Easy Reproducible Longitudinal Modelling
in Clinical Research

Pau Esteve Bricullé

GR-BIO Retreat 2026

July 1, 2026

Project Introduction

Industrial PhD project.

Project Introduction

Industrial PhD project.

BCNpeptides + UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

It responds to the **company's need** for improved and personalized analysis of complex clinical trial data.

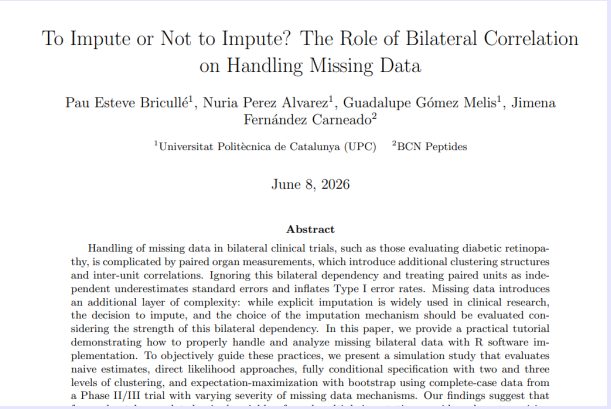
Conducted within the **clinical trials R&D environment** where:

Methods must align with business goals and regulatory expectations.

Project Status

Objective 2: Missing Data

Handling missing data and imputation for correlated higher tier cluster variables.



To Impute or Not to Impute? The Role of Bilateral Correlation on Handling Missing Data

Pau Esteve Bricullé[1], Nuria Perez Alvarez[1], Guadalupe Gómez Melis[1], Jimena Fernández Carneado[2]

[1]Universitat Politècnica de Catalunya (UPC) [2]BCN Peptides

June 8, 2026

Abstract

Handling of missing data in bilateral clinical trials, such as those evaluating diabetic retinopathy, is complicated by paired organ measurements, which introduce additional clustering structures and inter-unit correlations. Ignoring this bilateral dependency and treating paired units as independent underestimates standard errors and inflates Type I error rates. Missing data introduces an additional layer of complexity: while explicit imputation is widely used in clinical research, the decision to impute, and the choice of the imputation mechanism should be evaluated considering the strength of this bilateral dependency. In this paper, we provide a practical tutorial demonstrating how to properly handle and analyze missing bilateral data with R software implementation. To objectively guide these practices, we present a simulation study that evaluates naive estimates, direct likelihood approaches, fully conditional specification with two and three levels of clustering, and expectation-maximization with bootstrap using complete-case data from a Phase II/III trial with varying severity of missing data mechanisms. Our findings suggest that

Objective 1: Longitudinal modeling

Improving methods of diagnosis and prediction:

→ DR historical trial data focused.

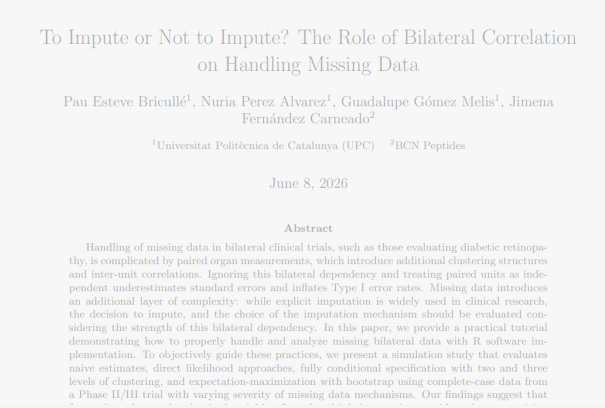
→ Tool for accessible and reproducible modeling.

Project Status

Objective 2: Missing data

Handling missing data and imputation for correlated higher tier cluster variables.

DISCARDED



To Impute or Not to Impute? The Role of Bilateral Correlation on Handling Missing Data

Pau Esteve Bricullé[1], Nuria Perez Alvarez[1], Guadalupe Gómez Melis[1], Jimena Fernández Carneado[2]

[1]Universitat Politècnica de Catalunya (UPC) [2]BCN Peptides

June 8, 2026

Abstract

Handling of missing data in bilateral clinical trials, such as those evaluating diabetic retinopathy, is complicated by paired organ measurements, which introduce additional clustering structures and inter-unit correlations. Ignoring this bilateral dependency and treating paired units as independent underestimates standard errors and inflates Type I error rates. Missing data introduces an additional layer of complexity: while explicit imputation is widely used in clinical research, the decision to impute, and the choice of the imputation mechanism should be evaluated considering the strength of this bilateral dependency. In this paper, we provide a practical tutorial demonstrating how to properly handle and analyze missing bilateral data with R software implementation. To objectively guide these practices, we present a simulation study that evaluates naive estimates, direct likelihood approaches, fully conditional specification with two and three levels of clustering, and expectation-maximization with bootstrap using complete-case data from a Phase II/III trial with varying severity of missing data mechanisms. Our findings suggest that

Objective 1: Longitudinal modeling

Improving methods of diagnosis and prediction:

→ DR historical trial data focused.

→ Tool for accessible and reproducible modeling.

Why?

Part of Industrial Doctorate (DI) Goals

Specified as a deliverable in the GenCat DI research plan.

Helping Clinical Teams

Removes the bottleneck of requiring R programming.

Data Reshaping

Built-in tool to pivot wide-format trial data into tidy long-format for analytical readiness.

Automated Fit

Fits LMM and GLMM outcomes with automated covariate screening via AIC/BIC selection.

Post-Hoc Power

Interaction term power for potential sample size calculations.

Why?

Part of Industrial Doctorate (DI) Goals

Specified as a deliverable in the GenCat DI research plan.

Helping Clinical Teams

Removes the bottleneck of requiring R programming.

Data Reshaping

Built-in tool to pivot wide-format trial data into tidy long-format for analytical readiness.

Automated Fit

Fits LMM and GLMM outcomes with automated covariate screening via AIC/BIC selection.

Post-Hoc Power

Interaction term power for potential sample size calculations.

Shiny app structure

1. Upload

Excel/CSV long-format data.

2. Mapping

ID, Time, Response mapping.

3. Fit

Automated model selection.

4. Post-Hoc

EMMs & Contrasts.

5. Export

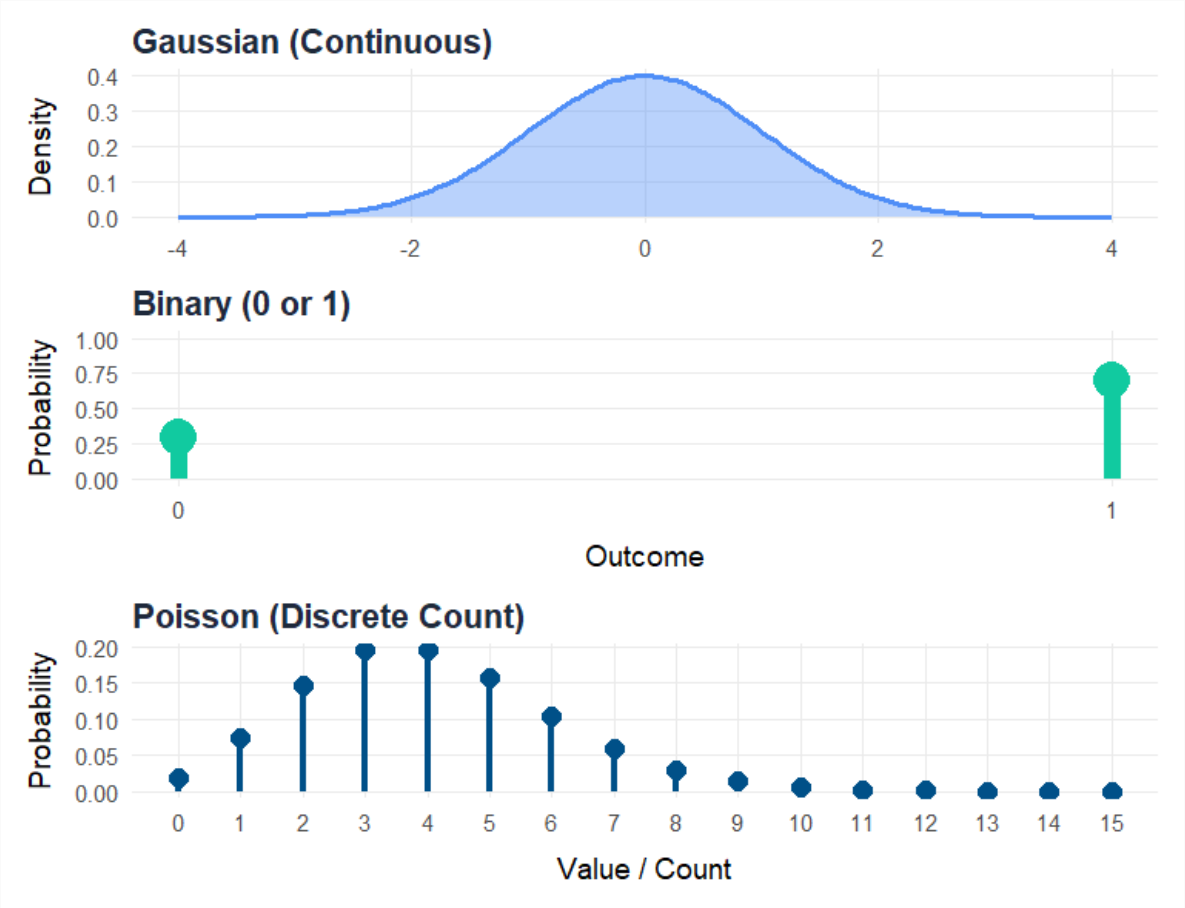
Complete Excel results.

Logic

Response Branching

Selection based on out-come distribution:

- `Gaussian`: Linear Mixed Models.
- `Binomial`: Logit-link GLMMs.
- `Count`: Poisson/NB fits.

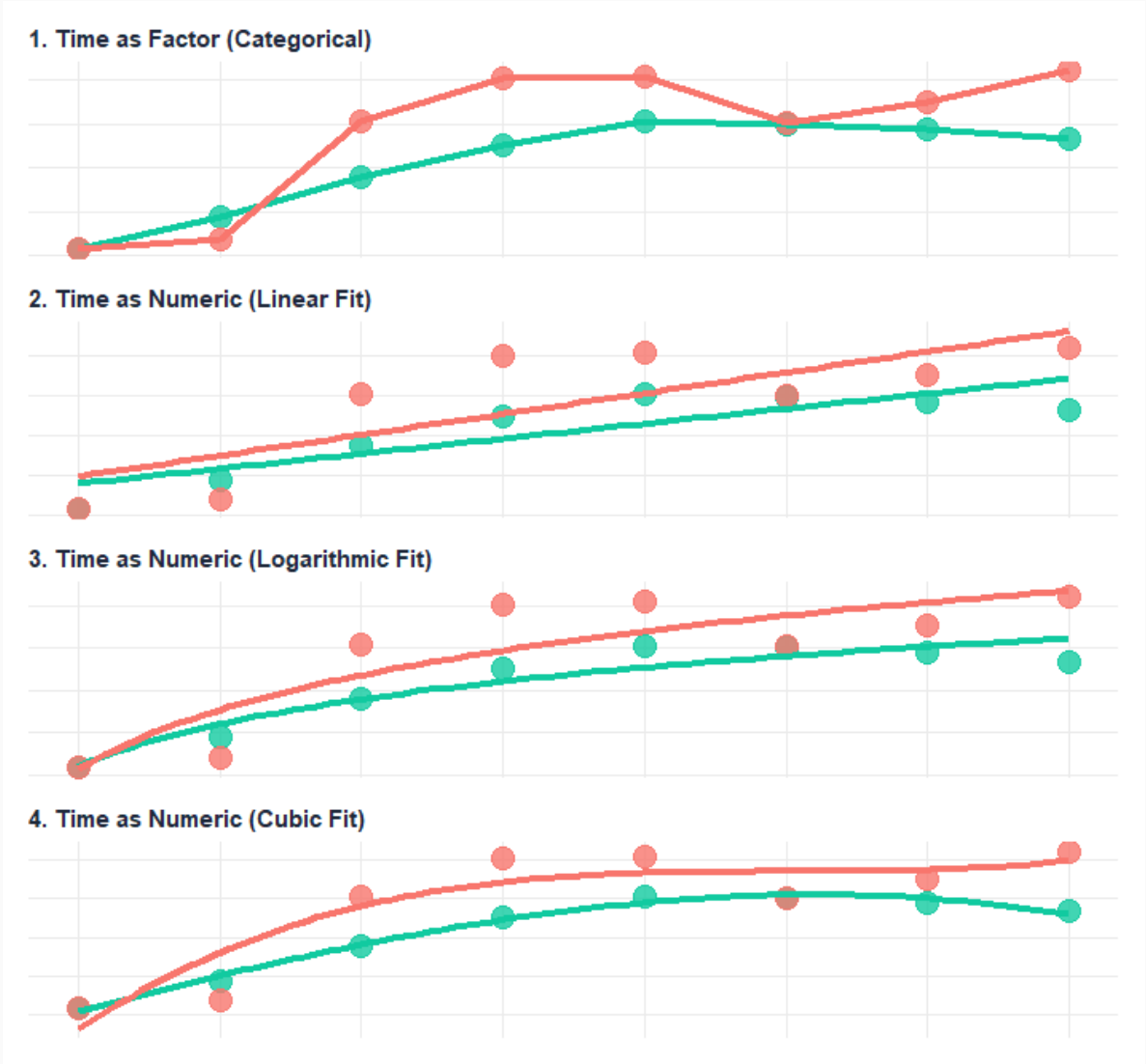


Logic

Temporal Control

Control over time-course and initial states:

- `Factor vs. Numeric`: Categorical visits or continuous growth curves.
- `Baseline Adjustment`: Correction for starting imbalances.



Time as Factor vs. Time as Numeric

Numeric Mode (Linear Assumption): 4 Parameters

$$Y = \beta_0 + \beta_1 \text{Treat} + \beta_2 \text{Time} + \beta_3 (\text{Treat} \times \text{Time})$$

Factor Mode (Categorical): $2J$ Parameters (J visits)

$$Y = \beta_0 + \beta_1 \text{Treat} + \sum_{j=2}^{J} \gamma_j I(\text{Time} = j) + \sum_{j=2}^{J} \delta_j (\text{Treat} \times I(\text{Time} = j))$$

	Factor	**Numeric**
Parameters mean	$2J$	4
Trajectory shape	Unconstrained	Constrained to assumed form
Estimand	Group diff at visit J	Slope difference β_3
Bias risk	Minimal	High if wrong functional form
df cost	High	Low
Irregular visit handling	Natural	Form-dependent

Time as Factor vs. Time as Numeric

Numeric Mode (Linear Assumption): 4 Parameters

$$Y = \beta_0 + \beta_1 \text{Treat} + \beta_2 \text{Time} + \beta_3 (\text{Treat} \times \text{Time})$$

Factor Mode (Categorical): $2J$ Parameters (J visits)

$$Y = \beta_0 + \beta_1 \text{Treat} + \sum_{j=2}^{J} \gamma_j I(\text{Time} = j) + \sum_{j=2}^{J} \delta_j (\text{Treat} \times I(\text{Time} = j))$$

	Factor	**Numeric**
Parameters mean	$2J$	4
Trajectory shape	Unconstrained	Constrained to assumed form
Estimand	Group diff at visit J	Slope difference β_3
Bias risk	**Minimal**	**High if wrong functional form**
df cost	High	Low
Irregular visit handling	Natural	Form-dependent

Other Settings

Other Functions and Settings

Added Covariates and validity checks:

- **AIC/BIC Automated Selection**:
  - → **Correlation**: Comparison of AR1, CS, and Unstructured.
  - → **Variable**: Automated screening of covariates.
- **Missing Data**: Classification of intermittent vs. monotone patterns.
- **Random Slope**: Inclusion of random time effects per cluster.
- **Residual plots**: Diagnostic checks for model assumptions.
- **Power**: Calculation of statistical power to detect group differences over time.

Other Settings

Other Functions and Settings

Added Covariates and validity checks:

- **AIC/BIC Automated Selection**:
  - → **Correlation**: Comparison of AR1, CS, and Unstructured.
  - → **Variable**: Automated screening of covariates.
- **Missing Data**: Classification of intermittent vs. monotone patterns.
- **Random Slope**: Inclusion of random time effects per cluster.
- **Residual plots**: Diagnostic checks for model assumptions.
- **Power**: Calculation of statistical power to detect group differences over time.

Power Calculations for Correlated Data

Classical tests

Power depends on:

- Effect size N α residual variance

Closed-form formulas. One N to find.

Power Calculations for Correlated Data

Classical tests

Power depends on:

- Effect size N α residual variance

Closed-form formulas. One N to find.

Longitudinal & Clustered Data

All of the previous, plus:

- # clusters, cluster size, σ_u^2, σ_e^2, ICC, random slope variances

No single N: power differs per parameter.

Correlated observations

ICC reduces effective N below nominal count

Approximate df

Such as Satterthwaite (normal fails with few clusters)

Random slopes

Extra variance components assumed from pilot data

Power Calculations for Correlated Data

Classical tests

Power depends on:

- Effect size N α residual variance

Closed-form formulas. One N to find.

Longitudinal & Clustered Data

All of the previous, plus:

- # clusters, cluster size, σ_u^2, σ_e^2, ICC, random slope variances

No single N: power differs per parameter.

Correlated observations

ICC reduces effective N below nominal count

Approximate df

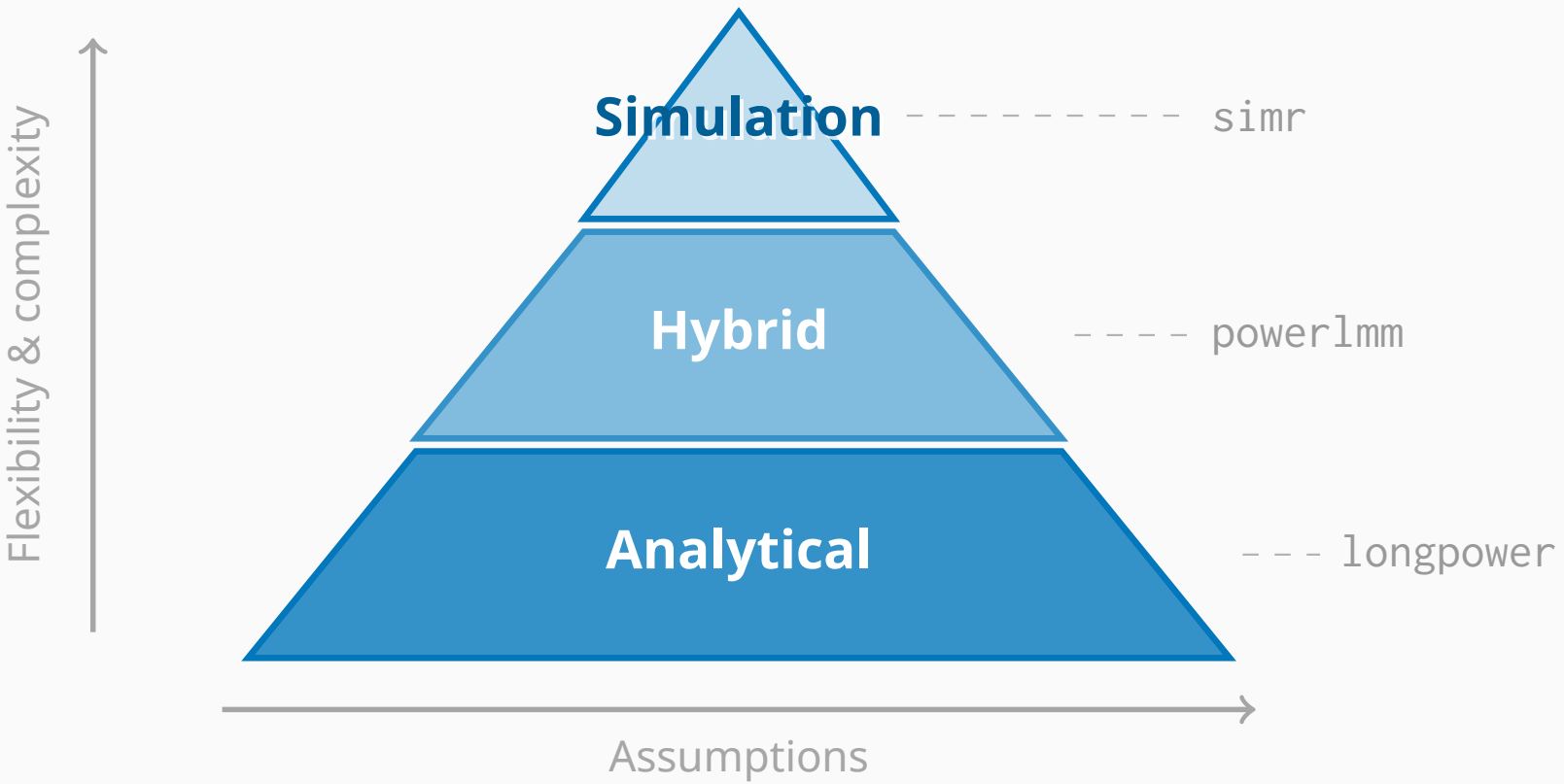
Such as Satterthwaite (normal fails with few clusters)

Random slopes

Extra variance components assumed from pilot data

$\rightarrow$ Reliable power often requires simulation

Power Calculations with Random Effects



Simulation Approach

Generative Model

Baseline Model ($\hat{M}$):

$$g(\mu_{ij}) = \mathbf{x}_{ij}^{\top}\boldsymbol{\beta} + \mathbf{z}_{ij}^{\top}\mathbf{u}_i$$

Random effects:

$$\mathbf{u}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

Simulated responses:

$$\tilde{y}_{ij} \sim \mathcal{F}\left(\mu_{ij}, \hat{\phi}\right)$$

$g(\cdot)$: link $\mathcal{F}$: distribution
$\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\Sigma}}$, $\hat{\phi}$ from pilot data

Monte Carlo Loop

Repeated B times:

1. **Simulate** $\tilde{y}^{(b)}$ from $\hat{M}$
2. **Refit** $\hat{M}^{(b)}$ to $\tilde{y}^{(b)}$
3. **Test** H_0: record result

Estimated power:

$$\hat{\pi} = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left[p^{(b)} < \alpha\right]$$

Wilson 95% CI reported around $\hat{\pi}$

Analytical Power Calculations

$$Z_P = \sqrt{\frac{mJs_t^2\Delta^2}{2\sigma^2(1-\rho)}} - Z_{1-\alpha/2}$$ **Diggle (2002)**

$$Z_P = \sqrt{\frac{N\pi_A\pi_B\beta_1^2(\mathbf{1}^\top R^{-1}\mathbf{1})}{\sigma^2}} - Z_{1-\alpha/2}$$ **Liu & Liang (1997)**

$$Z_P = \sqrt{\frac{n_{A1}\delta_j^2}{(\psi_A+\lambda\psi_B)\sigma_j^2}} - Z_{1-\alpha/2}$$ **Lu, Luo & Chen (2008)**

Final Statistical Power = $\Phi(Z_P)$

Comparison of Analytical Power Approaches

Framework	Strengths	Limitations
Diggle (2002) *Linear Mixed Model*	✓ Compares overall linear slope (Δ)	× Continuous outcomes only (σ^2) × Strict Compound Symmetry (ρ) × Assumes complete follow-up (*mJ*)
Liu & Liang (1997) *GEE Framework*	✓ Multiple outcome types ✓ Flexible correlation matrix (*R*) ✓ Unequal group allocation ($\pi_A \neq \pi_B$)	× Marginal (population-averaged) model × Less robust to dropout
Lu, Luo & Chen (2008) *Mixed Model*	✓ Unconstrained trajectory (categorical time) ✓ Handles MAR dropout (ψ) ✓ Retains partial data (Higher Power) ✓ Optimal allocation for attrition (λ)	× Continuous outcomes only (σ_J^2) × No time as continuous flexibility (δ_J)

Backend vs Frontend

Backend

→ Statistical computing

→ Data manipulation

→ Model fitting & power calculations

→ Runs server-side

Frontend

→ Structure & layout (HTML)

→ Visual styling (CSS)

→ Interactivity & logic (JS)

Backend vs Frontend

Backend

→ Statistical computing

→ Data manipulation

→ Model fitting & power calculations

→ Runs server-side

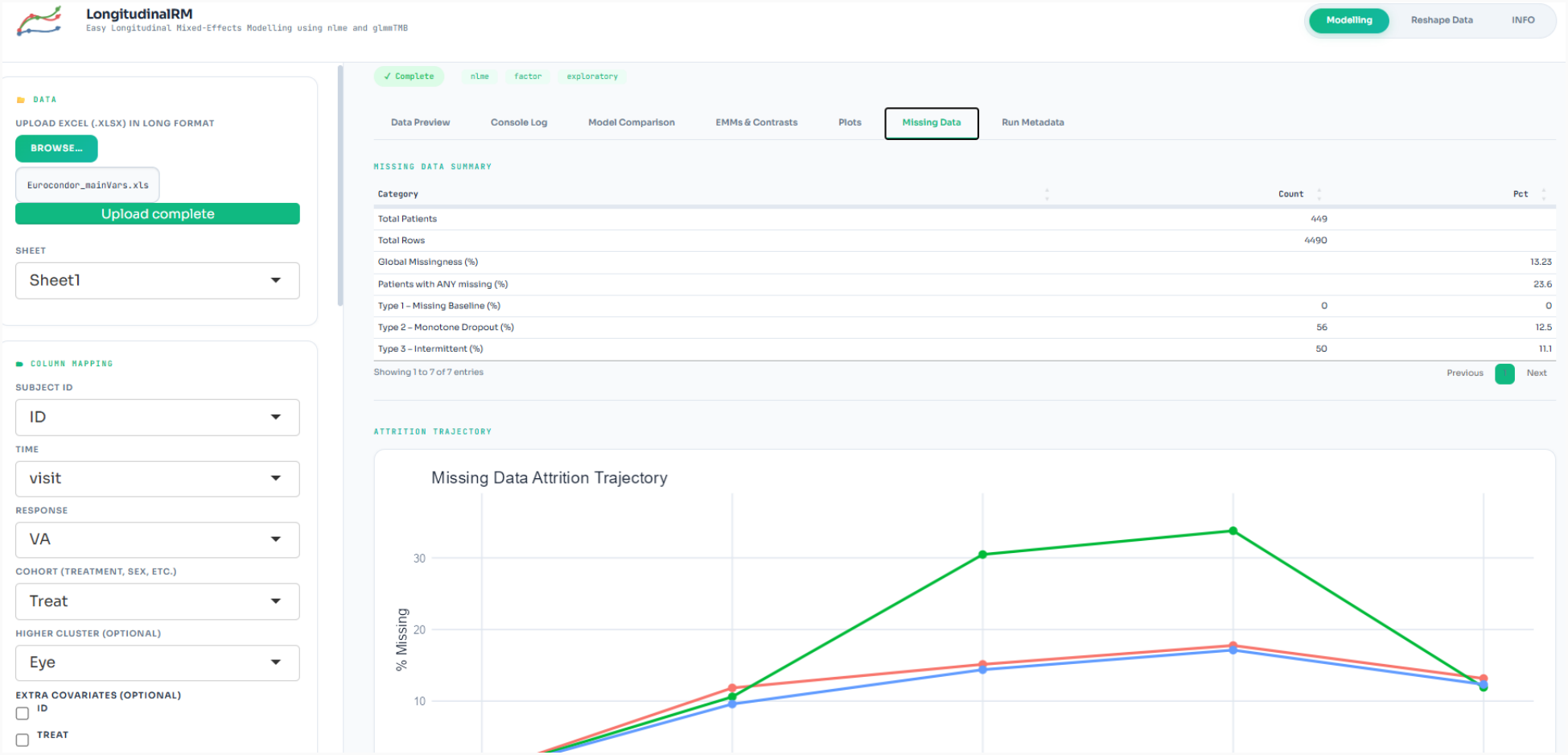
Frontend

→ Structure & layout (HTML)

→ Visual styling (CSS)

→ Interactivity & logic (JS)

User Interface



The Vibe coding Dilemma

What is Vibe Coding?

Intent-Driven Development:

- Practice of **prompting AI tools** to generate logic rather than writing manual syntax.
- Introduced by Andrej Karpathy (Feb 2025).
- Shift from *syntax/languages* to *context, intention, and orchestration*.

Human Role:

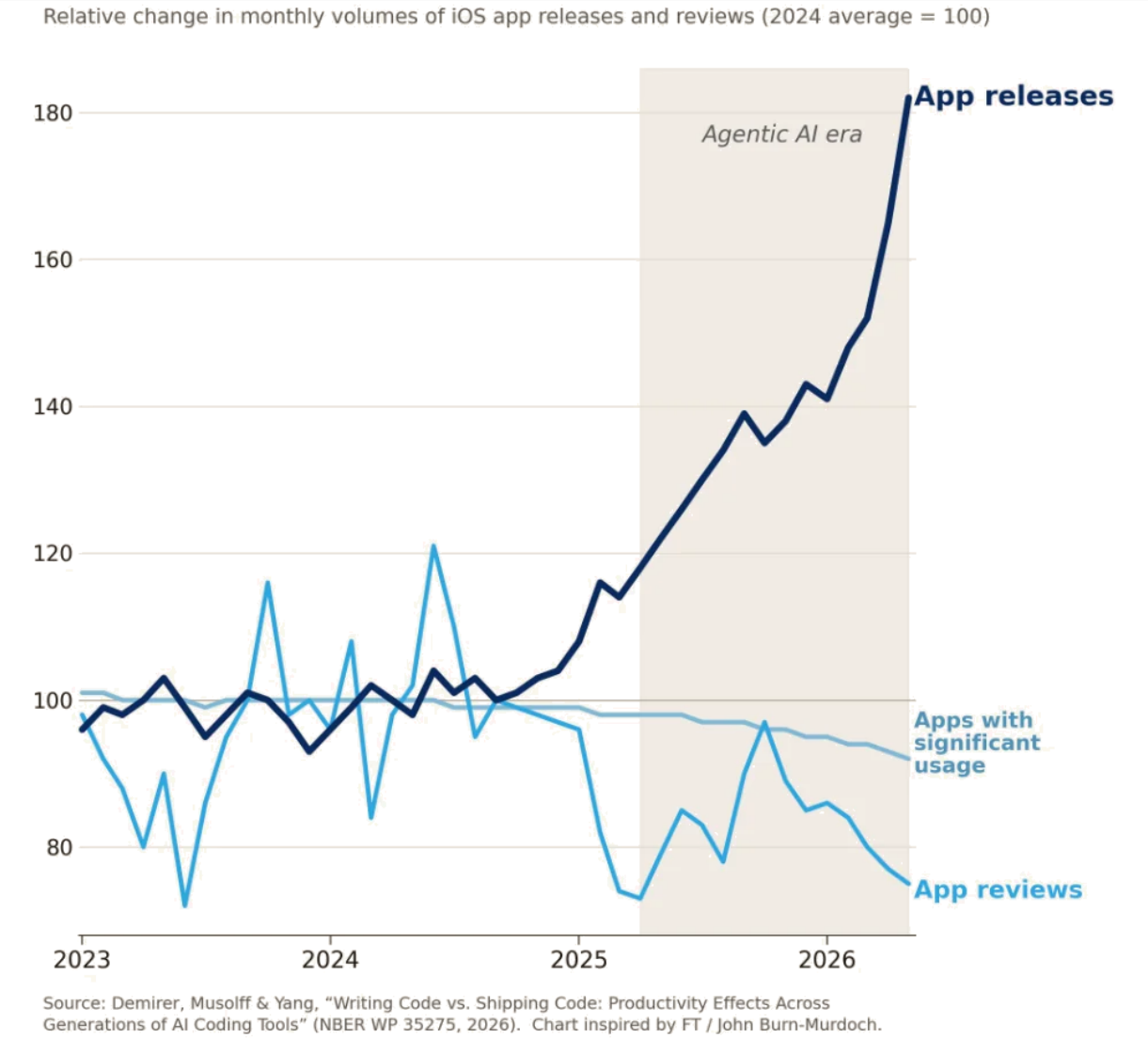
→ Minimal manual coding.

→ Focus on **result outcomes**.

→ Primary tasks: **Testing, fine-tuning, and specification**.

The transition from "How to code" to "What to achieve"

Vibe coding Effects



Closing Remarks

→ An accessible app can make longitudinal modeling practical for clinical researchers.

→ Context determines the appropriate methodology for modeling time.

→ Power estimation approaches can differ heavily depending on trial settings.

→ Vibe coding can be both a bad and a good thing at the same time.

Thank you for your attention

Any Questions or Suggestions welcome

This project is funded by the AGAUR Ajuts per a doctorats industrials from the Generalitat de Catalunya.

References

- Green, Peter and Catriona J MacLeod (Apr. 2016). “SIMR: an R package for power analysis of generalized linear mixed models by simulation”. en. In: *Methods Ecol. Evol.* 7.4, pp. 493–498.
- Iddi, Samuel and Michael C Donohue (Mar. 2022). “Power and sample size for longitudinal models in R - the longpower package and Shiny app”. en. In: *R J.* 14.1, pp. 264–281.
- Liu, G and K Y Liang (Sept. 1997). “Sample size calculations for studies with correlated observations”. en. In: *Biometrics* 53.3, pp. 937–947.
- Lu, Kaifeng, Xiaohui Luo, and Pei-Yun Chen (2008). “Sample size estimation for repeated measures analysis in randomized clinical trials with missing data”. en. In: *Int. J. Biostat.* 4.1, Article 9.

Diggle et. al. (2002) Power Calculation

$$Z_P = \sqrt{\frac{mJs_t^2\Delta^2}{2\sigma^2(1-\rho)}} - Z_{1-\alpha/2}$$

Final Statistical Power = $\Phi(Z_P)$

Design & Effect Parameters

- m Sample size *per group* (total $N = 2m$).
- J Total number of repeated measurements per participant.
- s_t^2 Within-participant variance of time points (how spread out the assessments are).
- Δ Target effect size (the difference in slopes / mean rate of change between groups).

Variance & Error Parameters

- σ^2 Residual (error) variance of the continuous outcome.
- ρ Assumed correlation between any two repeated measurements (compound symmetry).
- $Z_{1-\alpha/2}$ Quantile for chosen Type I error rate α (e.g., 1.96 for $\alpha = 0.05$).

Liu & Liang (1997) Power Calculation

$$Z_P = \sqrt{\frac{N\pi_A\pi_B\beta_1^2(\mathbf{1}^\top R^{-1}\mathbf{1})}{\sigma^2}} - Z_{1-\alpha/2}$$

Final Statistical Power = $\Phi(Z_P)$

Design & Effect Parameters

N Total sample size across all groups.

π_A, π_B Proportion of participants in Group A and Group B (e.g., 0.5 and 0.5).

β_1 Target effect size (difference in average response between groups).

Variance & Matrix Parameters

σ^2 Residual (error) variance of the continuous outcome.

R Working correlation matrix ($J \times J$) for repeated measurements.

$\mathbf{1}^\top R^{-1}\mathbf{1}$ Sum of the inverse correlation matrix elements (scales effective info).

$Z_{1-\alpha/2}$ Quantile for chosen Type I error rate α (e.g., 1.96 for $\alpha = 0.05$).

Lu, Luo & Chen (2008) Power Calculation

$$Z_P = \sqrt{\frac{n_{A1}\delta_J^2}{(\psi_A + \lambda\psi_B)\sigma_J^2}} - Z_{1-\alpha/2}$$

Final Statistical Power = $\Phi(Z_P)$

Design & Effect Parameters

n_{A1} Initial sample size for treatment group 1.

λ Allocation ratio at the first time point (n_{A1}/n_{B1}).

δ_J Target effect size specifically at the *last* time point *J*.

Variance & Attrition Parameters

ψ_A, ψ_B Variance inflation factors (quantifies information loss due to missing data/dropout).

σ_J^2 Residual variance of the outcome at the last time point *J*.

$Z_{1-\alpha/2}$ Quantile for chosen Type I error rate α (e.g., 1.96 for $\alpha = 0.05$).

The Degrees of Freedom in Mixed Models

- ✗ **Unknown df:** Correlated data means within-cluster observations aren't fully independent.
- ✗ **Infinite df assumption:** Software often defaults to a *z*-distribution (Standard Normal).
- ✗ **Small *N* risk:** With few clusters, the *z*-distribution is too narrow and underestimates uncertainty.
- ✗ Inflated Type I error (too many false positives).

- ✓ **Approximation:** Mathematically estimates the *effective* degrees of freedom using variance components.
- ✓ **Switch to *t*-test:** Allows the use of a *t*-distribution, which has wider tails for small sample sizes.
- ✓ **Kenward-Roger:** Similar robust alternative.
- ✓ Accurate, trustworthy p-values.

simr Notation Definitions

1. Simulation Variables

$\hat{M}$ **(Fitted Model):**
The estimated pilot model containing baseline coefficients ($\hat{\beta}$) and variances ($\hat{\Sigma}$). Estimates derived from real data.

$\tilde{y}^{(b)}$ **(Simulated Data):**
The pseudo-response data generated during Monte Carlo loop

2. GLMM to LMM Link

$\mathcal{F}$ **(Exponential Family):**
Demonstrates the approach supports multiple distributions (Gaussian, Binomial, Poisson), not just normal data.

$g(\cdot)$ **(Link Function):**
Connecting linear predictors to the non-linear probability distribution (e.g., logit for binomial).

Simplification for LMMs:
For standard continuous data, $g(\cdot)$ is the **Identity Link** ($g(x) = x$), simplifying the equation to standard linear form:

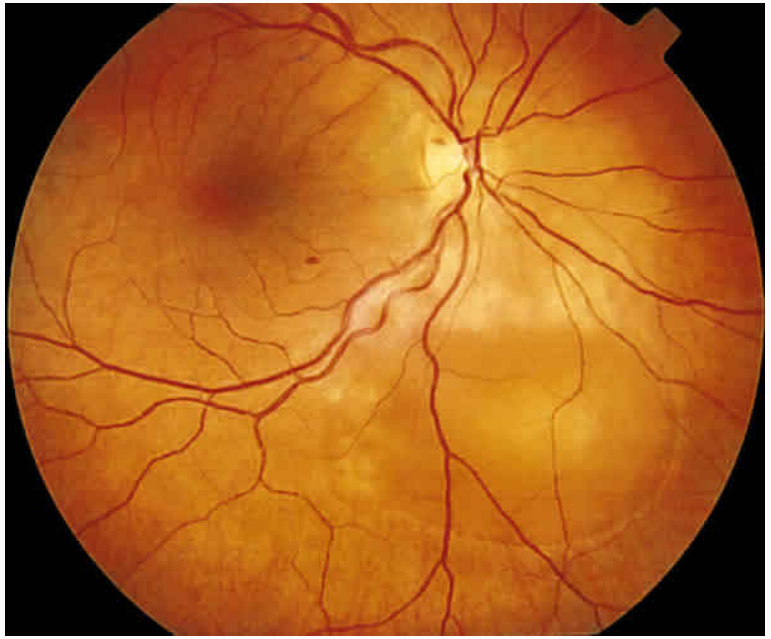
$$\mu_{ij} = \mathbf{x}_{ij}^{\top} \boldsymbol{\beta} + \mathbf{z}_{ij}^{\top} \mathbf{u}_i$$

3. Wilson Confidence Interval

Estimated power is a binomial proportion, the Wilson score interval is used. Superior to the standard Wald interval, accurate even when power approaches 100% or 0%.

The Project Objectives

Improve Diabetic Retinopathy (DR) methods of diagnosis and prediction using historical trial data.



The Project Objectives

Improve Diabetic Retinopathy (DR) methods of diagnosis and prediction using historical trial data.

Longitudinal measurements from *both eyes* across multiple visits.

→ Highly nested, correlated data structures.
→ Temporal and inter-eye correlation modeling.
→ Missing data on higher tier clusters.