Karim Naguib — Curriculum Vitae

Economist/Data Scientist | Bayesian Methods & Causal Inference
karim.a.naguib@gmail.com | karimn.github.io | github.com/karimn

Summary

Economist and applied Bayesian statistician with 10+ years building **hierarchical models** for **longitudinal and survival data**. Current focus: **patient-level digital-twin models** of oncology tumor dynamics and real-world evidence integration at AstraZeneca; prior work spans field experiments in East Africa and South Asia, real-estate market design, and health-behavior interventions.

Skills

Methods: Bayesian hierarchical modeling, state-space and longitudinal models, multistate survival and competing risks, joint longitudinal–survival ("digital twin") models, cross-validation and predictive assessment (LFO-CV, PSIS-LOO), MCMC diagnostics and prior sensitivity, causal inference and randomized experiment design, Gaussian processes

Languages: R (tidyverse), Stan, Julia, SQL, C/C++

Agentic Workflows for Statistical Modeling: Built and maintain an agentic-workflow layer on Claude Code, shared across the AstraZeneca modeling team via an internal plugin marketplace. Collaborative mathematical derivation with Claude as derivation partner, Stan diagnostic sub-agents, prior sensitivity review, `targets` pipeline explorer, MCP-integrated skills for Domino job management, reusable plugin and skill templates codifying team conventions for Bayesian workflow.

Infrastructure: CmdStan / cmdstanr, `targets` pipelines, Domino, Docker, AWS, Databricks, Snowflake, PostgreSQL, git, SLURM, RStudio / VS Code

Education

Boston University | Boston, MA | **Ph.D. in Economics** | 2014 *Fields: development, health, and experimental economics*

The American University in Cairo | Cairo, Egypt | **B.S. in Computer Science, Minor: Mathematics** | 1999

Experience

Senior Data Scientist | **AstraZeneca** | Aug 2023 – present

- Bayesian digital-twin modeling of oncology tumor dynamics for patient-level forecasting of Phase 3 trial outcomes, integrating emerging Phase 2 evidence with historical trial data and real-world data to support Phase 3 initiation decisions (Ph3ID)
- Built hierarchical state-space models of longitudinal tumor size (Stein-Fojo framework) with population/trial/patient hierarchy, non-centered parameterization, and Student-t priors to handle prior-data conflict in sparse cohorts — surfacing early tumor-dynamics signals predictive of downstream survival, validated via leave-future-out cross-validation

- Developed multistate (illness-death) survival models decomposing progression-free survival into distinct progression and death hazards, jointly fit with the tumor-dynamics model so the patient-level latent tumor state drives burden-related progression risk mechanistically
- Contributed modular Stan model architecture and a reproducible `targets`-based pipeline used across multiple trial analyses; routine work on MCMC diagnostics (R-hat, ESS, E-BFMI, divergences) and prior sensitivity
- Built and shared an agentic-workflow layer — collaborative mathematical derivation, diagnostic sub-agents, and MCP-integrated pipelines — adopted by the modeling team for deriving, implementing, and debugging hierarchical Bayesian models

Senior Data Scientist | **Opendoor** | May 2021 – May 2022

- Designed randomized experiments to estimate demand and supply elasticities informing pricing policy
- Built hierarchical Bayesian models and Gaussian-process priors over spatial/temporal regional structure to pool information across markets while preserving local heterogeneity
- Worked across experiment design, inference, and decision-facing deployment

Independent Researcher / Consultant | May 2019 – May 2021

- Built structural Bayesian models of social-experiment outcomes to recover primitive behavioral parameters from reduced-form treatment effects
- Developed hierarchical models for intervention effectiveness across study contexts
- Estimated time-varying epidemiological models of COVID-19 transmission during the early pandemic

Economist | **Evidence Action** | May 2014 – May 2019

- Designed and analyzed large-scale randomized evaluations for Evidence Action's flagship programs: deworming across East Africa and the at-scale replication of the No Lean Season seasonal-migration subsidy program in Bangladesh
- On No Lean Season, contributed to the analysis that failed to replicate the earlier pilot's effect on migration and consumption; the finding informed Evidence Action's decision to close the program — an instance of rigorous evidence driving a hard organizational decision
- Work across projects spanned study design and power analysis, field data-collection supervision, and Bayesian and frequentist analysis for operational and policy audiences

Software Design Engineer in Test | **Microsoft**, Windows Debugging Tools | Apr 1999 – Mar 2007

- Test framework and automation for Windows debugging tools; C/C++, systems-level testing

Projects

Hierarchical state-space "digital twin" model for longitudinal tumor dynamics

Problem. Oncology trials generate repeated tumor-size measurements that carry early predictive signal for survival, but standard analyses either reduce them to crude categorical response (RECIST) or fit patient-by-patient curves that ignore between-study structure. A Phase 3 initiation decision needs patient-level forecasts that borrow appropriately from hierarchically related trials.

Method. Built a Bayesian state-space "digital twin" model of log tumor size under a Stein-Fojo decay-plus-growth framework with a three-level hierarchy (population → trial → patient), non-centered parameterization, and Student-t priors on raw effects to absorb prior-data conflict in sparse strata (small subgroups). Each patient has a personalized posterior trajectory that borrows strength from their trial and the population. Implemented in Stan with modular include architecture; diagnostics included R-hat, ESS, E-BFMI, and prior-posterior conflict checks.

Impact. Extracts tumor-dynamics signal from Phase 2 and historical trial data to forecast Phase 3 outcomes at both population and patient levels, supporting Phase 3 initiation decisions on programs combining emerging trial evidence with real-world data.

Multistate illness-death survival model for competing-risk decomposition

Problem. Progression-free survival bundles multiple distinct events — disease progression and death — under mechanisms that typically differ in hazard shape and covariate dependence. Conventional PFS analyses collapse this structure and can mislead trial-vs-historical comparisons when the event mix shifts between cohorts. And although much of progression is driven by underlying tumor-burden dynamics (SLD, PSA), important events are *not* automatically explained by burden — non-target lesion progression and clinical progression are clinician-adjudicated and not mechanically tied to the modelled burden trajectory.

Method. Built a Bayesian illness-death multistate model with three transitions (progression, death-without-progression, death-after-progression) **jointly fitted with the mechanistic tumor-dynamics model** (above). The patient-level latent tumor state enters the progression hazard directly, so burden-driven progression is captured mechanistically; the multistate layer adds the structure needed for progression events not driven by modelled burden (non-target and clinical progression) and for competing death hazards. The same framework supports treating informative dropout — non-administrative censoring and line-of-therapy switches — as explicit transitions rather than ignorable censoring. Trial-level hierarchy; interval censoring for assessment-gap uncertainty; weighted likelihood for downstream real-world-data borrowing.

Impact. Makes competing-risk and informative-censoring structure explicit in survival comparisons, enabling defensible hazard-level comparisons between trial arms and historical or real-world cohorts where event mix and follow-up structure differ — supporting Phase 3 initiation decisions on programs that combine emerging trial evidence with historical and real-world data.

Modular Bayesian inference infrastructure for clinical trial analyses

Problem. Bayesian trial models accumulate components (tumor dynamics, multistate survival, covariate effects, data-borrowing schemes) that are repeatedly recombined for different trial analyses; hand-rolled Stan models drift in interface and diagnostics across analyses, making results hard to reproduce and compare.

Method. Designed a modular Stan architecture in which each sub-model (tumor regression, growth fraction, initial state, multistate hazards, measurement error, RWD borrowing) exposes a uniform 7-file interface — flags, data, hyperparameters, transformed data, parameters, transformed parameters, priors. Wired into a `targets`-based R pipeline with standardized diagnostics reports, prior-sensitivity checks, and reproducible artifact storage.

Impact. Same components now feed multiple concurrent trial analyses with consistent diagnostics and priors; new model variants are composed rather than rewritten.

Leave-future-out cross-validation for hierarchical longitudinal-survival ("digital twin") models

Problem. Assessing the predictive accuracy of a patient-level tumor-dynamics and survival model requires scoring forecasts against data the model has not seen, but standard cross-validation (leave-one-out over visits or patients) conflates two kinds of held-out information — *future* observations from seen patients and *unseen patients at seen times* — and can make a model look better than it actually is at the forward-in-time forecasting task that matters for Phase 3 decision support. Leave-future-out cross-validation is well-established for time-series models (Bürkner, Gabry, Vehtari 2020) but is not routine for hierarchical longitudinal-plus-survival models, where validation typically leans on leave-one-subject-out or survival-specific summaries (time-dependent ROC, Brier scores).

Method. Adapted LFO-CV to a hierarchical state-space digital-twin model of tumor dynamics and survival: refit at successive truncation times and score out-of-sample predictive density on every held-out quantity with a log-likelihood contribution — longitudinal tumor-size visits *and* survival/multistate events — then summarize forecast performance as a function of prediction horizon. Primary workflow uses full-refit LFO-CV; PSIS-based approximate LFO is also supported for faster iteration. Built as a separate Stan variant sharing the base model's modular structure so validation tracks the production model exactly.

Impact. Produces horizon-stratified predictive-accuracy curves across both biomarker and survival outcomes — a more honest answer to "how well does the digital twin forecast?" than aggregate LOO — and makes forecast calibration directly auditable for Phase 3 decision reviews.

Agentic-workflow layer for Bayesian statistical modeling

Problem. Bayesian trial-modeling work has two hard parts that off-the-shelf AI assistants don't address. First, the *mathematical* layer: multistate survival with joint longitudinal submodels, non-centered hierarchical parameterizations, competing risks, and interval censoring produce dense derivations where a single person working by hand is both slow and error-prone — small algebra mistakes propagate into Stan code and surface weeks later as pathological sampling. Second, the *development-loop* layer: MCMC runs take hours to days, diagnostics and reparameterization require domain judgment, and state across the feedback loop is hard to carry. Generic coding assistants miss the statistical context for both.

Method. Designed and built an agentic-workflow layer on Claude Code, shared across the modeling team via an internal plugin marketplace. Core components: **collaborative mathematical derivation** — working through the underlying statistical mathematics with Claude as a derivation partner to produce accurate, well-documented, and computationally tractable formulations before they become Stan code; a persistent memory system for carrying model-specific context across sessions; domain-specific sub-agents (Stan diagnostic triage, prior sensitivity review, `targets` pipeline explorer); MCP-integrated skills for pipeline operations and job management on Domino; and reusable skill and plugin templates that codify team conventions for Bayesian workflow. Version-controlled and iterated like any other team codebase.

Impact. Shifts the mathematical-modeling step from error-prone solo derivation to collaboratively-verified, fully-documented model specifications; compresses the hand-off between "MCMC finished" and "I know what to do next," which is where most of the latency in Bayesian model development actually lives. Adopted across the modeling team for reproducible Stan development, MCMC diagnostics triage, and pipeline operations.

Publications

Peer-reviewed

Barker, N., Davis, C. A., López-Peña, P., Mitchell, H., Mobarak, A. M., Naguib, K., Reimão, M., Shenoy, A., & Vernot, C. "Migration and Resilience during a Global Crisis." *European Economic Review* (accepted).

Working papers

Mitchell, H., Mobarak, A. M., Naguib, K., Reimão, M., & Shenoy, A. "External Validity and Implementation at Scale: Evidence from a Migration Loan Program in Bangladesh."

Jee, E., Karing, A., & Naguib, K. "Optimal Incentives in the Presence of Social Norms: Experimental Evidence from Kenya."