Advanced Statistical Modeling to Address Variability Arising From Complex Experimental Design

Danni Liu
Department of Epidemiology and Biostatistics
Joe C. Wen School of Population & Public Health
UC Irvine

Wen-pin Chen
Biostatistics Shared Resources, CFCCC

UCI

趙 Chao Family
Comprehensive Cancer Center

UCI Health

Date: 12.12.2025

List of Contents

The Hidden Sources of Variation in Your Experiments

An example: The researchers collected data from 10 mice (5 treated, 5 controls), each with 3 tissue sections. They processed samples from 5 treated mice in one batch and 5 controls in another batch the next day. Finally, they captured 100 spots*500 cells/spot in spatial transcriptomics (ST) for each sample. After analysis, they found 50,000 differentially expressed genes significant in all 10 mice AND all 3 tissue sections, but not consistently...

What are the sources of variation?
Anything went wrong?

The Hidden Sources of Variation in Your Experiments

An example: The researchers collected data from **10 mice (5 treated, 5 controls)**, each with 3 tissue sections. They processed samples from 5 treated mice in one batch and 5 controls in another batch the next day. Finally, they captured 100 spots*500 cells/spot in spatial transcriptomics (ST) for each sample. After analysis, they found 50,000 differentially expressed genes significant in all 10 mice AND all 3 tissue sections, but not consistently...

- **Phenotype**
- **Biological replicates**

The Hidden Sources of Variation in Your Experiments

An example: The researchers collected data from 10 mice (5 treated, 5 controls), each with **3 tissue sections**. They processed samples from 5 treated mice in one batch and 5 controls in another batch the next day. Finally, they captured 100 spots*500 cells/spot in spatial transcriptomics (ST) for each sample. After analysis, they found 50,000 differentially expressed genes significant in all 10 mice AND all 3 tissue sections, but not consistently...

- **Phenotype**
- **Biological replicates**

Technical replicates

The Hidden Sources of Variation in Your Experiments

An example: The researchers collected data from 10 mice (5 treated, 5 controls), each with 3 tissue sections. They processed samples from **5 treated mice in one batch and 5 controls in another batch the next day**. Finally, they captured 100 spots*500 cells/spot in spatial transcriptomics (ST) for each sample. After analysis, they found 50,000 differentially expressed genes significant in all 10 mice AND all 3 tissue sections, but not consistently...

- **Phenotype**
- **Biological replicates**

Technical replicates

Batch/technical variations

The Hidden Sources of Variation in Your Experiments

An example: The researchers collected data from 10 mice (5 treated, 5 controls), each with 3 tissue sections. They processed samples from 5 treated mice in one batch and 5 controls in another batch the next day. Finally, they captured **100 spots*500 cells/spot in spatial transcriptomics (ST)** for each sample. After analysis, they found 50,000 differentially expressed genes significant in all 10 mice AND all 3 tissue sections, but not consistently...

- **Phenotype**
- **Biological replicates**

Technical replicates

Spatial variations

Batch/technical variations

Repeated measures within each mouse

The Hierarchy of Variations

I. Phenotype: Disease vs control, treated vs untreated
(What we care about)

II. Samples/biological replicates:
(independent patients/mice 1,2,3…)

III. Tissue sections/technical replicates:
(same individual, measured multiple times)

IV. Spatial/cell variation:
(spots physically close, cells from same patient)

V. Technical variations:
(Batch, day of processing, technician, instruments, etc.)

The Hidden Sources of Variations in Your Experiments

An example: The researchers collected data from 10 mice (5 treated, 5 controls), each with 3 tissue sections. They processed samples from 5 treated mice in one batch and 5 controls in another batch the next day. Finally, they captured 100 spots*500 cells/spot in spatial transcriptomics (ST) for each sample. After analysis, they found 50,000 differentially expressed genes significant in all 10 mice AND all 3 tissue sections, but not consistently...

Anything went wrong?

The Hidden Sources of Variation in Your Experiments

An example: The researchers collected data from 10 mice (5 treated, 5 controls), each with 3 tissue sections. They processed samples from 5 treated mice in one batch and 5 controls in another batch the next day. Finally, they captured 100 spots*500 cells/spot in spatial transcriptomics (ST) for each sample. After analysis, they found 50,000 differentially expressed genes significant in all 10 mice AND all 3 tissue sections, but not consistently...

Anything went wrong?

- Bad experimental design: Confounded batch effects with biological signals;

The Hidden Sources of Variation in Your Experiments

An example: The researchers collected data from 10 mice (5 treated, 5 controls), each with 3 tissue sections. They processed samples from 5 treated mice in one batch and 5 controls in another batch the next day. Finally, they captured 100 spots*500 cells/spot in spatial transcriptomics (ST) for each sample. After analysis, they found 50,000 differentially expressed genes significant in all 10 mice AND all 3 tissue sections, but not consistently...

Anything went wrong?

- Bad experimental design: Confounded batch effects with biological signals;
- Statistical analysis: No statistical control for pseudo-replication problem

The Hidden Sources of Variation in Your Experiments

An example: The researchers collected data from 10 mice (5 treated, 5 controls), each with 3 tissue sections. They processed samples from 5 treated mice in one batch and 5 controls in another batch the next day. Finally, they captured 100 spots*500 cells/spot in spatial transcriptomics (ST) for each sample. After analysis, they found 50,000 differentially expressed genes significant in all 10 mice AND all 3 tissue sections, but not consistently...

Anything went wrong?

- Bad experimental design: Confounded batch effects with biological signals;
- Statistical analysis: No statistical control for **pseudoreplication** problem

Replicated tissue sections and cell measurements from the same mouse are autocorrelated, while they were treated as independent units in analysis.

→ 10 x 3 x 5000=150,000 data points, but only 10 TRUE biologically independent samples (mice)

What is Confounding?

Question: Does Disease affect gene X's expression?

Bad design (confounded):

- Disease samples → All in Batch 1
- Control samples → All in Batch 2

Results: Gene X has significant higher expression in Disease samples

But is this because of DISEASE or BATCH?

Can't tell! Batch and Disease are confounded.

What is Confounding?

Question: Does Disease affect gene X's expression?

Good design:

- Batch 1: 5 Disease + 5 Control
- Batch 2: 5 Disease + 5 Control
- Batch 3: 5 Disease + 5 Control

Now can separate Batch effect from Disease effect

Common Confounding Variables

Common confounding variables in omics:

- **Batch/Run:** Different sequencing runs, flow cytometry days, imaging dates
- **Operator:** Different technicians process samples
- **Processing:** Different protocols, instruments, reagent lots
- **Sample order:** First samples processed differently than last
- **Environmental:** Temperature, humidity, time of day
- **Biological covariates:** Age, sex, post-mortem interval (can't randomize but must measure!)

Complete vs Block Randomization

Complete randomization: All samples randomly assigned to batches

- Problem: Can randomly end up with all disease in batch 1 (by chance)
- Advantage: Simple, unbiased
- Use when: **Very large number of samples** (unlikely to accidentally confound)

Block randomization: Restrict randomization to ensure balanced allocation

- Approach: Divide samples into strata, randomize within strata
- Example: Within disease strata, randomly allocate to batches. Within control strata, randomly allocate.
- Result: Guaranteed balanced, no accidental confounding
- Advantage: Prevents confounding by design, not luck
- Use when: **Smaller studies**, multiple important variables

Complete vs Block Randomization

If for a small sample size study:

WRONG (complete randomization, unlucky):

- Batch 1: 10 Disease patients
- Batch 2: 10 Control patients
- Batch 3: 0 Disease, 0 Control

→ Can't separate batch from disease!

RIGHT (block randomization):

First, stratify by disease status:

- Disease strata: 10 patients
- Control strata: 10 patients

Then randomize within strata:

- Batch 1: 4 Disease + 3 Control = 7
- Batch 2: 3 Disease + 4 Control = 7
- Batch 3: 3 Disease + 3 Control = 6

→ Guaranteed balance, disease and batch independent

"experDesign" R package can do this automatically

Including Technical Controls

What are technical controls?

- Technical controls are standard samples processed in every batch (positive control)
- **Purpose:** Quantify batch-to-batch variation in standardized sample
- **Example:** Use same reference RNA in every 10X run

Implementation:

- Include 1-2 reference samples in every batch
- Reference should be biologically relevant (e.g., cell line from same tissue)
- Estimate batch effect from controls: How much does reference shift between batches?
- Validate that biological effects > batch effects

The Hidden Sources of Variation in Your Experiments

An example: The researchers collected data from 10 mice (5 treated, 5 controls), each with 3 tissue sections. They processed samples from 5 treated mice in one batch and 5 controls in another batch the next day. Finally, they captured 100 spots*500 cells/spot in spatial transcriptomics (ST) for each sample. After analysis, they found 50,000 differentially expressed genes significant in all 10 mice AND all 3 tissue sections, but not consistently...

Anything went wrong?

- Bad experimental design: Confounded batch effects with biological signals;
- Statistical analysis: No statistical control for **pseudoreplication** problem

Replicated tissue sections and cell measurements from the same mouse are autocorrelated, while they were treated as independent units in analysis.
→ 10 x 3 x 5000=150,000 data points, but only 10 TRUE biological samples (mice)

Pseudoreplication & Type I Error Inflation

Pseudoreplication: Samples from a single source were treated as separate experiments, while they share common genetic and environmental backgrounds.

→ 10 x 3 x 5000=150,000 data points, but only 10 TRUE biological samples (mice)

- Intra-individual correlation (r=0.3-0.5) is higher than inter-individual correlation (r=0.1-0.2). **(Figure 1)**

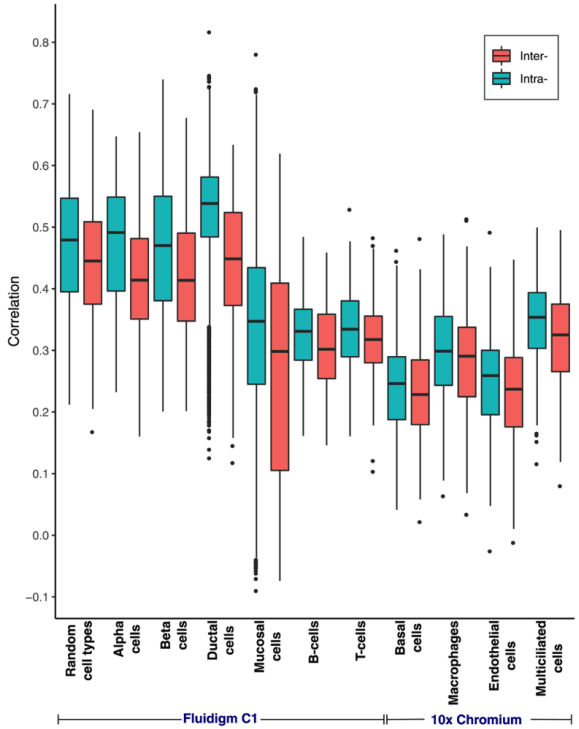


Figure 1. Inter- and intra-individual correlation for gene expression across 10 cell types

Pseudoreplication & Type I Error Inflation

Pseudoreplication: Samples from a single source were treated as separate experiments, while they share common genetic and environmental backgrounds.

→ 10 x 3 x 5000=150,000 data points, but only 10 TRUE biological samples (mice)

- Intra-individual correlation (r=0.3-0.5) is higher than inter-individual correlation (r=0.1-0.2).
- Spatial spots physically close have highly correlated expression. (**Figure 2**)

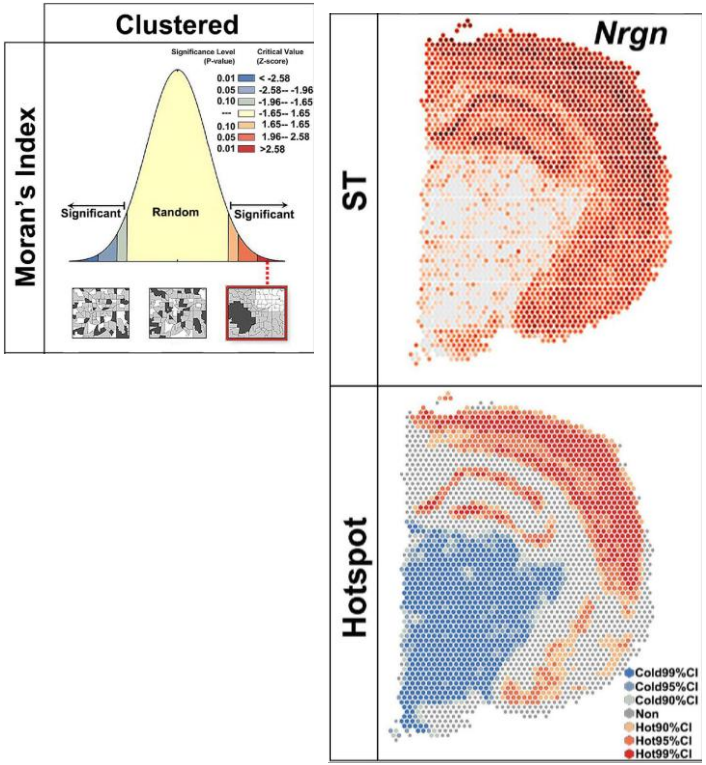


Figure 2. Global Moran's index to evaluate spatial autocorrelation patterns

Qiu et al. 2022

Pseudoreplication & Type I Error Inflation

Standard differential testing methods:

- t-test, DESeq2 (Wald test), edgeR (ExactTest)
- MAST (Likelihood-Ratio test) without random effects
- FindMarkers (Wilcoxon rank-sum test) in Seurat

Important assumption for these tests:

- Each observation is INDEPENDENT
  → They treat cells as independent observations. ✖

Problem:

- Underestimate true variance (didn't account for individual variation);
- Reject null hypothesis too often → many false positive DE genes;
- Type I error INFLATES! → FALSE high power (falsely detects too many significant genes)

Pseudoreplication & Type I Error Inflation

- Inflated false positive rate if individual variation was not accounted.

N_{ind}	N_{cells}	Two-part hurdle			Tweedie		GEE1	Pseudo-bulk		Tobit	Modified *t*
		Default	Corrected	RE	GLMM	GLM		Mean	Sum		
5	50	0.561	0.637	0.069	0.082	0.340	0.114	0.023	0.035	0.353	0.400
	100	0.677	0.719	0.064	0.084	0.463	0.110	0.022	0.032	0.471	0.510
	250	0.798	0.778	0.066	0.083	0.609	0.103	0.023	0.028	0.628	0.644
	500	0.862	0.803	0.065	0.081	0.705	0.104	0.023	0.026	0.725	0.718
10	50	0.563	0.611	0.055	0.064	0.350	0.076	0.024	0.021	0.345	0.397
	100	0.689	0.718	0.053	0.065	0.462	0.077	0.024	0.020	0.470	0.502
	250	0.810	0.793	0.049	0.064	0.610	0.074	0.022	0.019	0.624	0.635
	500	0.875	0.827	0.049	0.061	0.705	0.073	0.021	0.018	0.722	0.717
20	50	0.562	0.606	0.051	0.056	0.344	0.063	0.024	0.016	0.343	0.393
	100	0.687	0.705	0.048	0.056	0.459	0.064	0.024	0.014	0.466	0.503
	250	0.817	0.805	0.042	0.058	0.610	0.060	0.022	0.011	0.619	0.637
	500	0.884	0.844	0.042	0.055	0.705	0.062	0.021	0.010	0.720	0.716
30	50	0.563	0.604	0.053	0.054	0.341	0.058	0.025	0.013	0.344	0.395
	100	0.691	0.698	0.049	0.056	0.463	0.058	0.025	0.012	0.469	0.504
	250	0.818	0.803	0.044	0.055	0.608	0.057	0.022	0.010	0.624	0.636
	500	0.886	0.853	0.041	0.055	0.707	0.058	0.022	0.009	0.719	0.706
40	50	0.561	0.602	0.051	0.054	0.345	0.055	0.025	0.013	0.340	0.393
	100	0.689	0.699	0.049	0.053	0.455	0.055	0.026	0.012	0.467	0.502
	250	0.820	0.803	0.044	0.053	0.607	0.053	0.022	0.010	0.622	0.639
	500	0.888	0.856	0.042	0.053	0.704	0.054	0.022	0.008	0.721	0.713

Table 1. Type I error rate of different models
(Default: MAST without random effect; Corrected: batch corrected without random effect; RE: with random effect)

Zimmerman et al. 2021

Pseudoreplication & Type I Error Inflation

- When number of cells is low, batch correction without accounting for individual variation will further boost the Type I error rate.

N_{ind}	N_{cells}	Two-part hurdle			Tweedie		GEE1	Pseudo-bulk		Tobit	Modified *t*
		Default	Corrected	RE	GLMM	GLM		Mean	Sum		
5	50	0.561	0.637	0.069	0.082	0.340	0.114	0.023	0.035	0.353	0.400
	100	0.677	0.719	0.064	0.084	0.463	0.110	0.022	0.032	0.471	0.510
	250	0.798	0.778	0.066	0.083	0.609	0.103	0.023	0.028	0.628	0.644
	500	0.862	0.803	0.065	0.081	0.705	0.104	0.023	0.026	0.725	0.718
10	50	0.563	0.611	0.055	0.064	0.350	0.076	0.024	0.021	0.345	0.397
	100	0.689	0.718	0.053	0.065	0.462	0.077	0.024	0.020	0.470	0.502
	250	0.810	0.793	0.049	0.064	0.610	0.074	0.022	0.019	0.624	0.635
	500	0.875	0.827	0.049	0.061	0.705	0.073	0.021	0.018	0.722	0.717
20	50	0.562	0.606	0.051	0.056	0.344	0.063	0.024	0.016	0.343	0.393
	100	0.687	0.705	0.048	0.056	0.459	0.064	0.024	0.014	0.466	0.503
	250	0.817	0.805	0.042	0.058	0.610	0.060	0.022	0.011	0.619	0.637
	500	0.884	0.844	0.042	0.055	0.705	0.062	0.021	0.010	0.720	0.716
30	50	0.563	0.604	0.053	0.054	0.341	0.058	0.025	0.013	0.344	0.395
	100	0.691	0.698	0.049	0.056	0.463	0.058	0.025	0.012	0.469	0.504
	250	0.818	0.803	0.044	0.055	0.608	0.057	0.022	0.010	0.624	0.636
	500	0.886	0.853	0.041	0.055	0.707	0.058	0.022	0.009	0.719	0.706
40	50	0.561	0.602	0.051	0.054	0.345	0.055	0.025	0.013	0.340	0.393
	100	0.689	0.699	0.049	0.053	0.455	0.055	0.026	0.012	0.467	0.502
	250	0.820	0.803	0.044	0.053	0.607	0.053	0.022	0.010	0.622	0.639
	500	0.888	0.856	0.042	0.053	0.704	0.054	0.022	0.008	0.721	0.713

Table 1. Type I error rate of different models
(Default: MAST without random effect; Corrected: batch corrected without random effect; RE: with random effect)

Zimmerman et al. 2021

Solutions

- Modeling individual variation as random effects in a mixed effect model
- Aggregate data into individual level (need large sample size)

N_{ind}	N_{cells}	Two-part hurdle		Tweedie			GEE1	Pseudo-bulk		Tobit	Modified *t*
		Default	Corrected	RE	GLMM	GLM		Mean	Sum		
5	50	0.561	0.637	0.069	0.082	0.340	0.114	0.023	0.035	0.353	0.400
	100	0.677	0.719	0.064	0.084	0.463	0.110	0.022	0.032	0.471	0.510
	250	0.798	0.778	0.066	0.083	0.609	0.103	0.023	0.028	0.628	0.644
	500	0.862	0.803	0.065	0.081	0.705	0.104	0.023	0.026	0.725	0.718
10	50	0.563	0.611	0.055	0.064	0.350	0.076	0.024	0.021	0.345	0.397
	100	0.689	0.718	0.053	0.065	0.462	0.077	0.024	0.020	0.470	0.502
	250	0.810	0.793	0.049	0.064	0.610	0.074	0.022	0.019	0.624	0.635
	500	0.875	0.827	0.049	0.061	0.705	0.073	0.021	0.018	0.722	0.717
20	50	0.562	0.606	0.051	0.056	0.344	0.063	0.024	0.016	0.343	0.393
	100	0.687	0.705	0.048	0.056	0.459	0.064	0.024	0.014	0.466	0.503
	250	0.817	0.805	0.042	0.058	0.610	0.060	0.022	0.011	0.619	0.637
	500	0.884	0.844	0.042	0.055	0.705	0.062	0.021	0.010	0.720	0.716
30	50	0.563	0.604	0.053	0.054	0.341	0.058	0.025	0.013	0.344	0.395
	100	0.691	0.698	0.049	0.056	0.463	0.058	0.025	0.012	0.469	0.504
	250	0.818	0.803	0.044	0.055	0.608	0.057	0.022	0.010	0.624	0.636
	500	0.886	0.853	0.041	0.055	0.707	0.058	0.022	0.009	0.719	0.706
40	50	0.561	0.602	0.051	0.054	0.345	0.055	0.025	0.013	0.340	0.393
	100	0.689	0.699	0.049	0.053	0.455	0.055	0.026	0.012	0.467	0.502
	250	0.820	0.803	0.044	0.053	0.607	0.053	0.022	0.010	0.622	0.639
	500	0.888	0.856	0.042	0.053	0.704	0.054	0.022	0.008	0.721	0.713

Table 1. Type I error rate of different models
(Default: MAST without random effect; Corrected: batch corrected without random effect; RE: with random effect)

Zimmerman et al. 2021

Mixed Effect Models to Capture Hierarchy of Variation

Model = **Fixed Effects** + Random Effects + Error

- Phenotype (what we care about)
- Biological covariates (Age, gender, race, etc.)
- Known technical confounders (library size, post-mortem interval, etc.)

Mixed Effect Models to Capture Hierarchy of Variation

Model = Fixed Effects + **Random Effects** + Error

- Batch effect
- Patient and any nested variation
- Spatial correlation

Mixed Effect Models to Capture Hierarchy of Variation

Model = Fixed Effects + Random Effects + Error

(1|batch)

(1|patient)
(1|patient:tissue)

$$Gene = \underbrace{\beta \cdot Treatment}_{\text{Fixed effects}} + \underbrace{\gamma \cdot batch + b \cdot patient}_{\text{Random effects}} + S_{(x,y)} + \epsilon_{ij}$$

Fixed effects:
What you want to test

Random effects:
Batch/patient variation
(can also account for nested variation within patient)

Spatial structure

Random noise

Solve both confounding and pseudoreplication issue!

R Packages for Mixed Models

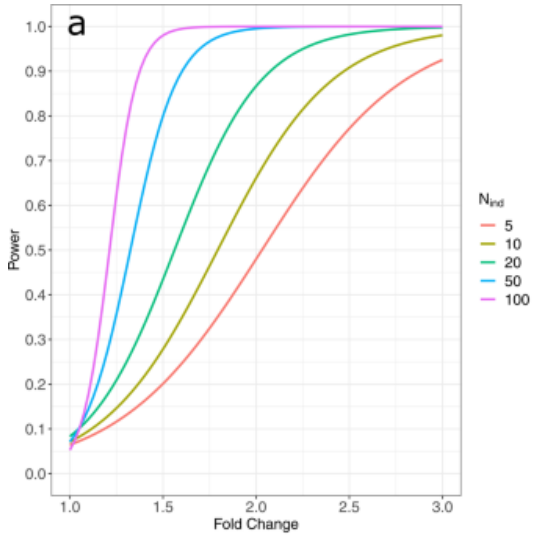
R packages:

- **lme4, glmmTMB** for general blocked designs
- **standR** for spatial blocked designs (GeoMx)
- **FLASHMM** for large datasets, fast speed
- **MAST** is flexible for complex experimental design
- **GLMES** also accounts for zero-inflation distribution.

Additional Note: Sample Size vs Cell Number

- Increasing sample size provides higher power boost than increasing cells.

N_{cell}=250, increase N_{ind} from 5-100



N_{ind} =20, increase N_{cell} per sample from 100-1000

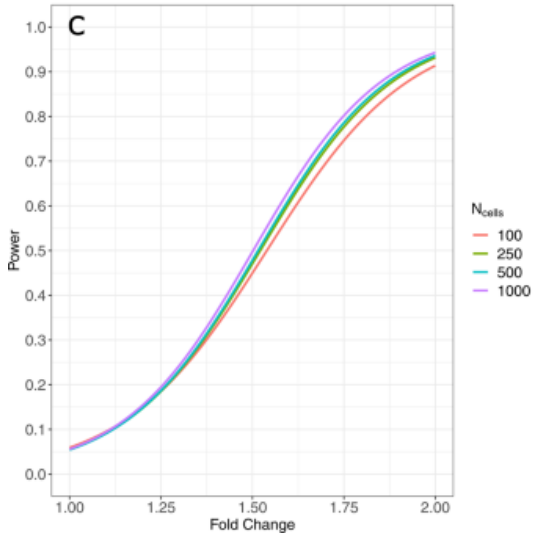


Figure 3. Power calculation for mixed models with different sample size or cell numbers.

When Mixed Models May NOT Be Needed

- **Scenario 1:** Very large number of individuals (N ≥ 100 individuals)
- **Scenario 2:** Very weak intra-class correlation (ICC<0.05)

$$ICC = \frac{\sigma^2_{between}}{\sigma^2_{between} + \sigma^2_{within}}$$

- **Scenario 3:** Aggregated Data Only
  - Averaging cells within individual first (pseudo-bulk)
  - Number of replicates = number of individuals

Key Take-Away

The problem:

- Confounding technical effects can mask the true biological signal.
- Cells and tissue sections from same patient are correlated. If you ignore this, you get 30-88% more false positives—that's a disaster.

The solution: Mixed models that include individual and batch as random effect. This resolves confounding effects, controls false positives caused by autocorrelation AND maintains power.

The outcome: Design experiments that prevent this problem, analyze data with appropriate statistical models, and report results that are actually reproducible.

Design matters more than analysis.

Fix problems at design time, not with statistics afterward.

References

- Zimmerman, K.D., Espeland, M.A. & Langefeld, C.D. A practical solution to pseudoreplication bias in single-cell studies. Nat Commun 12, 738 (2021). https://doi.org/10.1038/s41467-021-21038-1
- Qiu, Z. et al. Detection of differentially expressed genes in spatial transcriptomics data by spatial analysis of spatial transcriptomics: A novel method based on spatial statistics. Front. Neurosci. 16, 1086168. https://doi.org/10.3389/fnins.2022.1086168 (2022).
- Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M, Linsley PS, Gottardo R. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. 2015 Dec 10;16:278. doi: 10.1186/s13059-015-0844-5. PMID: 26653891; PMCID: PMC4676162.
- Ospina, O.E., Soupir, A.C., Manjarres-Betancur, R. et al. Differential gene expression analysis of spatial transcriptomic experiments using spatial mixed models. Sci Rep 14, 10967 (2024). https://doi.org/10.1038/s41598-024-61758-0
- Changjiang Xu, Delaram Pouyabahar, Veronique Voisin, Hamed Heydari, Gary D. Bader. FLASH-MM: fast and scalable single-cell differential expression analysis using linear mixed-effects models. bioRxiv 2025.04.08.647860; doi: https://doi.org/10.1101/2025.04.08.647860
- Wu CH, Zhou X, Chen M. Exploring and mitigating shortcomings in single-cell differential expression analysis with a new statistical paradigm. Genome Biol. 2025 Mar 17;26(1):58. doi: 10.1186/s13059-025-03525-6. PMID: 40098192; PMCID: PMC11912664.
- Ning Liu, Dharmesh D Bhuva, Ahmed Mohamed, Micah Bokelund, Arutha Kulasinghe, Chin Wee Tan, Melissa J Davis, standR: spatial transcriptomic analysis for GeoMx DSP data, Nucleic Acids Research, Volume 52, Issue 1, 11 January 2024, Page e2, https://doi.org/10.1093/nar/gkad1026
- Su, H., Wu, Y., Chen, B. et al. STANCE: a unified statistical model to detect cell-type-specific spatially variable genes in spatial transcriptomics. Nat Commun 16, 1793 (2025). https://doi.org/10.1038/s41467-025-57117-w

Biostatistics Shared Resource (BSR)

Wen-Pin Chen, MS
Senior Statistician, Manager of BSR

UCI

趙 Chao Family Comprehensive Cancer Center

UCI Health

BSR Services

- **Grant preparation/New Study Design/Power & Sample Calculation**
  - Statistical plan
  - Power and sample size calculation
  - Reviews for re-submission when addressing criticism of original submission
  - Budgets should be included for future statistical services
    - Percent effort for one or more named statisticians from the BSR, and
    - A budget allocation for statistical service to be paid on the basis of recharge.
  - Future statistical services: Data analysis, abstract, manuscript and presentation preparation
- **Data analysis and abstract, manuscript, and presentation preparation**
  - A recharge will be made for these services and standard hourly rates will apply.

BSR Service: Advanced Statistical Analysis

Omics Data Analysis

- Genomic (SNP, WGS, WES) data analysis (including GWAS, PheWAS)
- Transcriptomic (bulk/single cell RNA-seq) including eQTL
- Epigenetics (ChIP-seq; ATAC-seq)
- Single-cell multi-omics
- Functional (pathway, GO)
- Metabolomics

Research Computing

- HIPAA-compliant computational needs, cloud computing technologies
- Setup and run computationally intensive jobs on Cloud
- Programming assistance
- Database design, creation and management

Consulting

- Bioinformatics
- Database
- Machine learning
- Statistical genetics and genomics

BSR Faculty and Staff

- Min Zhang, M.D.,Ph.D., Director of BSR & Professor, Dept. of Epidemiology & Biostats
  minzhang@hs.uci.edu,
- Thomas Taylor, Ph.D., Principal Statistician, Dept. of Neurology & Dept. of Medicine
  thtaylor@hs.uci.edu, 949-824-6903
- Argyrios (Al) Ziogas, Ph.D., Associate Professor, Dept. of Medicine
  aziogas@hs.uci.edu, 949-824-1587
- Baolin Wu, Ph.D., Professor of Biostatistics, Dept. of Epidemiology & Biostats
  baolinw1@hs.uci.edu,
- Piyanuch Kongtim, M.D., Ph.D. HS Associate Clinical Professor, Dept. of Medicine
  pkongtim@hs.uci.edu ,714-456-5153

https://cancer.uci.edu/bsr

General inquiries: Ms. Wen-Pin Chen, MS (BSR Manager) wenpinc@hs.uci.edu

Training and Education

bigcare.uci.edu

Questions?

STOP!

Statistical Consultation - An iterative step/process

1. **Primary Endpoint**: What will be measured and how will it be quantified?
   - Tissue putrescine levels, Tumor volume – a continuous variable → Mean, Standard Deviation (SD)
   - A participant has experienced the event (Yes/No) → Proportion
   - Progression free survival, Overall Survival

2. **Primary Comparison**:
   a. Each participant to him/herself (e.g., baseline to off study)
   b. One study group to another (e.g., exposed to unexposed)
   c. Both of the above
   d. Values from the literature

Example:
Zell JA, Taylor TH, Albers CG, Carmichael JC, McLaren CE, Wenzel L, Stamos MJ. Phase IIa clinical biomarker trial of dietary arginine restriction and aspirin in colorectal cancer patients. Cancers, in press 28 March 2023. ID: cancers-2291650

The primary endpoint was percent reduction in rectal mucosa tissue putrescine levels from baseline to end of study.

Sample size calculations posited that percent reduction in putrescine would be 30% under the null and 60% under the alternative hypotheses, yielding a requirement for 20 subjects, using a 2-sided test (alpha = 0.05, power = 80%); 24 patients were enrolled to allow for at least 15% attrition.

Statistical Consultation - An iterative step/process

3. Source(s) of Data:
    a. Chart/record review: data exist already
    b. Prospective data gathering
        i. One-shot questionnaire
        ii. Participants followed over time

4. Inclusion/Exclusion Criteria: Person, place, time

5. Randomization (if applicable):
    Number of recruiting centers
    Ratio of assignments to conditions (e.g., 1 to 1, 1 to 2, ,,,)
    Expected rate of accrual at each center: block size
    Stratification variables (e.g., MGMT mutation status, …)

6. Selection of Controls per Case (if applicable):
    Variables on which to match (e.g., none, age, treatment history)
    Max time difference on important covariates (e.g., age, diagnosis, treatment history…)
    If there are multiple potential controls per case, plan for choosing

7. Justification of Sample Size:
    a. Constraints of resources, participants (pilot study)
    b. Desire to assert non inferiority (separate checklist)
    c. Desire to assert superiority
        i. Benchmark value (Ho: plausible range of values)
        ii. Smallest, clinically meaningful difference: plausible range of values)
        iii. Appropriate values for:
            1. Type-1 error
            2. Power
            3. Accrual time
            4. Follow-up time
    d. Estimation
            1. Half-width of confidence interval
            2. Appropriate interval (e.g., 95%, 80%,…)
            3. Value to be excluded (benchmark)

Based on the discussions of the above, a draft on sample sizes can be produced, the which may fuel further discussion. When there is general agreement on the design, including sample size and dependent measures, a draft statistical-analysis plan can be prepared.

Pseudoreplication & Type I Error Inflation

Pseudoreplication: Samples from a single source were treated as separate experiments, while they share common genetic and environmental backgrounds.

Intraclass Correlation Coefficient (ICC):

$$ICC = \frac{\sigma^2_{between}}{\sigma^2_{between} + \sigma^2_{within}}$$

Range: 0 (no correlation) to 1 (perfect correlation)

High ICC = significant fraction of total variance from individual differences