590.

Zhang, Jin-Ting and Chen, Jianwei. (2007). 函数型数据的统计推断。*TheAnnals of Statistics* **35**(3), 1052–1079.

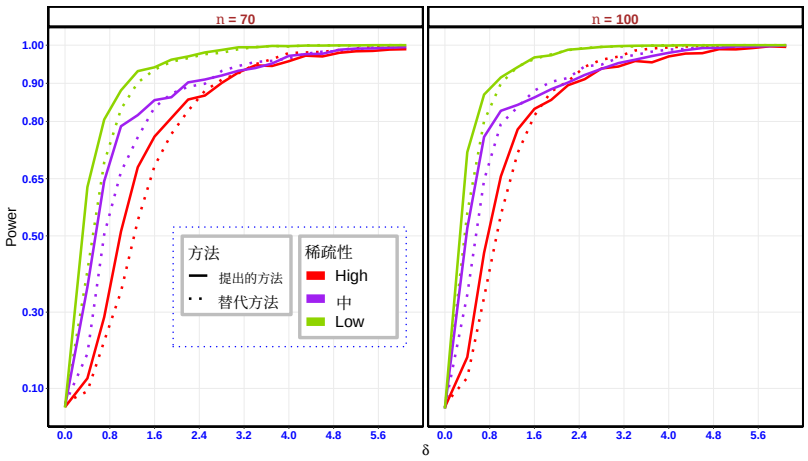
Supplementary Material

补充材料详细描述了多元函数主成分得分（多元函数主成分）的估计方法，该方法是我们基于投影的检验的构建基石；此外，还包含支持TOMORROW 研究分析的补充文件。

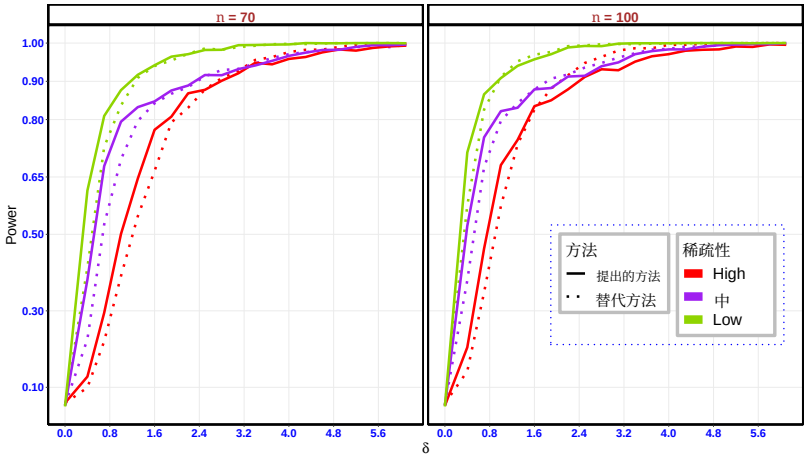
S7.1 数据驱动投影的估计

对于稀疏观测函数型数据，在混合模型下通过最佳线性无偏预测（BLUP）获得得分（Yao 等人，2005年）

$$\mathbf{Y}_i = \boldsymbol{\mu}_i + \sum_{k=1}^{K} \boldsymbol{\psi}_{i,k}\xi_{i,k} + \boldsymbol{\epsilon}_i, \tag{S8}$$



（a）高斯数据



（b）非高斯数据

图2. 所提出的检验的经验功效随 δ（横轴）变化的情况，并与 Pomann *and others*（2016）提出的多重检验程序进行比较。（a）上图对应高斯得分情形，（b）下图对应非高斯情形。量 δ 衡量与 H_0 的偏离程度。

其中，$\mathbf{Y}_i = (Y(1)_{i1}, \ldots, Y(1)_{im_i}, \ldots, Y(q)_{i1}, \ldots, Y(q)_{im_i})^\top$ 为第 i个受试者的响应 $q \times m_i$维堆叠向量，$\boldsymbol{\mu}_{i,k} = (\boldsymbol{\mu}_{i\,k}^{(1)^\top}, , \ldots, \boldsymbol{\mu}_{i\,k}^{(q)^\top},)^\top$（其中 $\boldsymbol{\mu}_i^{(\ell)}{}_{,k} = (\mu^{(\ell)}(t_{i1}) \ldots, \mu^{(\ell)}(t_{im_i}))^\top$）和$\boldsymbol{\psi}_{i,k} = (\boldsymbol{\psi}_{i\,k}^{(1)^\top}, , \ldots, \boldsymbol{\psi}_{i\,k}^{(q)^\top},)^\top$（其中 $\boldsymbol{\psi}_i^{(\ell)}{}_{,k} = (\psi^{(\ell)}(t_{i1}) \ldots, \psi^{(\ell)}(t_{im_i}))^\top$）分别是均值和特征函数在第 i 个受试者的观测点处取值后组成的 $q \times m_i$维堆叠向量，$\boldsymbol{\epsilon}_i$ 为测量误差。在得分和测量误差服从高斯假设的条件下，模型（S8）中 $\boldsymbol{\xi}_i$ 的最佳线性无偏预测（BLUP）形式为

$$\widetilde{\boldsymbol{\xi}}_i := \mathbb{E}(\boldsymbol{\xi}_i \mid \mathbf{Y}_i) = \mathrm{diag}(\lambda_1, \ldots, \lambda_K) \boldsymbol{\Psi}_i^\top \mathbf{G}_{\mathbf{Y}_i}^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i), \tag{S9}$$

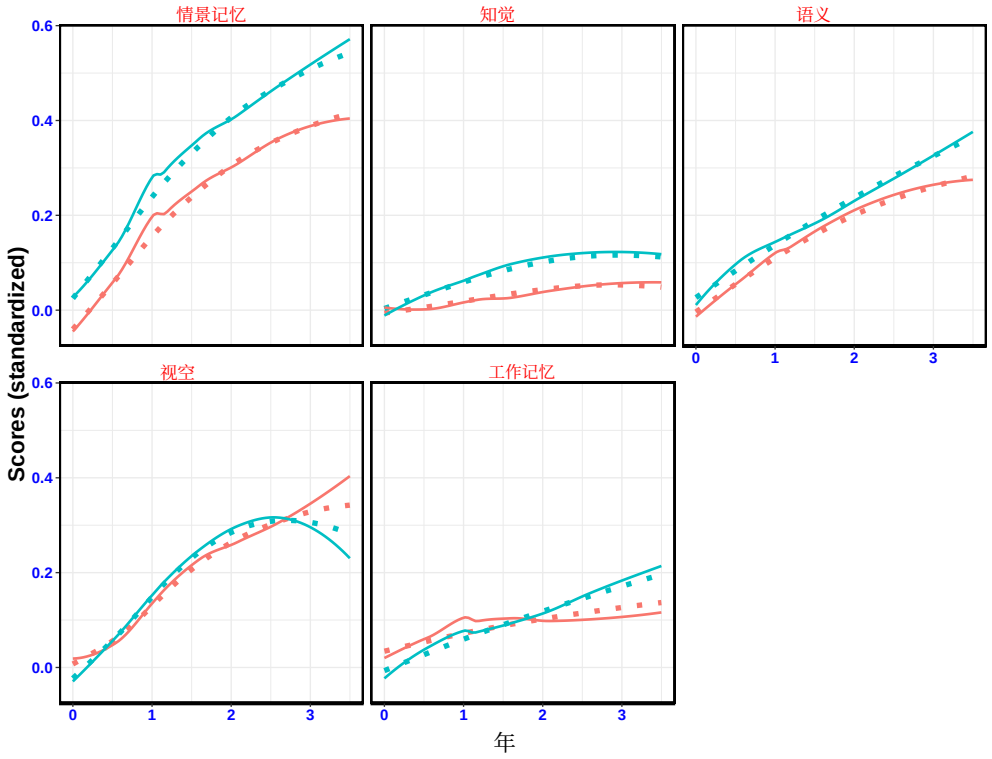


图 3. 在五个认知域中，吡格列酮组（红色）和安慰剂组（蓝色）的观测表现得分的局部加权散点图平滑曲线（实线）以及预测平滑得分（虚线）。预测得分由 R 中 mgcv 软件包的 gam()函数得出。

其中，$\mathbf{G}_{\mathbf{Y}_i} = \mathrm{Cov}(\mathbf{Y}_i) = \{\Sigma_{\ell,\ell'}(t_{ij}, t_{ij'}) + \tau_\ell^2 \mathbb{I}(j = j',\ \ell = \ell')\}_{1 \leqslant \ell,\ell' \leqslant q,\ 1 \leqslant j,j' \leqslant m_i}$ 为 $\mathbf{Y}_i$ 的协方差矩阵，而 $\mathbf{\Psi}_i = (\boldsymbol{\psi}_{i,1}, \ldots, \boldsymbol{\psi}_{i,K})$ 是 $\{\boldsymbol{\psi}_{i,k}\}_{k=1}^K$'s 的按列堆叠版本。最佳线性无偏预测$\tilde{}$估计量 $\boldsymbol{\xi}_i$ 是对真实得分 $\boldsymbol{\xi}_i$ 的一致估计量；随着每位受试者的观测次数增加且测量误差减小，这一性质成立。还需注意的是，与真实得分不同，$\{\widetilde{\xi}_{i,k}\}_{k=1}^K$ 在 k之间也不互不相关。

通过代入$\boldsymbol{\mu}_i$、$\mathbf{G}_{\mathbf{Y}_i}$、$\{\lambda_k\}_{k=1}^K$ 和 $\mathbf{\Psi}_i$的估计量，可以经验地得到方程（S9）中的最佳线性无偏预测估计量。这需要估计协方差函数 $\{\Sigma_{\ell,\ell'}(t,t')\}_{1 \leqslant \ell,\ell' \leqslant q}$、测量误差方差 $\{\tau_\ell^2\}_{\ell=1}^q$ 以及协方差的特征分量。首先，在工作独立假设下，对所有受试者的第 ℓ个坐标 $\{Y(\ell)_{ij}$的响应进行平滑，以得到均值函数 $\mu^{(\ell)}(t)$的平滑估计量：$j = 1, \ldots, m_i\}_{i=1}^n$，进而得到残差 $E_{ij}^{(\ell)} = Y(\ell)_{ij} - \widehat{\mu}^{(\ell)}(t_{ij})$。对于稀疏数据，需要汇集所有受试者的数据，以获得 $\mathbf{\Sigma}(t,t')$的平滑估计量。与单变量函数数据不同，这需要估计大型协方差矩阵。Li等人（2020）给出了一种计算高效的 $\mathbf{\Sigma}(t,t')$估计实现方法。简而言之，交叉协方差函数建模为单变量基函数的张量积，即$\Sigma_{\ell\ell'}(t,t') = \mathrm{B}^\top(t)\mathbf{\Gamma}_{\ell\ell'}\mathrm{B}(t')$，其中 $\mathrm{B}(t) = (B_1(t) \cdots, B_r(t))^\top$ 是一个长度为 r的基函数向量，$\mathbf{\Gamma}_{\ell\ell'}$ 是系数矩阵；通过将“原始协方差” $\{\widetilde{E}_{ij}^{\ell}\widetilde{E}_{ij'}^{\ell'}\}_{i,j}$作为伪响应，可以得到 $\mathbf{\Gamma}_{\ell\ell'}$ 的惩罚回归估计量

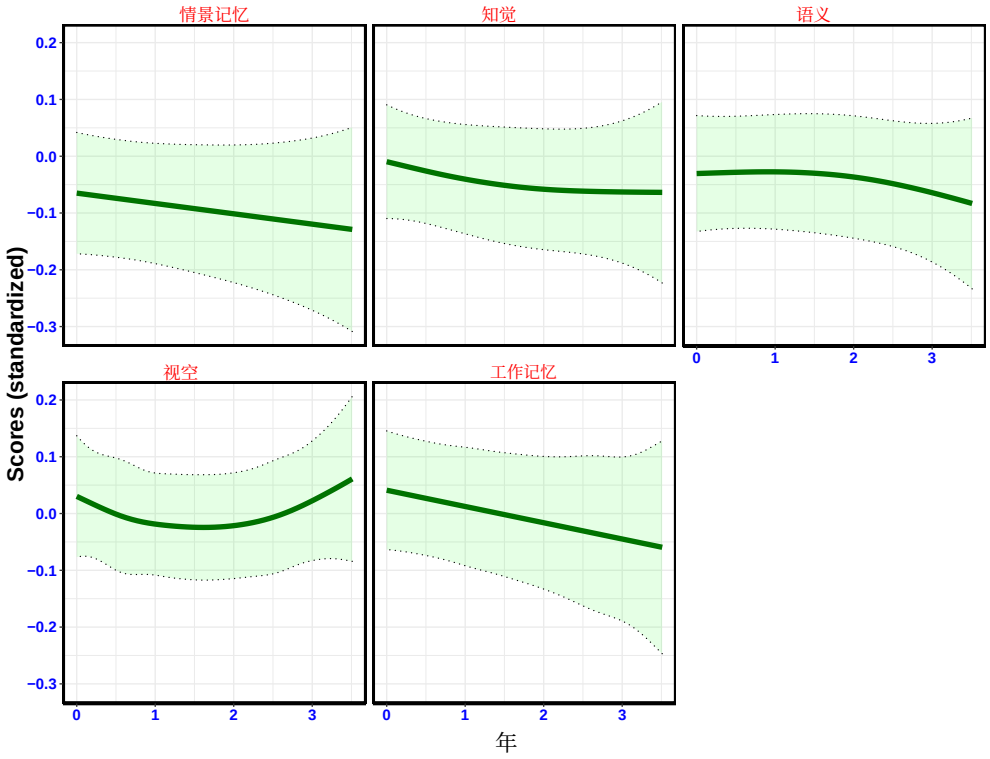


图 4. 吡格列酮组与安慰剂组在五个认知域上的均值表现得分差异的95%自助法置信带（以浅绿色阴影表示）。粗线表示估计的处理效应 $\widehat{\eta}^{(\ell)}(t)$。

对所有 $1 \leqslant \ell \leqslant \ell' \leqslant q$ 并令 $\mathbf{\Gamma}_{\ell\ell'} = \mathbf{\Gamma}_{\ell'\ell}^{\top}$。特征函数可高效地估计为 $\widehat{\psi}_k(t) = (\mathbf{I}_q \otimes \mathbf{B}^{\top}(t)\mathbf{S}^{-\frac{1}{2}})\widehat{\mathbf{V}}_k$，其中$\mathbf{S} = \int \mathbf{B}(u)\mathbf{B}^{\top}(u)du$，且 $\{\widehat{\mathbf{V}}_k\}_k$ 是块矩阵的特征向量，其元素为$\mathbf{S}^{\frac{1}{2}}\widehat{\mathbf{\Gamma}}_{\ell\ell'}\mathbf{S}^{\frac{1}{2}}$,，其中 $1 \leqslant \ell, \ell' \leqslant q$。特征函数的最优数量$\widehat{K} = K(n)$通过选择 K 的最小值来确定，使得解释方差比例（PVE）高于预先设定的阈值，其定义为 $\sum_{k=1}^{K}\widehat{\lambda}_k / \sum_{k=1}^{\infty}\widehat{\lambda}_k$。其他用于优化选择 K 的方法包括交叉验证、赤池信息准则（AIC）和贝叶斯信息准则（BIC）。

一旦特征分量 $\{\widehat{\lambda}_k, \widehat{\psi}_k(t)\}_{k=1}^{\widehat{K}}$ 从数据中估计得出后，得分通过将均值和方差成分的估计量代入公式 (S9) 进行经验估计，从而得到

$$\widehat{\boldsymbol{\xi}}_i = \widehat{\mathbf{V}}_k^{\top}(\mathbf{I}_q \otimes \mathbf{S}^{\frac{1}{2}})\widehat{\mathbf{\Gamma}}(\mathbf{I}_q \otimes \mathbf{B}_i^{\top})\widehat{\mathbf{G}}_{\mathbf{Y}_i}^{-1}(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i), \tag{S10}$$

其中 $\mathrm{B}_i = (\mathrm{B}(t_{i1}) \cdots, \mathrm{B}(t_{im_i}))^{\top}$, $\widehat{\mathbf{\Gamma}} = ((\widehat{\mathbf{\Gamma}}_{\ell,\ell'}))_{1\leqslant\ell,\ell'\leqslant q} \in \mathbb{R}^{qr\times qr}$和 $\widehat{\mathrm{G}}_{\mathbf{Y}_i} = (\mathrm{I}_q \otimes \mathrm{B}_i)\widehat{\mathbf{\Gamma}}(\mathrm{I}_q \otimes \mathrm{B}_i^{\top}) + \mathrm{blockdiag}(\widehat{\tau}_1^2\mathrm{I}_{m_i}, \ldots, \widehat{\tau}_q^2\mathrm{I}_{m_i})$。因此，该方法无需针对每个 $i = 1$、.....、n 通过求逆大型 $q \times m_i$ 乘 $q \times m_i$ 协方差矩阵，而只需对 $qr \times qr$ 矩阵 $\mathbf{\Gamma}$ 求逆一次，从而减轻了多元函数型数据得分估计的计算负担，尤其是在 q 相对较大时。

	高风险 吡格列酮 (N = 985)	高风险 安慰剂 (N = 962)	总体 (N = 1947)
年龄（年）			
均值（标准差）	74.7（5.20）	74.9 (5.19)	74.8 (5.19)
范围	65 - 83	65 - 83	65 - 83
Sex			
女性	543 (55.1%)	579 (60.2%)	1122 (57.6%)
Male	442 (44.9%)	383 (39.8%)	825（42.4%）
体重指数（单位：mg/m2）			
均值（标准差）	28.0（5.15）	28.0（5.35）	28.0（5.25）
范围	14 - 68	16 - 52	14 - 68
缺失	2 (0.2%)	1 (0.1%)	3 (0.2%)

表5. TOMORROW 研究中两组参与者的基线人口统计学特征。

S7.2 支持 *TOMORROW* 研究分析的其他结果

本节将介绍一些与 TO-MORROW 研究基线分析相关的补充表格。

基线分析： 研究纳入参与者的平均年龄为74年，标准差为5年；吡格列酮组和安慰剂组的情况与此相似。两组中，队列约有60%的参与者年龄超过75岁。吡格列酮组约55%的参与者为女性，而安慰剂组约60%为女性。各组的基线身高（cm）、体重（kg）和体重指数（BMI）（mg/m2）范围相近，平均身高约为167 cm（标准差10 cm），平均体重约为78 kg（标准差17 kg），平均BMI约为28 mg/m2（标准差5 mg/m2）。表5所列的基线分析表明，两组在关键基线人口统计学特征方面具有可比性。