arXiv:2302.00841v1 [stat.AP] 2 Feb 2023

# 纵向典型相关分析

Seonjoo Lee、Jongwoo Choi、Zhiqian Fang 和 F. DuBois Bowman 代表阿尔茨海默病神经影像学倡议*
哥伦比亚大学和纽约州精神病学研究所，美国纽约。密歇根大学生物统计学系，美国纽约。
电子邮箱：seonjoo.lee@nyspi.columbia.edu

**摘要**。本文探讨了两个纵向变量的典型相关分析，这些变量可能以不同的时间分辨率在不规则网格上采样。我们使用随机效应对多变量的轨迹进行建模，并在潜在空间中找到了相关性最高的线性组合。数值模拟表明，纵向典型相关分析（LCCA）能够有效恢复两个高维纵向数据集之间潜在的相关模式。我们将所提出的 LCCA 应用于阿尔茨海默病神经影像学倡议的数据，并识别出脑形态变化和淀粉样蛋白累积的纵向特征。

关键词：典型相关分析、纵向数据分析、阿尔茨海默病

## 1. 引言

典型相关分析（CCA）旨在寻找两组多变量之间的相关结构。CCA 在每组变量中寻找线性组合，使所得线性组合之间的相关性最大，同时与任一组中的所有其他线性组合正交。近年来，研究者通过降维（Song 等人，2016年）、惩罚化（Avants 等人，2010年；Bao 等人，2019年；Fang 等人，2016年；Gossmann 等人，2018年；Lin 等人，2014年；Wilms 和 Croux，2016年；Witten 等人，2009年）以及结合多个数据集（Deleus 和 Van Hulle，2011年；Zhang 等人，2014年；Kim 等人，2019年），将典型相关分析应用于高维多变量。

典型相关分析已广泛应用于医学领域，用于识别临床、行为、影像和遗传变量之间的关联。例如，典型相关分析已被用于探索临床症状与行为测量（Mihalik et al.，2019）、脑功能连接与认知缺陷（Adhikari et al.，2019）或临床症状（Kang et al.，2016；Grosenick et al.，2019）、两组影像数据（Avantset al.，2010）以及基因与影像关联（Lin etal.，2014；Kim et al.，2019）。

*本文所用数据来自阿尔茨海默病神经影像学计划（ADNI）数据库（adni.loni.usc.edu）。因此，ADNI 的研究人员参与了 ADNI 的设计与实施，和/或提供了数据，但未参与本报告的分析或撰写。ADNI 研究人员完整名单请见：http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf。

当其中一个多变量随时间测量时，可通过组套索惩罚的多组稀疏典型相关方法扩展CCA（Hao 等，2017）以及时间多任务 SCCA（T-MTSCCA）（Du 等，2019）。然而，现有方法无法处理缺失值、不规则时间采样以及两个变量之间的时间错位。

为弥补这一不足，我们提出一种新的 LCCA 方法，用于识别能够最大化纵向轨迹之间关联的典型变量模式。我们使用随机效应（例如随机截距和随机斜率）对每个变量的纵向轨迹进行建模。随后，求出能够最大化相关性的随机效应线性组合。由于两个多变量的维度可能大于样本量，我们通过特征分解进行降维，并在纵向主成分分析框架下实施估计。我们开展了大量模拟实验，以评估 LCCA 的性能。此外，我们还将 LCCA 应用于阿尔茨海默病神经影像学倡议（ADNI）队列的数据（Mueller等人，2005；Weiner等人，2012）。ADNI数据包含为期八年的纵向 $^{18}$F-AV- 45（氟贝他吡）正电子发射断层显像（PET）图像，用于量化脑淀粉样蛋白负荷，以及同一随访期间采集的结构磁共振成像数据。我们的 LCCA 方法在维度选择方面表现稳定，能够准确识别信号，并可识别ADNI数据中不同的阿尔茨海默病相关脑模式。

## 2. 方法

### *2.1.* 纵向典型相关分析

对于每位受试者 $i = 1, ..., n$ 以及各次访视 $j = 1, \ldots, J_i$，我们在时间 $t_{ij}$观测到一个 $p$ 维向量$\mathbf{X}_{ij}$。类似地，对于各次访视 $k = 1, \ldots, K_i$，我们在时间 $s_{ik}$观测到一个 $q$ 维向量 $\mathbf{Y}_{ik}$。为便于说明，我们考虑每个变量的线性轨迹，随后通过样条回归将其推广到非线性轨迹。我们使用随机截距和斜率模型对每次观测进行建模：

$$\mathbf{X}_{ij} = \mu_{ij}^X + \mathbf{Z}_{i,0}^X + t_{ij}\mathbf{Z}_{i,1}^X + \boldsymbol{\epsilon}_{ij}^X, \tag{1}$$

$$\mathbf{Y}_{ik} = \mu_{ik}^Y + \mathbf{Z}_{i,0}^Y + s_{ik}\mathbf{Z}_{i,1}^Y + \boldsymbol{\epsilon}_{ik}^Y, \tag{2}$$

其中，$\mu_{ij}^X$ 和 $\mu_{ik}^Y$ 为固定效应；$\mathbf{Z}_i^X$,0 和 $\mathbf{Z}_i^Y$,0 为随机截距；$\mathbf{Z}_i^X$,1 和 $\mathbf{Z}_i^Y$,1 为随机斜率；$\boldsymbol{\epsilon}_{ij}^X$ 和 $\boldsymbol{\epsilon}_{ik}^Y$ 为误差。随机效应按顺序排列为一个 $2p$ 维向量 $\mathbf{Z}_i^X := (\mathbf{Z}_i^X{}_{,0}', \mathbf{Z}_i^X{}_{,1}')' \in \mathbb{R}^{2p}$，以及一个 $2q$ 维向量 $\mathbf{Z}_i^Y := (\mathbf{Z}_i^Y{}_{,0}', \mathbf{Z}_i^Y{}_{,1}')' \in \mathbb{R}^{2q}$；假设它们服从均值为零、协方差为 $\boldsymbol{\Sigma}_c = \begin{bmatrix} \boldsymbol{\Sigma}_c^{(0,0)} & \boldsymbol{\Sigma}_c^{(0,1)} \\ \boldsymbol{\Sigma}_c^{(1,0)} & \boldsymbol{\Sigma}_c^{(1,1)} \end{bmatrix}$ 的分布，其中 $\boldsymbol{\Sigma}_c^{(a,b)} = \mathbb{E}\left(\mathbf{Z}_{i,a}^c \mathbf{Z}_{i,b}^{c\prime}\right)$, $a, b \in \{0, 1\}$, $c \in \{X, Y\}$，并且与 $\boldsymbol{\epsilon}_{ij}^X$ 和 $\boldsymbol{\epsilon}_{ik}^Y$ 不相关。假设随机截距、斜率和误差均服从高斯分布。

因此，定义典型相关的目标函数为

$$r_m^{XY} = max_{\mathbf{u}_m \in \mathcal{R}^{2p}, \mathbf{v}_m \in \mathcal{R}^{2q}} < \mathbf{u}_m' \mathbf{Z}_i^X, \mathbf{v}_m' \mathbf{Z}_i^Y > . \tag{3}$$

当 $p$ 大于 $n/2$时，我们可以采用主成分分析来表示$\mathbf{Z}_i^X \approx \mathbf{\Phi}_X \boldsymbol{\xi}_i^X$，其中 $\mathbf{\Phi}_X = \left(\mathbf{\Phi}_0^{X\prime}, \mathbf{\Phi}_1^{X\prime}\right)'$ 是一个由 $\mathbf{\Sigma}_X = Var(\mathbf{Z}_i^X)$的前 $N_X$ 个特征向量组成的 $2p \times N_X$ 矩阵，而 $\boldsymbol{\xi}_i^X = (\xi_{i1}^X, \ldots, \xi_{iN_X}^X)' \in \mathbb{R}^{N_X}$ 是相应的个体特异性特征得分。采用纵向主成分分析（LPCA）（Greven等人，2011年；Zipunnikov等人，2014年；Lee等人，2015年），每个项可表示为

$$\mathbf{X}_{ij} \approx \mu_{ij}^X + \sum_{l=1}^{N_X} \xi_{il}^X \left(\Phi_l^{X,0} + t_{ij}\Phi_l^{X,1}\right) + \epsilon_{ij}^X, \tag{4}$$

其中 $(\xi_{il_1}^X,\ \xi_{il_2}^X) \sim (0,\ 0;\ \lambda_X^{l_1},\ \lambda_X^{l_2};\ 0)$，其中"$\cdot \sim (\mu_1,\ \mu_2;\ \sigma_1^2,\ \sigma_2^2;\ \rho)$"表示一对高斯变量服从均值为（$\mu_1$，$\mu_2$）、方差为（$\sigma_1^2$，$\sigma_2^2$）且相关系数为 $\rho$的分布。类似地，$\mathbf{Y}_{ik}$ 表示为

$$\mathbf{Y}_{ik} \approx \mu_{ik}^Y + \sum_{l=1}^{N_Y} \xi_{il}^Y \left(\Phi_l^{Y,0} + s_{ik}\Phi_l^{Y,1}\right) + \epsilon_{ik}^Y, \tag{5}$$

其中 $(\xi_{il_1}^Y,\ \xi_{il_2}^Y) \sim (0,\ 0;\ \lambda_Y^{l_1},\ \lambda_Y^{l_2};\ 0)$。通过对协方差矩阵进行最小二乘估计，可以得到（4）和（5）的特征向量。不失一般性，假设$\mathbf{X}_{ij}$ 已去均值，均值为零。X × 的p阶协方差为$\mathbf{X}_{ij_1}$ 与$\mathbf{X}_{ij_2}$，其表达式为

$$E\mathbf{X}_{ij_1}\mathbf{X}_{ij_2}^\top = \mathbf{\Sigma}_X^{(0,0)} + t_{ij_2}\mathbf{\Sigma}_X^{(0,1)} + t_{ij_1}\mathbf{\Sigma}_X^{(1,0)} + t_{ij_1}t_{ij_2}\mathbf{\Sigma}_X^{(1,1)} + \delta_{j_1,j_2}\mathbf{\Sigma}_X^{\epsilon}, \tag{6}$$

其中，$\mathbf{\Sigma}_X^{\epsilon}$ 是 $\epsilon_{ij}^X$ 和$\delta_{j_1}$的协方差矩阵；若 $j_1 = j_2$ 和$\delta_{j_1}$，则为$j_2 = 1$；否则为$j_2 = 0$。将 $vec(\cdot)$定义为矩阵的向量化，即将矩阵的各列依次堆叠。然后，我们可以构造矩阵$\mathbf{K}_X = \{vec(\mathbf{\Sigma}_X^{(0,0)})\ vec(\mathbf{\Sigma}_X^{(0,1)})\ vec(\mathbf{\Sigma}_X^{(1,0)})\ vec(\mathbf{\Sigma}_X^{(1,1)})\ vec(\mathbf{\Sigma}_X^{\epsilon})\}$和$\mathbf{f}_{ij_1j_2} = (1, t_{ij_2}, t_{ij_1}, t_{ij_1}t_{ij_2}, \delta_{j_1,j_2})^\top$，使得 $Evec(\mathbf{X}_{ij_1}\mathbf{X}_{ij_2}^\top) = \mathbf{K}_X\mathbf{f}_{ij_1j_2}$。将所有受试者和访视的向量连接起来后，我们得到矩阵X的矩矩阵恒等式：$\mathbf{X} = \mathbf{K}_X\mathbf{F}$，其中 $J = \sum_{i=1}^N J_i^2$。随后可通过普通最小二乘法（OLS）对协方差参数$\mathbf{K}_X$进行无偏估计：$\hat{\mathbf{K}}_X = \mathbf{X}\mathbf{F}^\top(\mathbf{F}\mathbf{F}^\top)^{-1}$。给定估计得到的协方差矩阵后，通过特征分解计算特征向量。给定特征向量后，如Greven等人（2011年）和Zipunnikov等人（2014年）所述，根据方程（4）和（5）使用最佳线性无偏预测量（BLUP）估计特征得分。

记矩阵 $N \times N_X$ $\boldsymbol{\xi}^X = \{\xi_{il}^X\}_{i=1\ldots N; l=1\ldots N_X}$ $N \times N_Y$ $\boldsymbol{\xi}^Y = \{\xi_{il}^Y\}_{i=1\ldots N; l=1\ldots N_Y}$ 和矩阵 $\ldots$。那么，定义典型相关的目标函数为

$$r_m^{XY} = max_{\mathbf{u}_m \in \mathcal{R}^{N_X}, \mathbf{v}_m \in \mathcal{R}^{N_Y}} < \boldsymbol{\xi}^X\mathbf{u}_m, \boldsymbol{\xi}^Y\mathbf{v}_m > \tag{7}$$

在限制条件下最大化式（7）中的目标函数：每个 $\mathbf{u}_m$ 均与低阶 $\mathbf{u}_{m'}$ 正交，其中 $1 \le m < m' \le min(N_X, N_Y)$。第二组变量中的 $\mathbf{v}_m$ 也适用相同的限制。为实现识别，要求对典型变量 $\mathbf{u}_m^\top\boldsymbol{\xi}_i^X$、$\mathbf{v}_m^\top\boldsymbol{\xi}_i^Y, m = 1, \ldots, min(N_X, N_Y)$ 进行归一化，使其具有单位方差。在低维空间中估计典型向量，并将第 $m$ 个纵向典型向量计算为 $\left(\mathbf{\Phi}_0^X + t_{ij}\mathbf{\Phi}_1^X\right)\mathbf{u}_m$ 和 $\left(\mathbf{\Phi}_0^Y + s_{ik}\mathbf{\Phi}_1^Y\right)\mathbf{v}_m$，其中 $m = 1, ..., min(N_X, N_Y)$。

根据时间点的可用情况，这种构造可以自然地扩展到高阶趋势，或通过 B样条基展开进行建模。假设位置 $v$ 处成像测量的轨迹由一个样条函数表示，其节点序列 $\tau_1 < \cdots < \tau_D$ 和次数 $d$ 均固定。

$$\mathbf{X}_{ij}(v) = \mu^X(v) + \sum_{r=1}^{D+d+1} Z_{i,r}(v) b_r(t_{ij}) + \epsilon_{ij}^X, \tag{8}$$

其中， $b_r$ 为一组基函数，$Z_{i,r}(v)$为相应的样条系数。$\mathbf{Z_i} = (\mathbf{Z}_i^\top, 1, \ldots, \mathbf{Z}_{i,D+d+1}^\top)^\top$ 组成的$(D+d+1)V$维堆叠向量可通过主成分分析近似为：$\mathbf{Z_i} \approx \sum_{l=1}^{N_X} \xi_{il}^X \mathbf{\Phi}_l^X$，其每个元素表示为$Z_{i,r}(v) \approx \sum_{l=1}^{N_X} \xi_{il}^X \Phi^X{}_{,l}{}^r(v)$，其中 $r = 1, \ldots, D+d+1$。然后，将式（8）重组为

$$\mathbf{X}_{ij}(v) = \mu^X(v) + \sum_{l=1}^{N_X} \xi_{il}^X \sum_{r=1}^{D+d+1} \Phi_l^{X,r}(v) b_r(t_{ij}) + \epsilon_{ij}^X. \tag{9}$$

类似地， $Y$ 建模为

$$\mathbf{Y}_{ik}(v) = \mu^Y(v) + \sum_{l=1}^{N_Y} \xi_{il}^Y \sum_{r=1}^{D+d+1} \Phi_l^{Y,r}(v) b_r(s_{ik}) + \epsilon_{ik}^Y. \tag{10}$$

该方法不要求 $X$和 $Y$在同一时间被观测到。只要它们在研究问题所关注的时间范围内被观测到，所提出的方法就能提取关联模式。

### *2.2.* 典型协变量的数量

为确定典型相关分析的维数，我们在潜在空间中采用一种传统的似然比检验方法。从 $m = 0$ 开始，我们检验零假设 $H_0$: $d = m$ ，对照备择假设 $H_1$: $d > m$。如果 $H_0$ 被拒绝，则将 $m$ 加一，并进行新的检验。如此继续，直到 $H_0$ 未被拒绝，或 $m$ 达到 $M = min$（$N_X$，$N_Y$）。给定典型变量的数量 $m$，Wilks 检验统计量（Wilks，1935年；Bartlett，1947年；Friederichs和Hense，2003年）表示为

$$\Lambda_m = \Pi_{j=m+1}^{min(N_X, N_Y)} (1 - \hat{r}_j^2). \tag{11}$$

基于 Rao 的 F 近似（Rao等人，1973年），$F = df_2/df_1(1 - \Lambda_m/(\Lambda_m))^{1/\nu}$渐近服从 $F_{df_1,\ df_2}$，其中 $\nu = \sqrt{(df_1^2 - 4)/((N_X - m)^2 + (N_Y - m)^2 - 5)}$,$df_1 = (N_X - m) * (N_Y - m)$，且 $df_2 = (n - 1.5 - (N_X + N_Y)/2)\nu - df_1/2 + 1$。

LCCA 的性能取决于 $N_X$ 和 $N_Y$的选择。既往研究建议，根据个体成分所解释方差的 80% 阈值来确定数量（Greven等人，2011年；Lee等人，2015年），或预先指定数量（Zipunnikov等人，2014年）。实际上，性能取决于信噪比，这一点对成像应用尤为重要。数值实验表明，采用 80%–90% 的阈值作为经验法则时，性能良好。这些算法已实现为 R 软件包 LCCA，并可在 GitHub 上获取（https://seonjoo.github.io/lcca/）。

**表1.** 基线人口统计学特征及5年转化率。

| | 阿尔茨海默病 (N=24) | 轻度认知障碍 (N=365) | 认知正常 (N=291) | 总计 (N=680) | p值 |
|---|---|---|---|---|---|
| 性别，女性，n (%) | 10 (41.7) | 162 (44.4) | 147 (50.5) | 319 (46.9) | 0.257 |
| 年龄，岁，均值（标准差） | 76.75 (7.15) | 71.88 (7.74) | 74.62 (6.47) | 73.23 (7.35) | <0.001 |
| 非西班牙裔，n（%） | 22（91.7） | 351（96.2） | 279（95.9） | 652（95.9） | 0.245 |
| 种族，白人，n（%） | 23 (95.8) | 342 (93.7) | 267 (91.8) | 632 (92.9) | 0.713 |
| 简易精神状态检查量表[1],均值（标准差） | 18.83 (4.90) | 26.89 (3.34) | 28.81 (1.63) | 27.43 (3.39) | < 0.001 |
| 阿尔茨海默病评估量表[2],均值（标准差） | 37.35 (13.75) | 16.64 (11.06) | 9.05 (5.71) | 14.10 (10.93) | < 0.001 |
| 5 年阿尔茨海默病转变[3],n (%) | N/A | 104 (28.5) | 16 (5.5) | 120 (18.3) | < 0.001 |

[1] 八名参与者存在缺失值。[2] 九名参与者存在缺失值。[3] 仅分析了基线时的非痴呆参与者。

## 3. 阿尔茨海默病神经影像数据分析

### *3.1.* 参与者

我们将 LCCA 方法应用于 ADNI数据，以识别脑形态测量与淀粉样蛋白沉积之间的纵向关联。数据下载自 ADNI数据库（http://adni.loni.usc.edu）。初始阶段（ADNI-1）招募了 800 名参与者，其中约有 200 名健康对照者、400 名晚期轻度认知障碍（MCI）患者，以及 200 名临床诊断为疑似阿尔茨海默病的患者；这些参与者来自美国和加拿大 50 多个研究中心，并在 2–3 年内每隔 6–12 个月接受一次随访。此后，ADNI-GO 和 ADNI-2 对现有参与者进行了随访，并招募了包括早期轻度认知障碍患者在内的其他参与者。在 ADNI 中，要被归类为轻度认知障碍，受试者的简易精神状态检查得分须在 24 至 30 分之间，并有主观记忆障碍；根据按教育程度校正后的韦氏记忆量表逻辑记忆分测验 II得分，须有客观记忆受损证据；全球临床痴呆评定量表得分须为 0.5；不存在当前重度抑郁症等显著混杂状况；日常活动能力正常或接近正常；且无临床痴呆。

所有研究均经各自的机构审查委员会批准，且所有受试者或其代理人均提供了符合HIPAA法规的知情同意。分析总计纳入了680名受试者的PET-MRI扫描配对，平均每人接受2.7（± 0.79）次访视，平均随访时间为3.7（±1.66）年。基线时，参与者中有291名认知正常（CN）、365名轻度认知障碍（MCI）和24名阿尔茨海默病（AD）。这些个体的详细特征见表1。

### *3.2.* 结构成像处理

为确保内部一致性，采用了3.0 MPRAGE磁化准备快速梯度回波T1加权磁共振图像。使用FreeSurfer 6.0版进行横断面图像处理。使用Desikan-Killiany图谱（Desikan等，2006）对皮层区域进行自动FreeSurfer v6解剖分区，并提取感兴趣区（ROI）特异性的皮层厚度和体积测量值；共有68个感兴趣区（左、右大脑半球各34个），收集了纵向皮层厚度和体积测量值。计算了28个皮层下区域的体积测量值（Fischl等，2002），包括外侧脑室、下外侧脑室、

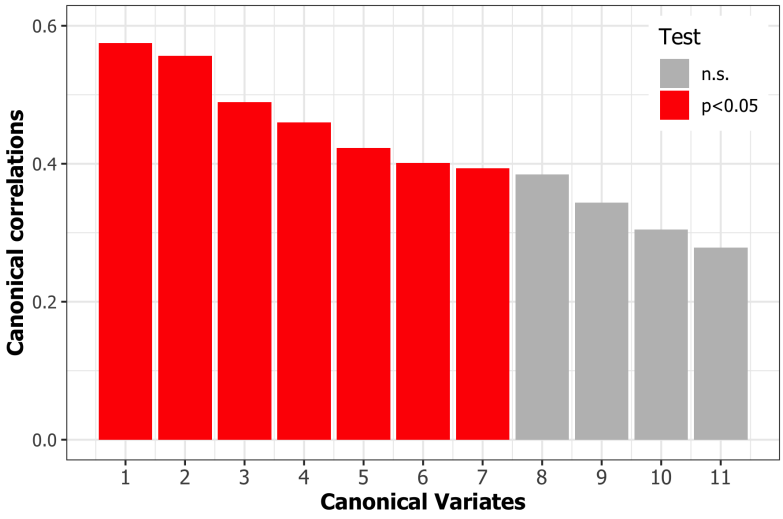


**图1**。 典型相关系数。基于 Rao 的 F 近似对 Wilks Lambda 进行 F检验，结果显示七个典型相关系数具有显著性（$p < 0.05$）。

丘脑、尾状核、壳核、苍白球、海马、杏仁核、伏隔核区、第三脑室、第四脑室以及五个胼胝体亚区。此外，还纳入了颅内容积总量。

我们使用 PetSurfer 流程计算淀粉样蛋白 SUVR 水平（Greve 等人，2014年，2016），该流程随 FreeSurfer 6 版本提供。PetSurfer流程首先将正电子发射断层显像图像与相应的 MRI 扫描进行配准，然后应用部分容积校正，最后将逐体素 SUVR 值重采样到皮层表面上。基于 Desikan 图谱，对皮层和皮层下区域计算了 81 个 ROI 层级汇总。

### *3.3.* 朴素方法

目前尚无现有的典型相关分析方法能够处理纵向数据中缺失的数据以及不规则的时间采样。因此，我们考虑了以下方法。首先，我们计算每个受试者每个变量的截距和斜率。然后，在每种模态内，将截距和斜率向量进行堆叠，并执行典型相关分析。最后，按照第 2.2 节所述的相同原则估计典型变量的数量。我们将此方法命名为朴素方法。

### *3.4.* 结果

纵向典型相关分析识别出 7 个典型变量。显著的典型相关系数显示在图 1中。纵向典型向量按时间 $\left(\mathbf{\Phi}_0^X + t\mathbf{\Phi}_1^X\right)\mathbf{u}_m$ 和 $\left(\mathbf{\Phi}_0^Y + s\mathbf{\Phi}_1^Y\right)\mathbf{v}_m$ 对于 $m = 1, \ldots, min(N_X, N_Y)$ 重新排列后的结果显示在图 2和 3 的时间点=0、1、2、3、4。为便于可视化，向量按其总方差进行标准化，并将向量中至少有一个元素大于

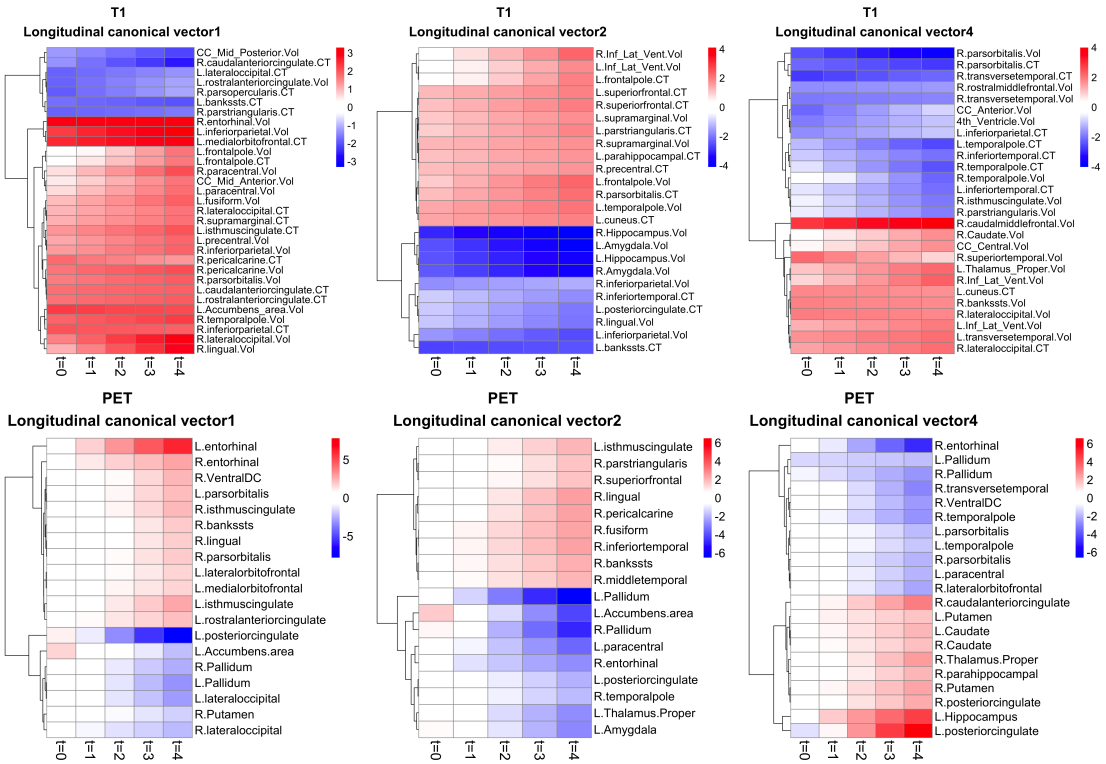


**图 2.** LCCA：纵向典型向量的配对结果显示于时间=0,1,2,3,4。

1.64 的感兴趣区纳入热图。使用完全连接法进行层次聚类分析，对各行进行聚类，以展示哪些变量随时间呈现相似变化。图2 和 3 展示了与基线诊断状态或阿尔茨海默病转变相关的六个纵向典型变量的纵向典型权重。我们还采用了朴素方法，并比较了其表现。

此外，我们还研究了纵向典型变量（CV）。首先，按各受试者的基线临床状态比较典型变量。进行组间比较时，以年龄、性别和受教育年限为协变量进行线性回归，随后进行控制假发现率的多重比较校正（Benjamini 和 Hochberg，1995）。表2 报告了 F统计量、未校正的p值和多重比较校正后的p值。所有分析均排除了超出第一和第三四分位数范围3个四分位距的极端离群值。事后两两组间比较采用最小二乘均值差。图 6 显示了抖动箱线图，以及多重比较校正后组间差异显著的典型变量的两两比较结果（未校正的p值）。

T1的典型变量2、4和7，以及正电子发射断层显像的典型变量2和5，在多重比较校正后显示出组间差异。典型变量2表明，基线时阿尔茨海默病患者的双侧海马和杏仁核出现萎缩。淀粉样蛋白沉积在双侧苍白球呈减少趋势，而在皮层则呈增加趋势

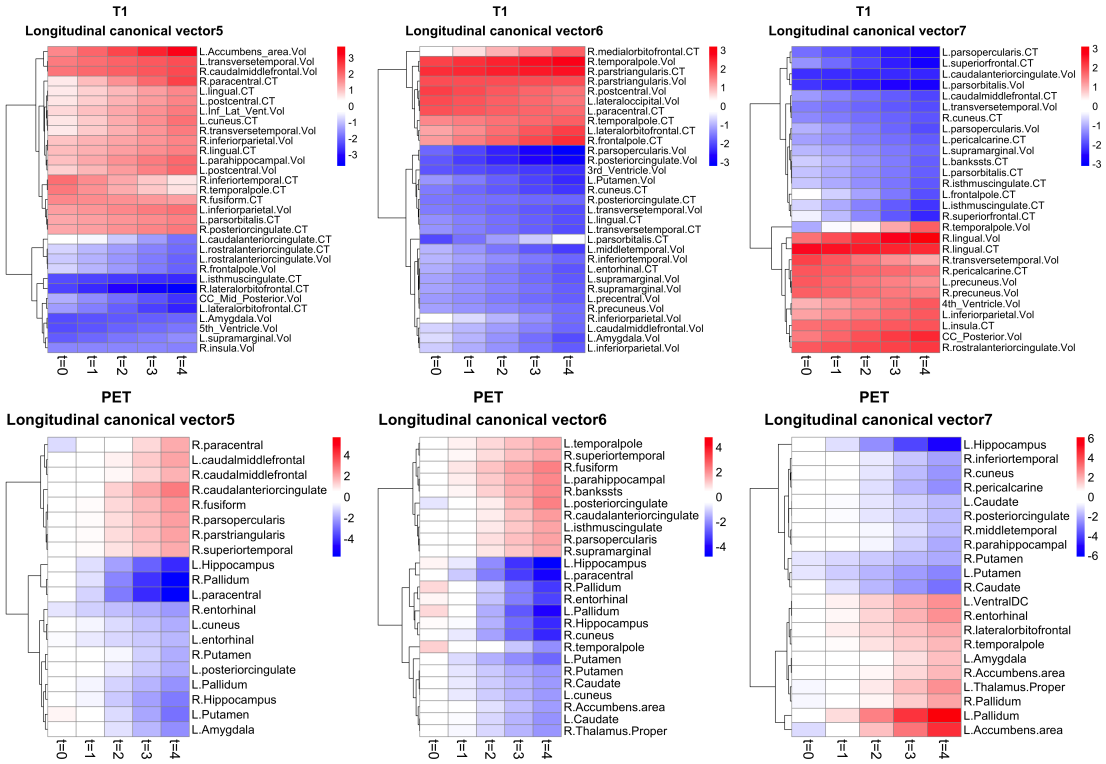


**图3.** LCCA：展示了时间=0、1、2、3、4时点的纵向典型向量对。

**表2.** 对诊断组比较和AD转化进行方差分析，并针对年龄、性别和教育程度进行调整。

| | | T1 | | | | PET | | | |
|---|---|---|---|---|---|---|---|---|---|
| | LCV | F | p | $\eta_p^2$ | 离群值 | F | p | $\eta_p^2$ | 离群值 |
| 组比较 | 1 | 3.47 | 0.032 | 0.010 | 0 | 0.57 | 0.566 | 0.002 | 1 |
| | 2 | 19.83 | <0.001 * | 0.056 | 0 | 38.73 | <0.001 * | 0.103 | 0 |
| | 3 | 2.13 | 0.120 | 0.006 | 1 | 1.08 | 0.342 | 0.003 | 2 |
| | 4 | 4.86 | 0.008 * | 0.014 | 1 | 3.34 | 0.036 | 0.010 | 2 |
| | 5 | 0.95 | 0.386 | 0.003 | 0 | 11.60 | <0.001 * | 0.033 | 2 |
| | 6 | 2.42 | 0.090 | 0.007 | 0 | 2.46 | 0.086 | 0.007 | 1 |
| | 7 | 8.26 | <0.001 * | 0.024 | 0 | 3.86 | 0.022 | 0.011 | 2 |
| AD转换 | 1 | 11.74 | 0.001 * | 0.018 | 0 | 5.76 | 0.017 * | 0.009 | 1 |
| | 2 | 102.21 | <0.001 * | 0.136 | 0 | 164.90 | <0.001 * | 0.202 | 0 |
| | 3 | 1.51 | 0.220 | 0.002 | 0 | 0.08 | 0.773 | 0.000 | 1 |
| | 4 | 6.11 | 0.014 * | 0.009 | 0 | 7.70 | 0.006 * | 0.012 | 1 |
| | 5 | 0.00 | 0.972 | 0.000 | 0 | 0.75 | 0.386 | 0.001 | 1 |
| | 6 | 3.19 | 0.074 | 0.005 | 0 | 7.05 | 0.008 * | 0.011 | 0 |
| | 7 | 2.80 | 0.095 | 0.004 | 0 | 3.44 | 0.064 | 0.005 | 2 |

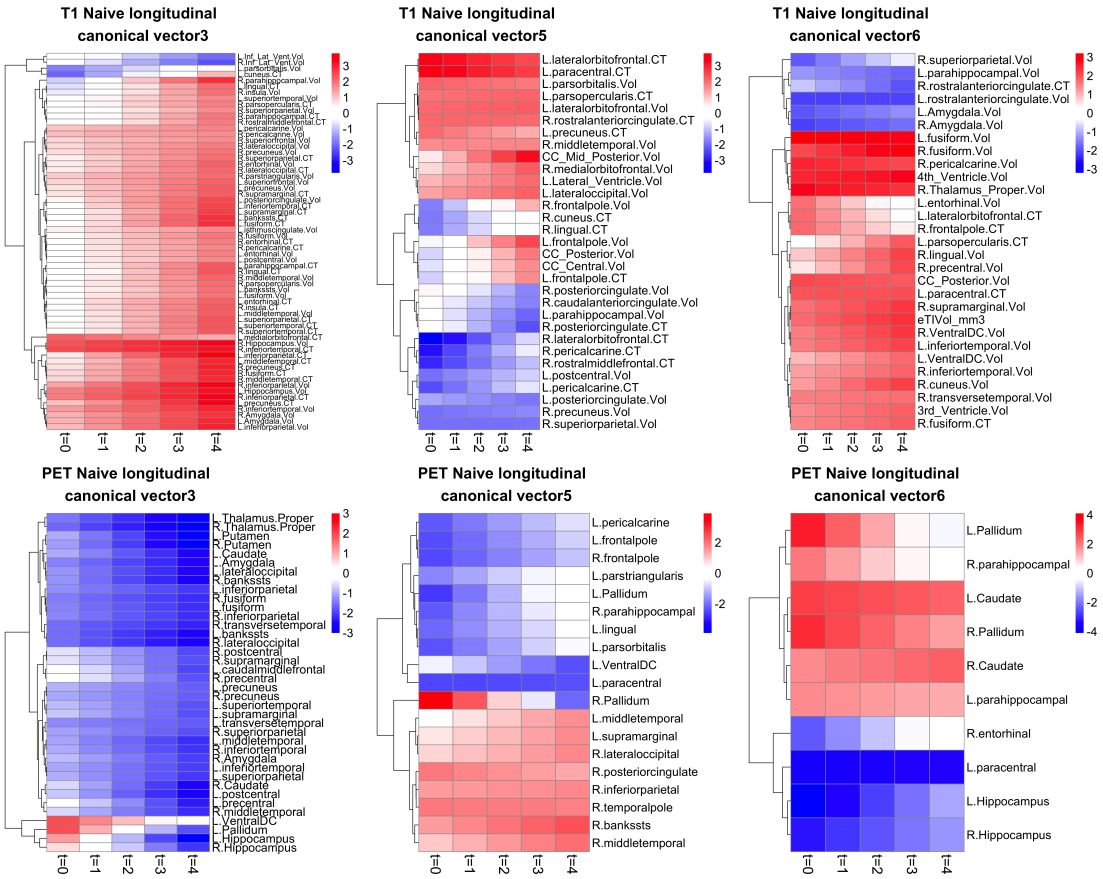


**图4.** 朴素：纵向典型向量的配对显示在时间=0、1、2、3、4。

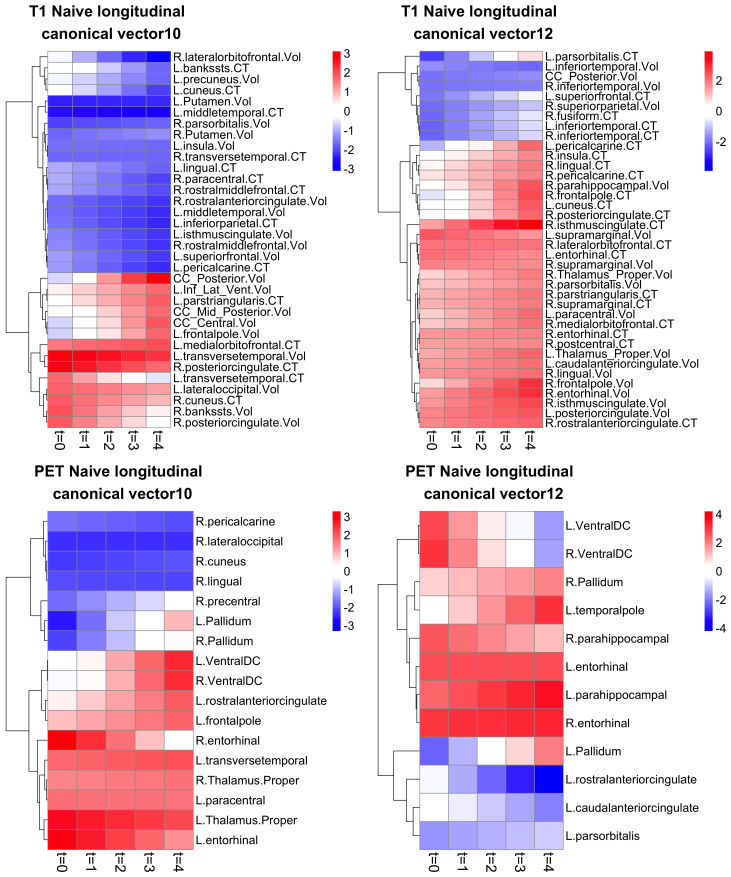


**图 5.** 朴素：展示了时间=0、1、2、3、4时的纵向典型向量对。

**表3** **.** 针对年龄、性别和教育程度进行调整后，诊断组比较及AD转化的方差分析

调整 后采用朴素方法。 *: 调整后的 p<0.05。

| | | T1 | | | | PET | | | |
|---|---|---|---|---|---|---|---|---|---|
| | CV | F | p | $\eta_p^2$ | 离群值 | F | p | $\eta_p^2$ | 离群值 |
| 组比较 | 1 | 0.02 | 0.984 | 0.000 | 3 | 3.23 | 0.040 | 0.010 | 4 |
| | 2 | 4.24 | 0.015 | 0.012 | 3 | 1.18 | 0.308 | 0.004 | 4 |
| | 3 | 47.83 | <0.001 * | 0.124 | 1 | 51.68 | <0.001 * | 0.133 | 0 |
| | 4 | 0.75 | 0.472 | 0.002 | 0 | 1.92 | 0.148 | 0.006 | 1 |
| | 5 | 8.74 | <0.001 * | 0.025 | 1 | 2.17 | 0.115 | 0.006 | 1 |
| | 6 | 2.03 | 0.132 | 0.006 | 2 | 8.47 | <0.001 * | 0.025 | 1 |
| | 7 | 0.24 | 0.787 | 0.001 | 1 | 1.19 | 0.304 | 0.004 | 3 |
| | 8 | 0.78 | 0.458 | 0.002 | 0 | 1.60 | 0.202 | 0.005 | 1 |
| | 9 | 0.53 | 0.589 | 0.002 | 0 | 1.71 | 0.183 | 0.005 | 2 |
| | 10 | 0.25 | 0.780 | 0.001 | 0 | 1.47 | 0.230 | 0.004 | 0 |
| | 11 | 1.00 | 0.367 | 0.003 | 0 | 1.71 | 0.182 | 0.005 | 2 |
| | 12 | 1.17 | 0.310 | 0.003 | 0 | 4.96 | 0.007 * | 0.015 | 0 |
| AD转换 | 1 | 0.64 | 0.424 | 0.001 | 2 | 1.89 | 0.170 | 0.003 | 3 |
| | 2 | 0.80 | 0.372 | 0.001 | 2 | 0.94 | 0.334 | 0.001 | 3 |
| | 3 | 248.28 | <0.001 * | 0.276 | 1 | 226.29 | <0.001 * | 0.258 | 0 |
| | 4 | 0.47 | 0.494 | 0.001 | 0 | 1.00 | 0.317 | 0.002 | 1 |
| | 5 | 0.23 | 0.635 | 0.000 | 1 | 0.03 | 0.874 | 0.000 | 1 |
| | 6 | 8.72 | 0.003 * | 0.013 | 2 | 13.11 | <0.001 * | 0.020 | 1 |
| | 7 | 0.38 | 0.541 | 0.001 | 1 | 0.81 | 0.368 | 0.001 | 3 |
| | 8 | 0.02 | 0.886 | 0.000 | 0 | 3.43 | 0.064 | 0.005 | 1 |
| | 9 | 0.09 | 0.769 | 0.000 | 0 | 1.16 | 0.282 | 0.002 | 2 |
| | 10 | 5.84 | 0.016 | 0.009 | 0 | 11.10 | 0.001 * | 0.017 | 0 |
| | 11 | 0.01 | 0.910 | 0.000 | 0 | 0.14 | 0.713 | 0.000 | 2 |
| | 12 | 0.31 | 0.578 | 0.001 | 0 | 1.21 | 0.271 | 0.002 | 0 |

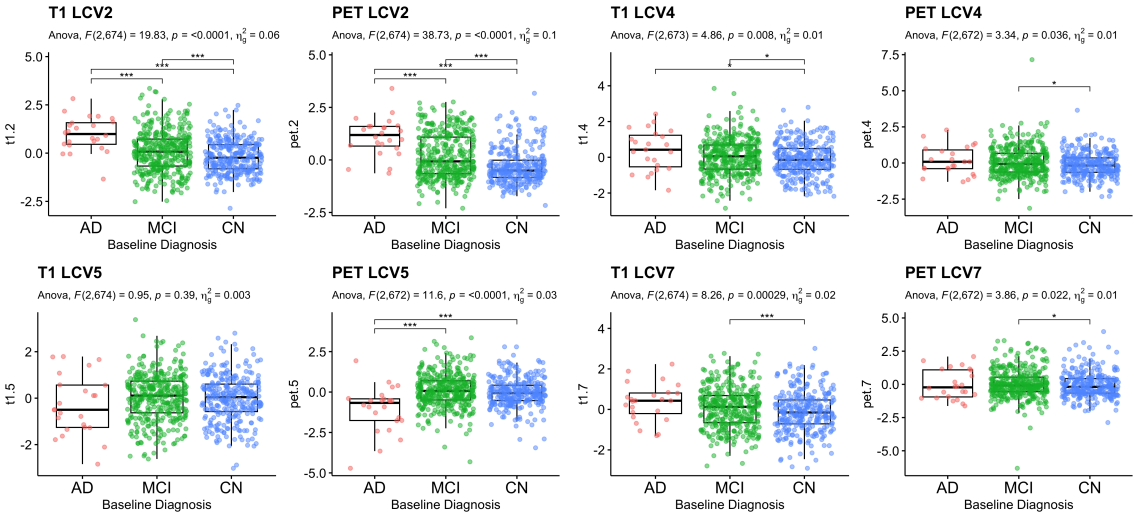


**图6.** LCCA：基线诊断状态比较。报告显著的典型变量。统计检验中已校正年龄、性别和教育程度。事后两两组间比较采用最小二乘均值差，并报告未校正的p值。

区域（右侧舌回、右侧距状皮层、右侧梭状回和下颞皮层）。与认知正常组相比，阿尔茨海默病和轻度认知障碍组的T1加权成像典型变量4显示右侧尾侧中额回体积增大，且右侧眶部呈纵向萎缩；而正电子发射断层显像未显示任何差异。T1加权成像典型变量5在不同基线临床状态之间未显示差异。阿尔茨海默病患者表现为左侧伏隔核区体积和右侧旁中央小叶厚度萎缩，左侧颞横回体积和右侧尾侧中额回体积较小，而左侧扣带回峡部和右侧外侧眶额回则较厚。阿尔茨海默病参与者的AV-45典型变量5显示左侧海马、右侧苍白球和左侧旁中央小叶纵向增加，而轻度认知障碍组和认知正常组参与者平均而言未呈现这一模式。典型变量7显示右侧舌回体积增大，但左侧尾侧前扣带皮层和眶部萎缩；淀粉样蛋白沉积则无显著差异。朴素方法识别出典型变量3和5存在显著的T1加权成像差异，以及典型变量9和12存在正电子发射断层显像差异。与LCCA不同，朴素方法未能识别出在T1加权成像和正电子发射断层显像中均显示诊断状态组间差异的典型变量。对于阿尔茨海默病转变，只有T1加权成像典型变量3和正电子发射断层显像典型变量9显示显著组间差异。

其次，在基线时非痴呆参与者中，我们比较了5年内转为阿尔茨海默病的参与者与仍未患痴呆的参与者之间的典型变量。656名非痴呆参与者中，有120名（18.3%）在5年内转为阿尔茨海默病。表2报告了F统计量、未校正的p值和经多重比较校正的p值。事后两两组间比较计算了最小二乘均值差。图 8 展示了抖动箱线图，以及多重比较校正后组间差异显著的典型变量的两两比较结果（未校正的p值）。T1加权成像的典型变量1、2和4，以及正电子发射断层显像的典型变量1、2、4和6，在转为阿尔茨海默病者与未转为阿尔茨海默病者之间存在显著差异。

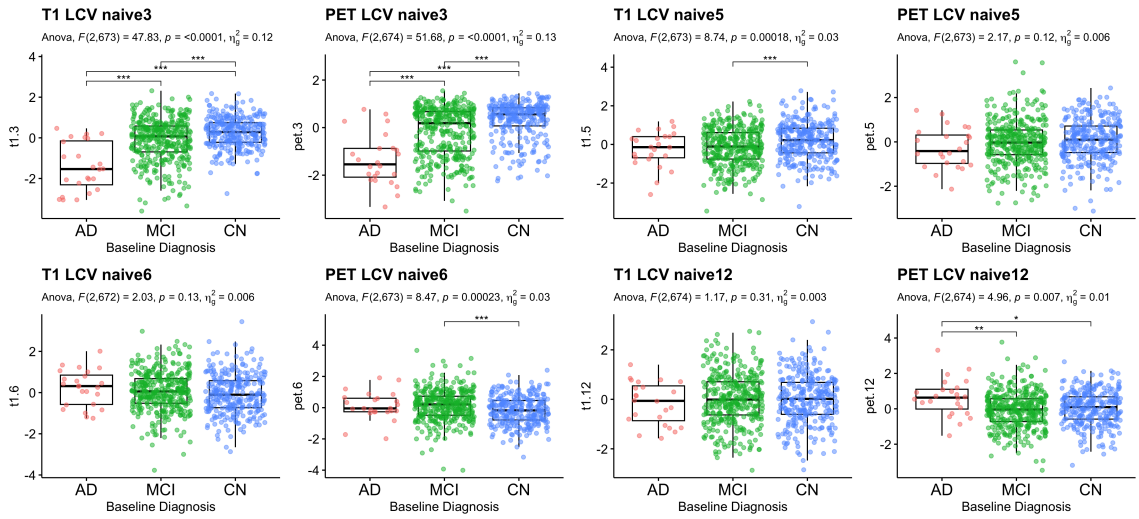


**图7**。 朴素方法：基线诊断状态比较。报告了显著的典型变量。统计检验中已校正年龄、性别和教育程度。对于事后两两组间比较，计算了最小二乘均值差，并报告了未校正的p值。

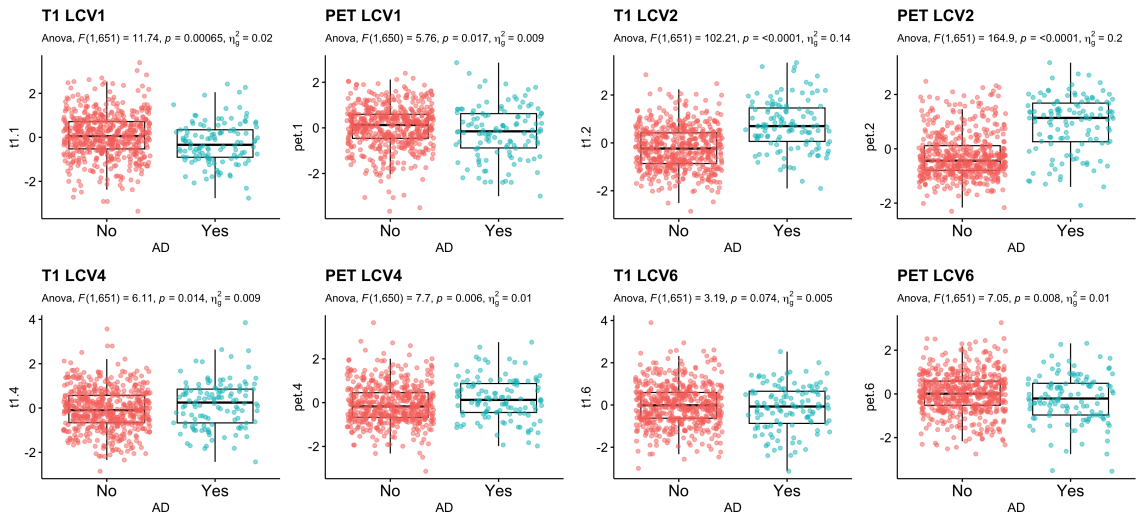


**图8**。 LCCA：比较5年内转为痴呆的参与者与未转为痴呆的参与者的典型变量。已校正年龄、性别和教育程度。

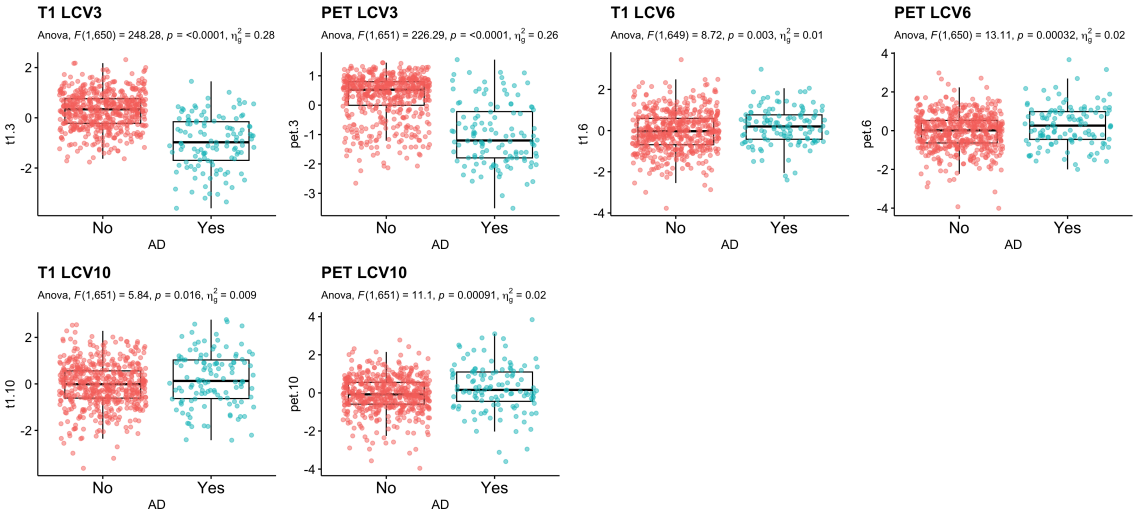


**图9.** 朴素方法：比较在5年内转为痴呆的参与者与未转为痴呆的参与者的典型变量。已对年龄、性别和教育程度进行调整。

## 4. 模拟

我们开展了大量数值实验，以评估LCCA相较于朴素方法的表现。第一种模拟设置旨在评估数据由LCCA模型生成时LCCA的整体表现。第二种模拟旨在考察数据在低信噪比条件下生成时的表现。我们还进行了第三种模拟，以确定在不同样本量和不同亚组不平衡程度下，LCCA恢复亚组的能力。

对于所有模拟设置，我们生成了100个独立数据集，并根据以下标准比较LCCA的表现：(1) 估计的典型变量数量；(2) 估计的典型相关系数的偏差；(3) 使用余弦相似度衡量估计典型载荷的准确性。对于第三种模拟，我们报告真实典型变量与估计典型变量之间的相关性，以替代标准(2)。由于真实典型变量的数量为1，即使LCCA或朴素方法选出的典型变量数量不同，我们仍比较第一典型变量的表现。

### *4.1.* 模拟设置*1*。

数据根据4 设置 $N_X$ 和 $N_Y$ 至 3 的模型生成。LPC载荷的生成方式如图10所示；我们设定 $p = 144$ 和 $q = 81$。LPC得分 $\xi_{ik}^X$ 和 $\xi_{ik}^Y$ 根据 $N(0, \lambda_k)$ 生成，$\lambda_1 = 8, \lambda_2 = 4, \lambda_3 = 2$，其中 $Cor(\xi_{i2}^X, \xi_{i2}^Y) = r$。值得注意的是，我们对LPC的第二对进行了相关性设置，以评估当相关信号解释的方差较小时LCCA的表现。对于每位受试者，访视次数由泊松分布(1)生成后加 3，使每位受试者平均有 4 个时间点。访视之间的时间间隔由

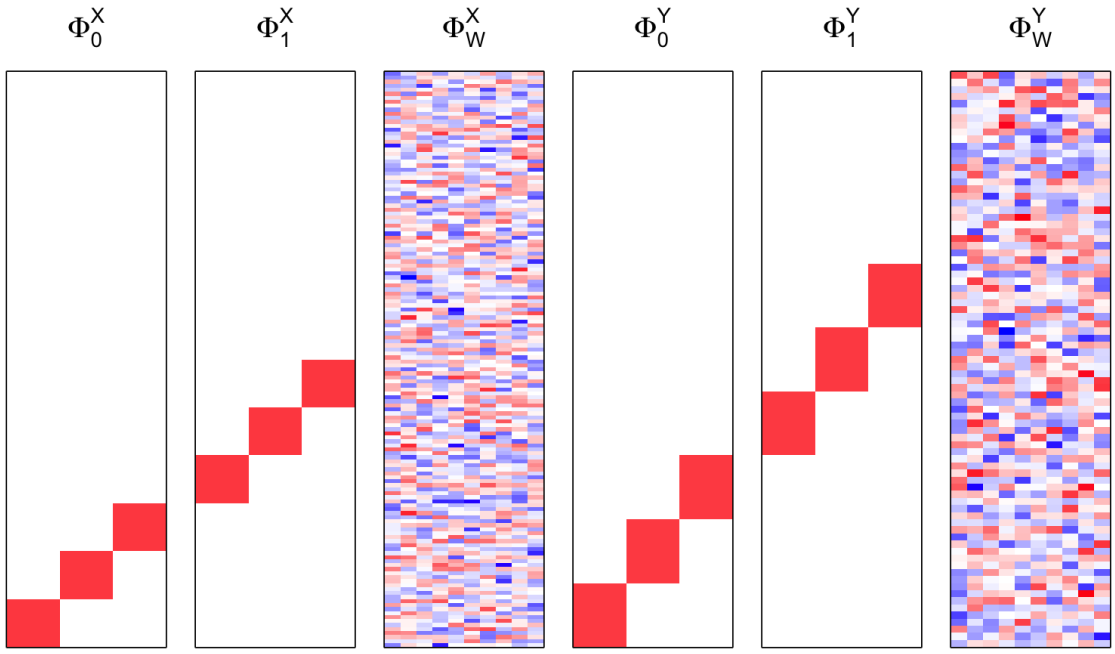


**图10**。 模拟1。用于模拟的LFPCA模型。

$U[1,2]$。图11 展示了某一模拟设置下前 10 位受试者的时间变量 $t_{ij}$和 $s_{ik}$示例。变量 $X$ 和 $Y$ 的时间点并不对齐，且访视次数各不相同。我们在 $r = 0.1$,0.3,0.5 下，取 $n = 100$,, 200、400，以及不同的LPC维数选择阈值（80%或90%），开展了模拟。每种情景均生成了 100 个独立数据集。

总体而言，在三个评价标准中，LCCA优于朴素方法。图12 显示，当样本量较小且真实典型相关较低时，LCCA和朴素方法均倾向于高估典型相关，而LCCA的表现优于朴素方法。结果表明，随着样本量增加以及潜在真实典型相关增大，基于近似的交叉验证维数估计的表现更好。与朴素方法相比，LCCA能够更准确地识别典型变量的数量。在余弦相似度测量中也发现了类似的模式。随着样本量增加以及潜在真实典型相关增大，估计的LCV载荷之间的余弦相似度表现更好。

*4.2.* 模拟设置*2*。

数据的生成方式与模拟1类似，但前五个LPC载荷取自第3节的ADNI数据（图 13）。LPC得分 $\xi_{ik}^X$ 和 $\xi_{ik}^Y$由 $N(0, \lambda_k)$、 $\lambda_1 = 45.4$、 $\lambda_2 = 17.2$、 $\lambda_3 = 7.0$、 $\lambda_4 = 4$、1, $\lambda_5 = 3.8$生成，其中 $Cor(\xi_{i2}^X, \xi_{i2}^Y) = r$。对于每位受试者，访视次数从泊松分布(1)生成，然后加上3，使每位受试者平均有4个时间点。该

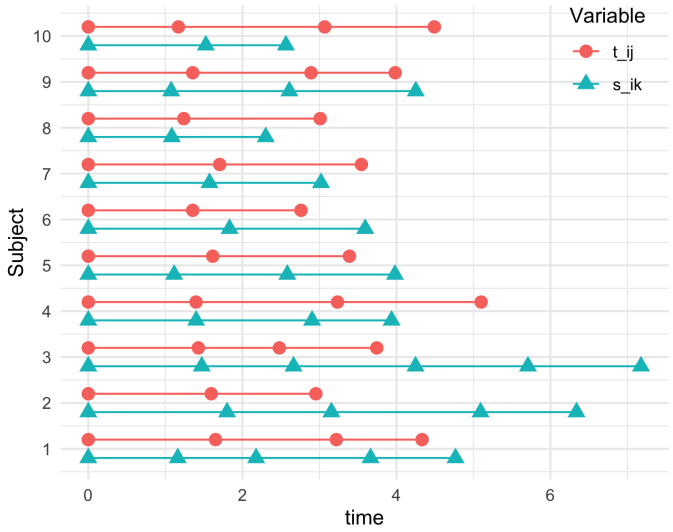


**图11.** 模拟设置中前10名受试者的时间变量 $t_{ij}$ 和 $s_{ik}$ 示例。 $X$ 和 $Y$ 变量之间的时间点未对齐，且访视次数不同。

访视之间的时间间隔由 $U[1,2]$生成。我们针对$r = 0.1$、$0.3$、$0.5$，在 $n = 100$、$200$、$400$ 的情况下，以及不同的LPC维数选择阈值（80%或90%）开展了模拟。每种情形均生成了100个独立数据集。我们发现结果与模拟1（图 14）非常相似，这表明我们在第3节中的数据分析结果在估计典型权重和相关性时偏倚较小。当阈值（$=0.8$）较小时，两种方法都倾向于低估典型相关；而当阈值较宽松时，这一问题则不会出现。

### *4.3.* 模拟设置*3*。

我们还考察了LCCA是否能够识别亚组。我们采用了与仿真2相同的模拟设置，但对纵向典型变量（LCV）进行了调整。我们没有从多元正态分布中生成LCV，而是设定了两个亚组。其中一组由 $MVN(0, \lambda \mathbf{I}_2)$生成，另一组由 $N(\mu \mathbf{1}_2, \mathbf{I}_2)$生成，如图15所示。我们将子样本比例从0.1变动到0.5，并评估LCCA能否识别亚组。

为评估LCCA在存在亚组时能否恢复真实纵向典型变量，我们计算了真实纵向典型变量与估计的纵向典型变量之间的相关性。图16 显示，即使样本量较小且亚组高度不平衡（子样本比例$=0.1$），LCCA仍能很好地识别纵向典型变量、纵向典型向量以及纵向典型变量的数量。

## 5. 结论

在本文中，我们提出了一种新的纵向典型相关分析方法，可以处理纵向数据中经常出现的不同时间采样和缺失值。

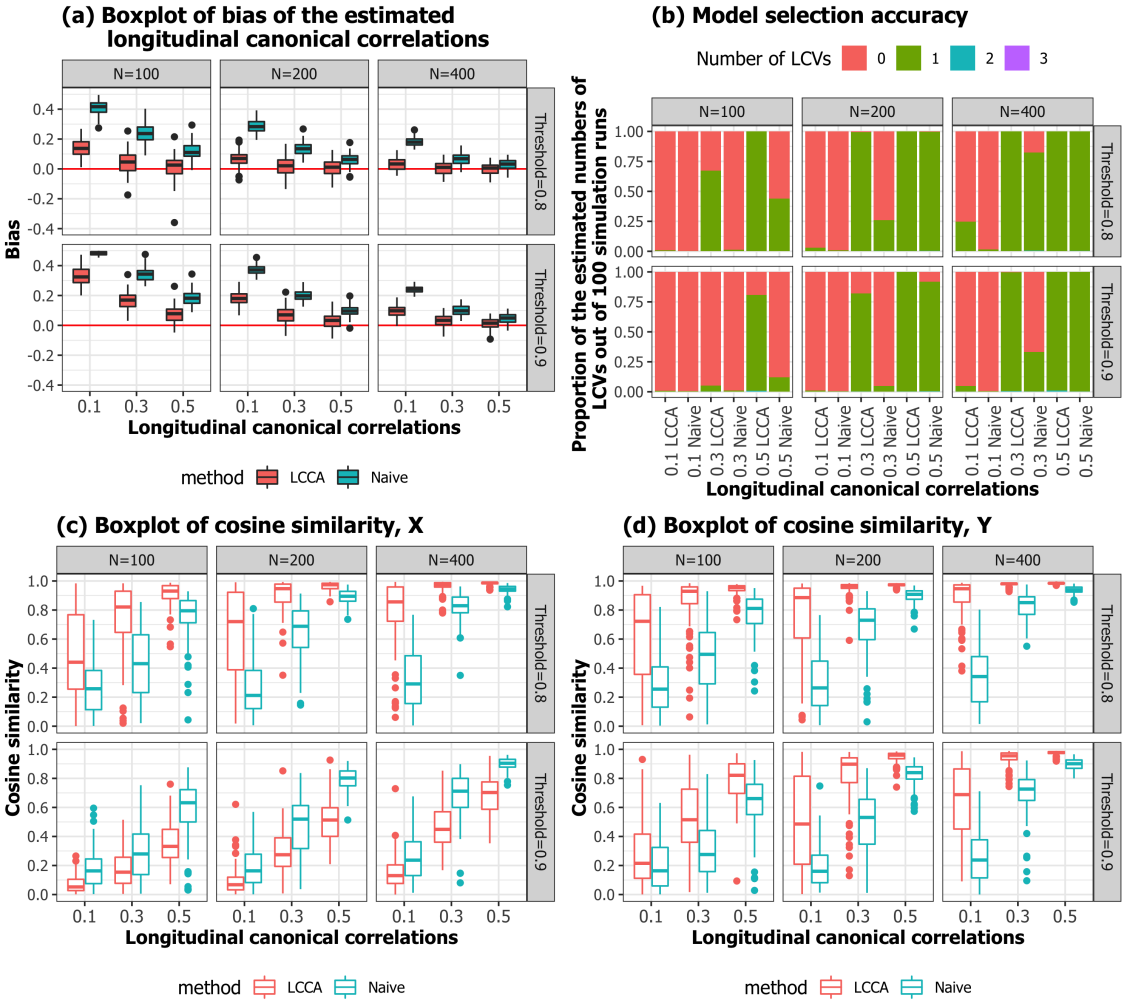
(a) Boxplot of bias of the estimated longitudinal canonical correlations
N=100
N=200
N=400
Threshold=0.8
Threshold=0.9
Bias
Longitudinal canonical correlations
method
LCCA
Naive
(b) Model selection accuracy
Number of LCVs
0
1
2
3
Proportion of the estimated numbers of LCVs out of 100 simulation runs
0.1 LCCA
0.1 Naive
0.3 LCCA
0.3 Naive
0.5 LCCA
0.5 Naive
Longitudinal canonical correlations
(c) Boxplot of cosine similarity, X
Cosine similarity
(d) Boxplot of cosine similarity, Y

**图12。** 模拟1。性能评估。

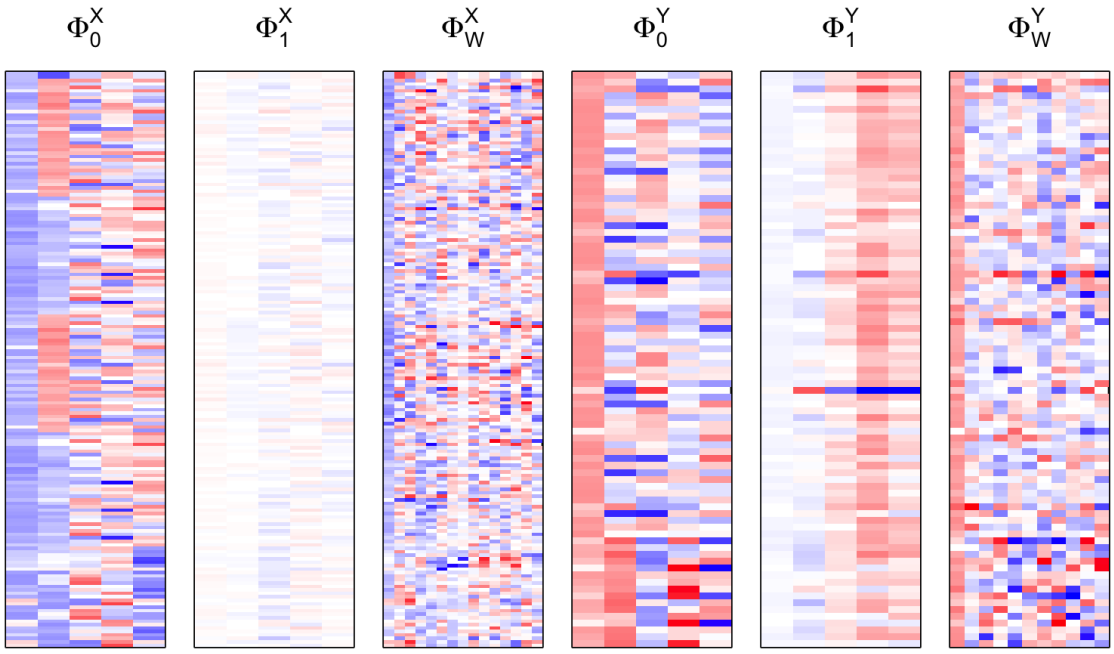


**图13**。 仿真2。用于仿真的LFPCA模型。

所提出的方法在处理线性和非线性轨迹方面具有很强的灵活性。应用于ADNI数据时，纵向典型相关分析揭示了淀粉样蛋白沉积相关神经元丢失与皮层厚度和皮层下体积之间最相关的模式。典型变量与基线临床状态相关，并可预测阿尔茨海默病转变。数值实验表明，LCCA的表现优于朴素方法；同时还显示，LCCA的性能对阈值的选择并不十分敏感，但需要足够的样本量或相关性的效应量。

我们的两阶段方法计算速度快，数值实验表明，按照经验法则使用80-90%的阈值时具有良好的性能。然而，使用LPCA进行降维可能会去除其他神经影像学模态（例如功能性磁共振成像（fMRI））中微小但重要的生物学信号。此外，与针对横断面高维数据的统一估计方法（Song 等人，2016）不同，两步法不会同时针对CCA优化降维和维度选择。尽管如此，我们新颖的LCCA方法提供了一种计算高效且性能良好的工具，可在神经影像学及其他产生高维、纵向、多变量数据的场景中发挥重要作用。

## 致谢

本研究得到美国国立卫生研究院 R01AG062578 项目资助（首席研究员：Lee）。

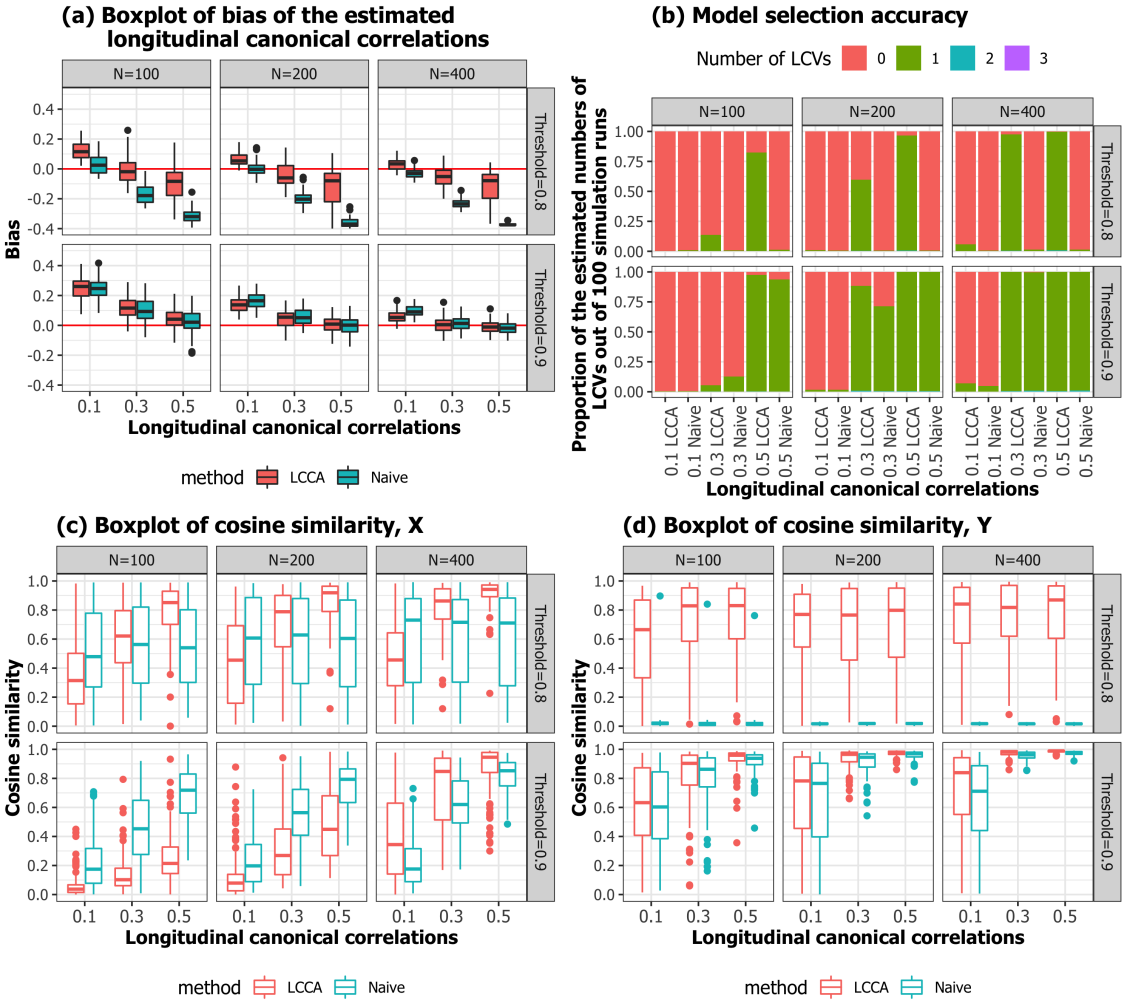
(a) Boxplot of bias of the estimated longitudinal canonical correlations
N=100
N=200
N=400
Threshold=0.8
Threshold=0.9
Bias
Longitudinal canonical correlations
method
LCCA
Naive
(b) Model selection accuracy
Number of LCVs
0
1
2
3
Proportion of the estimated numbers of LCVs out of 100 simulation runs
Longitudinal canonical correlations
(c) Boxplot of cosine similarity, X
Cosine similarity
(d) Boxplot of cosine similarity, Y

**图14。** 仿真2。性能评估。

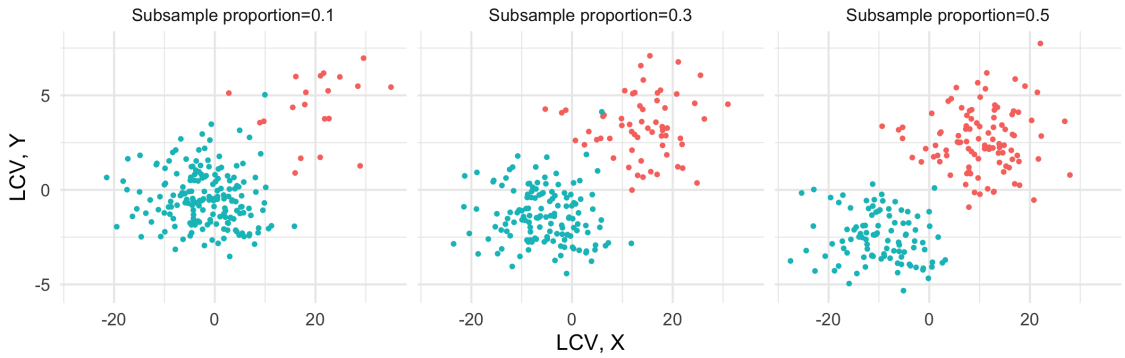


**图15.** 模拟3。不同亚组比例下纵向典型变量（LCV）的示例。

本项目的数据收集与共享由阿尔茨海默病神经影像学倡议（ADNI）（美国国立卫生研究院拨款 U01 AG024904）和国防部阿尔茨海默病神经影像学计划（美国国防部资助编号 W81XWH-12-2-0012）资助。ADNI由美国国家老龄研究所、美国国家生物医学影像与生物工程研究所，以及以下机构慷慨提供的捐款资助：艾伯维、阿尔茨海默病协会、阿尔茨海默病药物发现基金会、Araclon Biotech、Bio-Clinica, Inc.、渤健、百时美施贵宝公司、CereSpir, Inc.、Cogstate、卫材株式会社、Elan Pharmaceuticals, Inc.、礼来公司、EuroImmun、F. Hoffmann-La Roche Ltd及其关联公司基因泰克公司、富士瑞必欧、GE医疗、IXICO Ltd.、杨森阿尔茨海默病免疫疗法研发有限责任公司、Johnson & Johnson Pharmaceutical Research & Development LLC.、Lumosity、灵北、默克公司、Meso Scale Diagnostics, LLC.、NeuroRx Research、Neurotrack Technologies、诺华制药公司、辉瑞公司、Piramal Imaging、施维雅、Takeda Pharmaceutical Company和Transition Therapeutics。加拿大卫生研究院正提供资金，以支持ADNI在加拿大的临床研究中心。私营部门的捐款由美国国立卫生研究院基金会（www.fnih.org）协助管理。受资助机构为北加州研究与教育研究所；该研究由南加州大学阿尔茨海默病治疗研究所协调。ADNI数据由南加州大学神经影像实验室发布。

## 参考文献

Adhikari, B. M., Hong, L. E., Sampath, H., Chiappelli, J., Jahanshad, N., Thompson, P. M., Rowland, L. M., Calhoun, V. D., Du, X., Chen, S. 和 Kochunov, P.（2019）精神分裂症中的功能网络连接障碍和核心认知缺陷。《人脑图谱》。

Avants, B. B., Cook, P. A., Ungar, L., Gee, J. C. 和 Grossman, M.（2010）痴呆会导致白质完整性和皮层厚度相关性降低：一项多

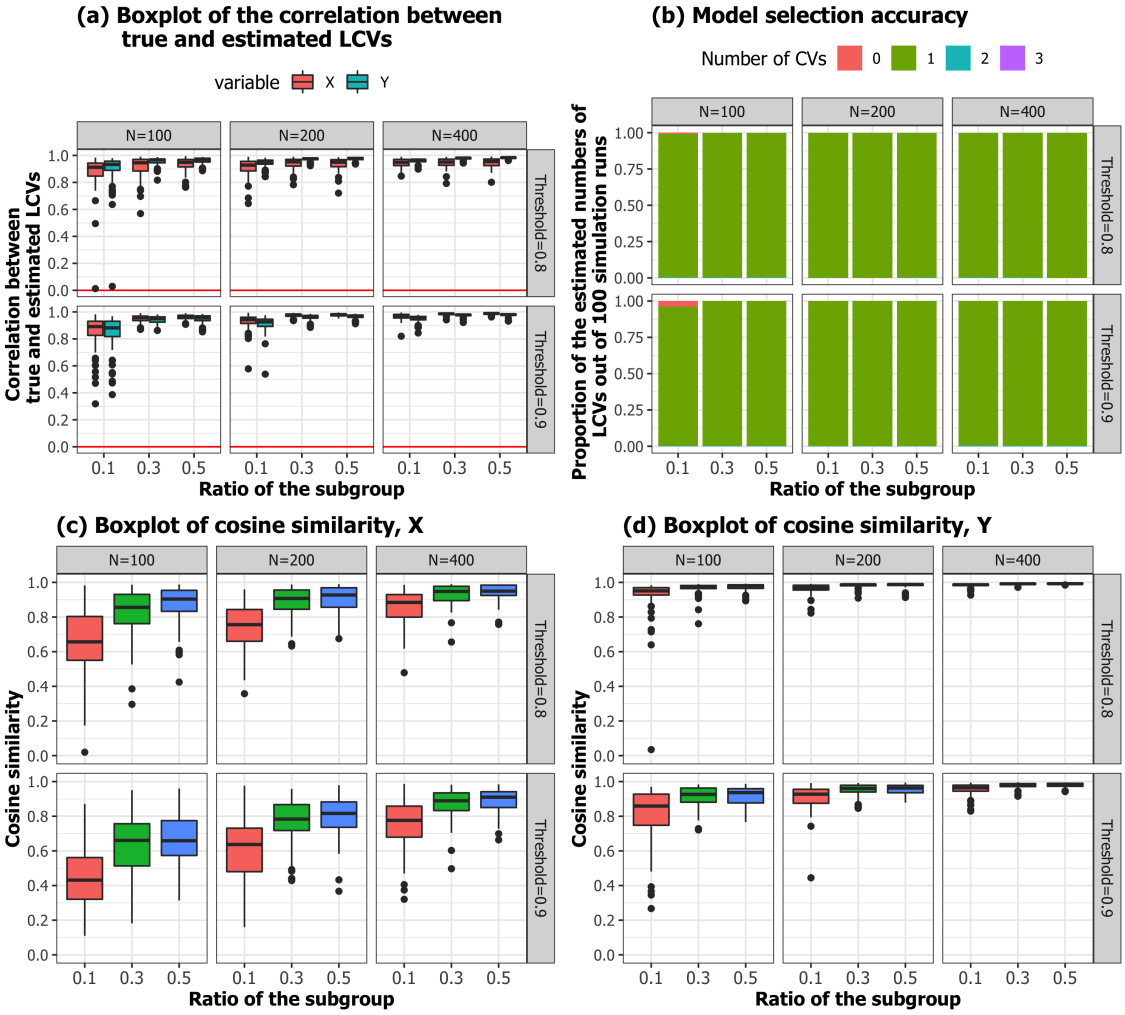
(a) Boxplot of the correlation between true and estimated LCVs
variable X Y
Correlation between true and estimated LCVs
N=100
N=200
N=400
Threshold=0.8
Threshold=0.9
Ratio of the subgroup
(b) Model selection accuracy
Number of CVs 0 1 2 3
Proportion of the estimated numbers of LCVs out of 100 simulation runs
(c) Boxplot of cosine similarity, X
Cosine similarity
(d) Boxplot of cosine similarity, Y

**图16.** 模拟3. 性能评估。

使用稀疏典型相关分析的神经影像学研究。*NeuroImage*，**50**，1004–1016。

Bao, Z., Hu, J., Pan, G. 和 Zhou, W.（2019）高维高斯向量的典型相关系数：有限秩情形。统计学年鉴，**47**，612–640。

Bartlett, M.（1947年）一般典型相关分布。数学统计学年鉴，1–17。

Benjamini, Y. 和 Hochberg, Y.（1995）控制假发现率：一种实用且有效的多重检验方法。英国皇家统计学会会刊：*B*辑（方法论），**57**，289–300。

Deleus, F. 和 Van Hulle, M. M.（2011年）基于正则化多集典型相关分析的功能性磁共振成像数据功能连接分析。神经科学方法杂志，**197**，143–157。

Desikan, R. S., S´egonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T. 等人（2006）一种自动标注系统，用于将磁共振成像扫描中的人类大脑皮层划分为基于脑回的感兴趣区域。*NeuroImage*, **31**, 968–980. Du, L., Liu, K., Zhu, L., Yao, X., Risacher, S. L., Guo, L., Saykin, A. J. 和 Shen, L.（2019）通过多任务稀疏典型相关分析识别渐进式影像遗传模式：ADNI队列的纵向研究。*Bioinformatics*, **35**, i474–i483. Fang, J., Lin, D., Schulz, S. C., Xu, Z., Calhoun, V. D. 和 Wang, Y.-P.（2016）用于检测差异影像遗传学模块的联合稀疏典型相关分析。*Bioinformatics (Oxford, England)*, **32**, 3480–3488. Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S. 等人（2002）全脑分割：人脑神经解剖结构的自动标注。*Neuron*, **33**, 341–355. Friederichs, P. 和 Hense, A.（2003）典型相关分析中的统计推断：以北大西洋海表温度对欧洲气候的影响为例。*Journal ofClimate*, **16**, 522–534. Gossmann, A., Zille, P., Calhoun, V. 和 Wang, Y.-P.（2018）假发现率校正的稀疏典型相关分析及其在影像基因组学中的应用。*IEEE*医学影像汇刊, **37**, 1761–1774. Greve, D. N., Salat, D. H., Bowen, S. L., Izquierdo-Garcia, D., Schultz, A. P., Catana, C., Becker, J. A., Svarer, C., Knudsen, G. M., Sperling, R. A. 和 Johnson, K. A.（2016）不同的部分容积校正方法会得出不同结论：一项关于衰老的18F-FDG-PET研究。*NeuroImage*, **132**, 334–343.

Greve, D. N., Svarer, C., Fisher, P. M., Feng, L., Hansen, A. E., Baare, W., Rosen, B., Fischl, B. 和 Knudsen, G. M.（2014）基于皮层表面的分析可降低脑PET数据动力学建模中的偏差和方差。*NeuroImage*, **92**, 225–236. Greven, S., Crainiceanu, C., Caffo, B. 和 Reich, D.（2011）纵向函数主成分分析。载于函数型数据分析及相关主题的最新进展, 149–154。Springer。Grosenick, L., Shi, T. C., Gunning, F. M., Dubin, M. J., Downar, J. 和 Liston, C.（2019）用于发现抑郁症中稳定亚型特异性回路机制的功能与光遗传学方法。*Biological Psychiatry. Cognitive Neuroscience andNeuroimaging*, **4**, 554–566. Hao, X., Li, C., Yan, J., Yao, X., Risacher, S. L., Saykin, A. J., Shen, L., Zhang, D. 和阿尔茨海默病神经影像学倡议（2017）通过时间约束组稀疏典型相关分析识别基因型与纵向表型之间的关联。*Bioinformatics (Oxford, England)*, **33**, i341–i349. Kang, J., Bowman, F. D., Mayberg, H. 和 Liu, H.（2016）通过多属性典型相关图发现的功能连接区域抑郁症网络。*NeuroImage*, **141**, 431–441. Kim, M., Won, J. H., Youn, J. 和 Park, H.（2019）基于联合连接的影像遗传学稀疏典型相关分析，用于检测帕金森病的生物标志物。*IEEE*医学影像汇刊。Lee, S., Zipunnikov, V., Reich, D. S. 和 Pham, D. L.（2015）纵向RAVENS图像的统计图像分析。*Frontiers in Neuroscience*, **9**, 368. Lin, D., Calhoun, V. D. 和 Wang, Y.-P.（2014）通过组稀疏典型相关分析建立功能性磁共振成像与SNP数据之间的对应关系。*Medical Image Analysis*, **18**, 891–902. Mihalik, A., Ferreira, F. S., Rosa, M. J., Moutoussis, M., Ziegler, G., Monteiro, J. M., Portugal, L., Adams, R. A., Romero-Garcia, R., V´ertes, P. E., Kitzbichler, M. G., V´aˇsa, F., Vaghi, M. M., Bullmore, E. T., Fonagy, P., Goodyer, I. M., Jones, P. B., Dolan, R. 和 Mour˜ao-Miranda, J.（2019）健康及临床抑郁的年轻人中的脑-行为协变模式。*Scientific Reports*, **9**. Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W., Trojanowski, J. Q., Toga, A. W. 和 Beckett, L.（2005）阿尔茨海默病神经影像学倡议。*Neuroimaging Clinics of North America*, **15**, 869–877. Rao, C. R., Rao, C. R., Statistiker, M., Rao, C. R. 和 Rao, C. R.（1973年）线性统计推断及其应用，第2卷。Wiley，纽约。Song, Y., Schreier, P. J., Ram´ırez, D. 和 Hasija, T.（2016）小样本支持下高维数据的典型相关分析。*Signal Processing*, **128**, 449–458.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Liu, E., Morris, J. C., Petersen, R. C., Saykin, A. J., Schmidt, M. E., Shaw, L., Siuciak, J. A., Soares, H., Toga, A. W., Trojanowski, J. Q. 和阿尔茨海默病神经影像学倡议（2012）阿尔茨海默病神经影像学倡议：成立以来发表论文综述。《阿尔茨海默病与痴呆》，**8**，S1–S68。Wilks, S.（1935年）论 k 组正态分布统计变量的独立性。《计量经济学》《计量经济学会杂志》，309–326。Wilms, I. 和 Croux, C.（2016）稳健稀疏典型相关分析。《*BMC* 系- 统生物学》，**10**，72。Witten, D. M., Tibshirani, R. 和 Hastie, T.（2009）惩罚矩阵分解及其在稀疏主成分和典型相关分析中的应用。《生物统计学》，**10**，515–534。Zhang, Y., Zhou, G., Jin, J., Wang, X. 和 Cichocki, A.（2014年）在基于稳态视觉诱发电位的脑机接口中使用多组典型相关分析进行频率识别。《国际神经系统杂志》，**24**，1450013。Zipunnikov, V., Greven, S., Shou, H., Caffo, B., Reich, D. S. 和 Crainiceanu, C.（2014年）纵向高维主成分分析及其在多发性硬化扩散张量成像中的应用。《应用统计年刊》，**8**，2175。