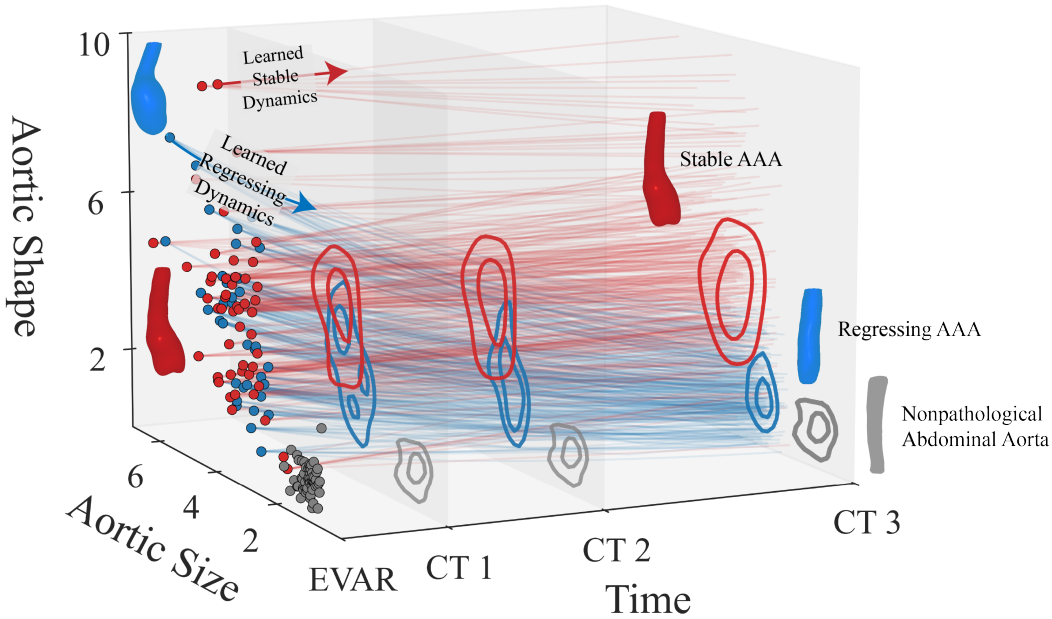
Dynamic Temporal Modeling of Abdominal Aortic Aneurysm Morphology with Z–SINDy

Joseph A. Pugar[1,†], Junsung Kim[1,2,†], Michael Mansour[1,3], Nhung Nguyen[1], Cheong Jun Lee[4], Hence Verhagen[5], Ross Milner[1], Andrei A. Klishin[6,*], and Luka Pocivavsek[1,*]

[1]Department of Surgery, University of Chicago, Chicago, IL, USA
[2]Harvard Medical School, Harvard University, Boston, MA, USA
[3]Department of Medical Physics, University of Chicago, Chicago, IL, USA
[4]Endeavor Health, Evanston, IL, USA
[5]Department of Vascular Surgery, Erasmus University Medical Center, Rotterdam, the Netherlands
[6]Department of Mechanical Engineering, University of Hawai‘i at Mānoa, HI, USA
[*]Authors to whom correspondence should be addressed.
[†]Co-first authors.



Abstract. Abdominal aortic aneurysm (AAA) is characterized by a localized enlargement of the aorta with risk of rupture. Endovascular aneurysm repair (EVAR) alters sac morphology in ways not captured by size alone and typically observed only at sparse follow-up times. We present a physically interpretable, noise-aware framework that models postoperative remodeling dynamics in a low-dimensional state space defined by normalized sac surface area ($\widetilde{A}$) and the normalized fluctuation in integrated Gaussian curvature ($\widetilde{\delta K}$). Using Sparse Identification of Nonlinear Dynamics (Z–SINDy), we infer ordinary differential equations governing the temporal evolution of $(\widetilde{A}, \widetilde{\delta K})$ for clinically defined cohorts of regressing and stable sacs. The learned models yield class-specific flow fields and fixed points that summarize long-term behavior: regressing sacs converge toward a low-size/low-shape attractor, whereas stable sacs maintain near-constant size with persistently elevated shape. Embedding these dynamics in a Bayesian decision framework enables complementary static (coordinate-based) and dynamic (derivative-based) classifiers. Across cohorts, the dynamic classifier separates outcomes earlier and with higher confidence, demonstrating that rates of change become informative before static anatomical distributions separate. Stress tests injecting spatial noise and imposing realistic, irregular follow-up cadences quantify performance degradation and sensitivity to measurement fidelity and scheduling. Together, these results provide a principled route from interpretable geometric features to individualized, probabilistic forecasts of AAA remodeling post-EVAR and offer actionable guidance for surveillance design.

Introduction

Abdominal aortic aneurysms (AAA) are localized dilations of the aorta that can develop into a balloon-like *sac* with substantial rupture risk if left untreated. Endovascular aneurysm repair (EVAR) is the dominant minimally invasive intervention used to exclude the sac from arterial flow, promote thrombus formation, and ideally induce sac regression over time (Fig. 1) [1, 2, 3]. Despite widespread adoption, predicting long-term outcomes post-intervention remains challenging [4, 5, 6]. Clinicians primarily rely on size assessments at discrete time points (typically cross-sectional diameter or sac volume) to characterize status [7, 8, 9]. In addition, pathological shape, captured by the fluctuation in integrated Gaussian curvature δK, contains information about disease progression [10].

However, instantaneous anatomical descriptors are often insufficient to infer the underlying trajectory of disease progression. Differentiating a stable post-EVAR sac from one undergoing regression or expansion requires dynamic characterization of morphology over time. The importance of dynamics is reflected in supplemental guidance that characterizes regression by annual rates of maximal diameter change [11], with thresholds such as $\geq$5–10 mm at 1 year used in practice [4, 12]. Failure to regress at 1 year and sac expansion are associated with higher long-term mortality independent of endoleaks or reinterventions [4, 5], and a "stable" sac carries worse prognosis than a regressing sac [6, 13]. Early sac shrinkage (e.g., $\geq$10 mm at 1 year) is linked to freedom from fatal adverse events [12], whereas delayed and persistent type II endoleaks correlate with sac growth and adverse late outcomes [14, 15]. The ability to model and anticipate trajectories could inform surveillance frequency, enable early identification of treatment failure, and guide personalized intervention strategies. Substantial regression to $\leq$40 mm has been associated with very low risk and may permit extended follow-up intervals in selected patients [16].

The dynamics of AAA remodeling are shaped by biomechanical, anatomical, and patient-specific factors. Real-world datasets are limited in size and irregularly sampled, yet they encode latent patterns of morphological change. Rather than reducing outcomes to discrete labels, it is valuable to view post-EVAR behavior as a continuum governed by rates and directions of change in biologically meaningful features. Clinical imaging is sparse: patients may contribute as few as two usable time points (typically one pre- or peri-operative scan and one postoperative scan) and at most about five, separated by months or years. This sparsity—compounded by variation in imaging frequency, segmentation algorithms, and device type—demands models that are robust to noise and heterogeneity and capable of extracting population-level trends from underdetermined individual trajectories. Heterogeneity is further amplified by device-specific effects and repair complexity: most sac regression occurs within the first two years [17]; regression patterns vary across endografts [12]; and complex repairs (fenestrated/branched) exhibit distinct sac dynamics requiring intensified surveillance when growth occurs [18, 19].

Traditional approaches to temporal modeling in clinical science—spanning biostatistics, time-series machine learning [20, 21], and black-box deep learning [22, 21]—are often inadequate here. Linear models underfit nonstationary biological processes, while flexible nonlinear methods risk overfitting noise and spurious fluctuations. These frameworks also rarely accommodate uncertainty from variable sampling intervals or missing data [23, 20, 24]. Black-box models may extract unimportant patterns while overlooking core mechanistic relationships [22, 21]. Effective clinical models must balance complexity and interpretability, capturing coarse-grained system dynamics without overfitting high-dimensional noise. This calls for frameworks that incorporate physical structure, account for data corruption, and

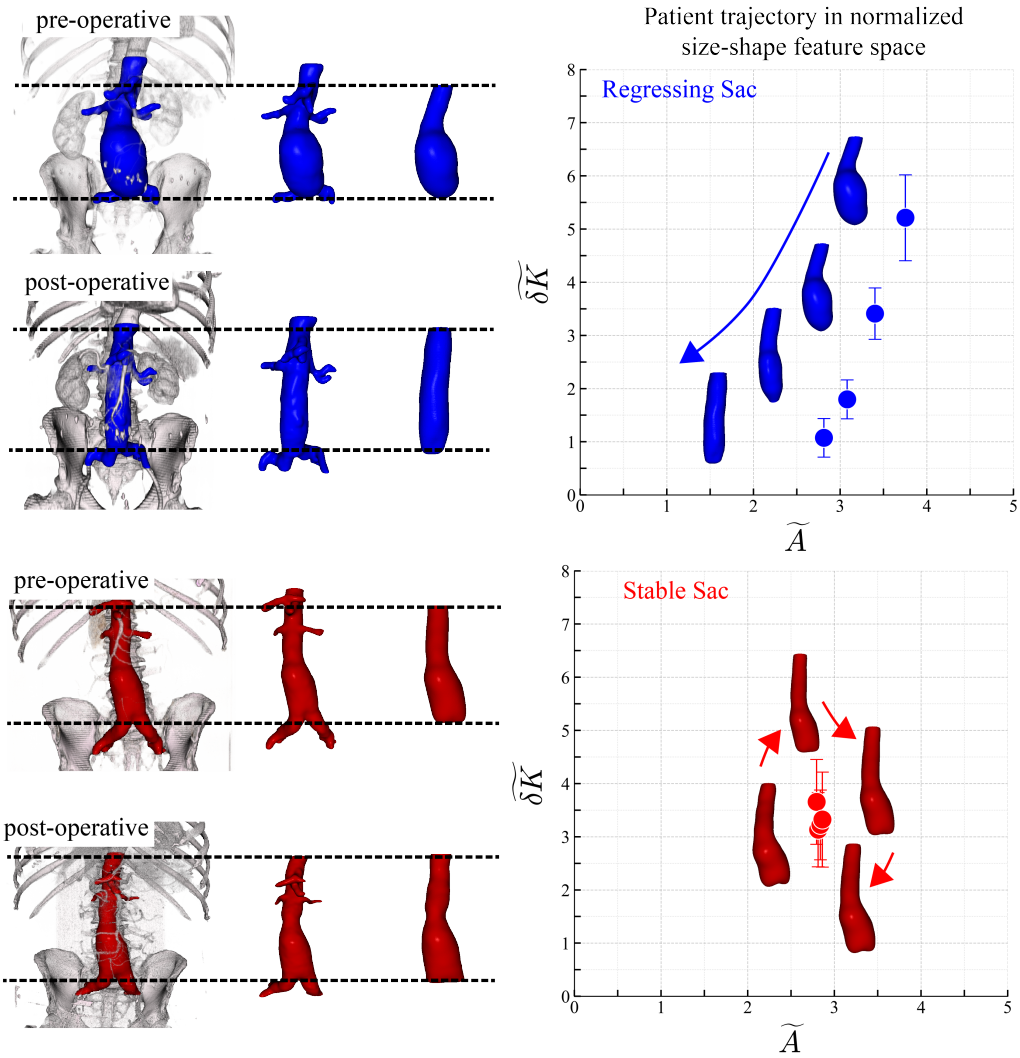


Figure 1: Defining aortic anatomy and mapping trajectories into size–shape space. **Left:** 3D segmentations of AAA scans before and after EVAR. Dashed horizontal lines mark the superior and inferior cut planes used to standardize the analyzed sac segment across time points. **Right:** Each scan is projected into a normalized size–shape space ($\widetilde{A}$, $\widetilde{\delta K}$), where $\widetilde{A}$ denotes normalized surface area and $\widetilde{\delta K}$ the normalized fluctuation in integrated Gaussian curvature. Points are serial scans connected chronologically; arrows indicate temporal progression. Top row: regressing sac (blue; $> 10\%$/year decrease in surface area). Bottom row: stable sac (red; $< 10\%$/year change). Together, the panels show the workflow—from standardized segmentation via consistent cut planes to quantitative embedding—used to characterize postoperative sac behavior in a common, interpretable feature space.

remain interpretable to clinicians, particularly when predicting forward trajectories under real-world constraints such as follow-up attrition and irregular scan timing. In parallel, morphovolumetric metrics can be more sensitive than diameter alone for detecting clinically meaningful sac change, underscoring the need to target features that better reflect underlying biomechanics [7].

In this work, we develop a physically interpretable, noise-aware framework for modeling post-EVAR AAA remodeling using a sparse identification approach. Sparse Identification of Nonlinear Dynamics (SINDy) [25, 26] has successfully identified governing equations from synthetic data and reduced-order models, but has seen more limited success with noisy experimental data [27, 28, 29]. We use Z–SINDy [30], a version equipped with uncertainty quantification rooted in statistical mechanics. We infer sparse, interpretable ordinary differential equations (ODEs) governing the temporal evolution of sac surface area A and sac shape δK, features that compactly encode aneurysm morphology. More broadly, our approach provides a general framework for learning and validating low-order dynamical models from sparse, irregularly sampled clinical time series under realistic noise and attrition.

We curated a dataset of 75 patients and 220 CT scans across two clinical cohorts in different countries. Our framework integrates clinical CT time-series data with finite-element–based synthetic up-sampling (FEA) [31] to mitigate sparsity and enhance signal. We demonstrate that Z–SINDy reproducibly identifies dynamic signatures associated with sac regression versus stabilization, provides interpretable steady-state fixed points with clinical meaning, and enables individualized probabilistic forecasts of anatomical evolution. We further examine how sampling frequency, noise, and attrition affect model fidelity, and we propose a principled way to assess confidence in outcome prediction using a Bayesian classification model, thereby aligning physical interpretability with clinical utility.

Results

Data curation

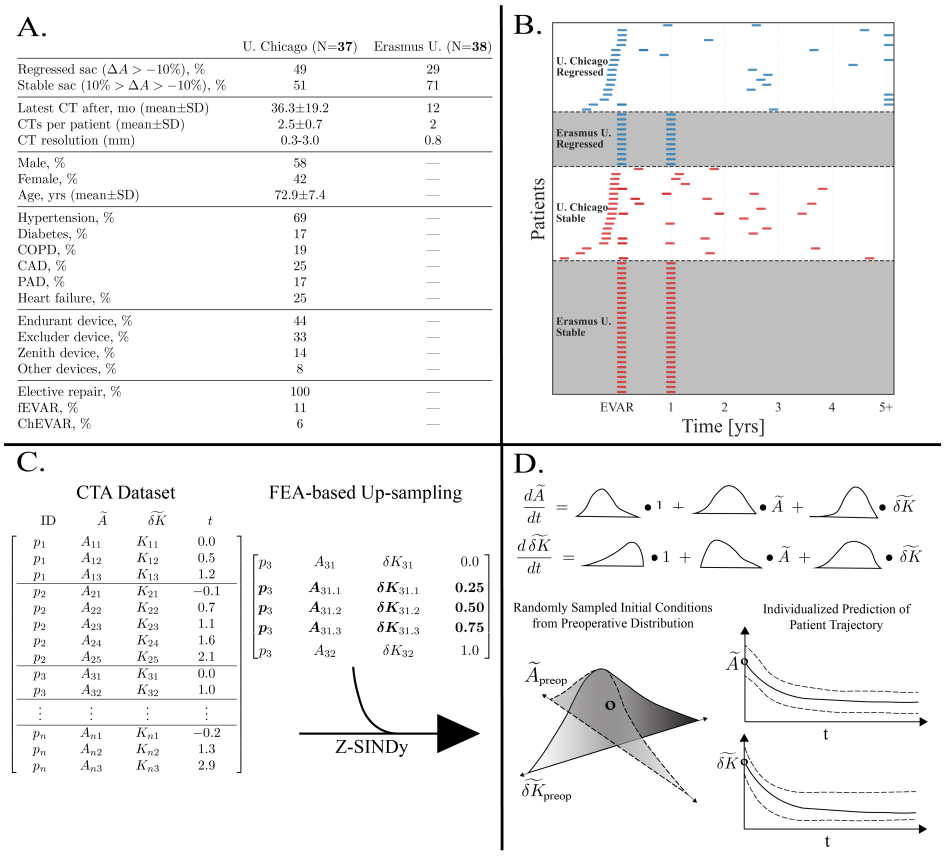
We curated a dataset of 75 patients and 220 CT scans across two clinical cohorts from different institutions: the University of Chicago (USA) and Erasmus University Medical Center (the Netherlands), with basic demographic statistics in Fig. 2A. Owing to differences in healthcare systems, the acquisition cadence differs substantially: in the Erasmus cohort each patient typically has one preoperative CT scan immediately before EVAR and one follow-up exactly 12 months later, whereas in the University of Chicago cohort the preoperative scan may precede EVAR by several months and follow-ups are irregular in both spacing and number (Fig. 2B). The scan-based data points were upsampled with the FEA procedure to increase data density (Fig. 2C).

Modeling post-EVAR dynamics with Z–SINDy

We first asked whether coarse-grained dynamics of postoperative sac remodeling can be captured by low-order, interpretable models in the normalized size–shape space ($\widetilde{A}$, $\widetilde{\delta K}$):

$$\frac{d\widetilde{A}}{dt} = f(\widetilde{A}, \widetilde{\delta K}; \boldsymbol{\Xi}), \qquad \frac{d\widetilde{\delta K}}{dt} = g(\widetilde{A}, \widetilde{\delta K}; \boldsymbol{\Xi}),$$

where $\boldsymbol{\Xi}$ are model coefficients inferred from the training data. We fitted Z–SINDy models separately to subjects clinically labeled regressing and stable to obtain posterior distribu-



	U. Chicago (N=**37**)	Erasmus U. (N=**38**)
Regressed sac ($\Delta A > -10\%$), %	49	29
Stable sac ($10\% > \Delta A > -10\%$), %	51	71
Latest CT after, mo (mean±SD)	36.3±19.2	12
CTs per patient (mean±SD)	2.5±0.7	2
CT resolution (mm)	0.3-3.0	0.8
Male, %	58	—
Female, %	42	—
Age, yrs (mean±SD)	72.9±7.4	—
Hypertension, %	69	—
Diabetes, %	17	—
COPD, %	19	—
CAD, %	25	—
PAD, %	17	—
Heart failure, %	25	—
Endurant device, %	44	—
Excluder device, %	33	—
Zenith device, %	14	—
Other devices, %	8	—
Elective repair, %	100	—
fEVAR, %	11	—
ChEVAR, %	6	—

Figure 2: Dataset, sampling, and Z–SINDy pipeline. **A** Demographic and clinical metadata for the two study cohorts (University of Chicago and Erasmus University) used in all analyses. **B** Barcode plot summarizing data-acquisition cadence: each row is a patient; horizontal position is time post-EVAR; ticks indicate CT acquisitions. This view makes temporal sparsity explicit and shows where derivative information for SINDy can be estimated. **C** Input data matrix for modeling, consisting of discrete observations of normalized size–shape variables ($\widetilde{A}$, $\widetilde{\delta K}$, t) per patient. Patients contribute a variable number of rows reflecting the sparsity seen in panel B. To reduce sparsity, the true observations are augmented with FEA-based surrogates so that each patient has samples on a 3-month grid between their first and last scan. **D** Schematic of the Z–SINDy methodology. The matrix in panel C is converted into a sparse system of ODEs with coefficient distributions. The ODEs are integrated forward from initial conditions sampled from the preoperative distribution ($\widetilde{A}_{\text{preop}}$, $\widetilde{\delta K}_{\text{preop}}$), yielding probabilistic trajectories for the regressing and stable dynamical models.

tions of coefficients (Fig. 2D). We then resampled both coefficients and initial conditions and integrated the equations forward, yielding ensembles of trajectories with uncertainty bands around the mean solutions (Fig. 3). For the regressing cohort (blue), both $\widetilde{A}(t)$ and $\widetilde{\delta K}(t)$ decline monotonically toward a low-size/low-shape fixed point; for the stable cohort (red), $\widetilde{A}(t)$ remains approximately constant while $\widetilde{\delta K}(t)$ remains elevated relative to the regressing group. Fixed-point analysis of the mean ODEs gives $\left(\widetilde{A}^*, \widetilde{\delta K}^*\right) \approx (2.1,\ 0.2)$ for the regressing model and $\approx (3.4,\ 2.8)$ for the stable model (Fig. 3, dashed lines), providing compact, clinically interpretable summaries of long-term behavior for each class.

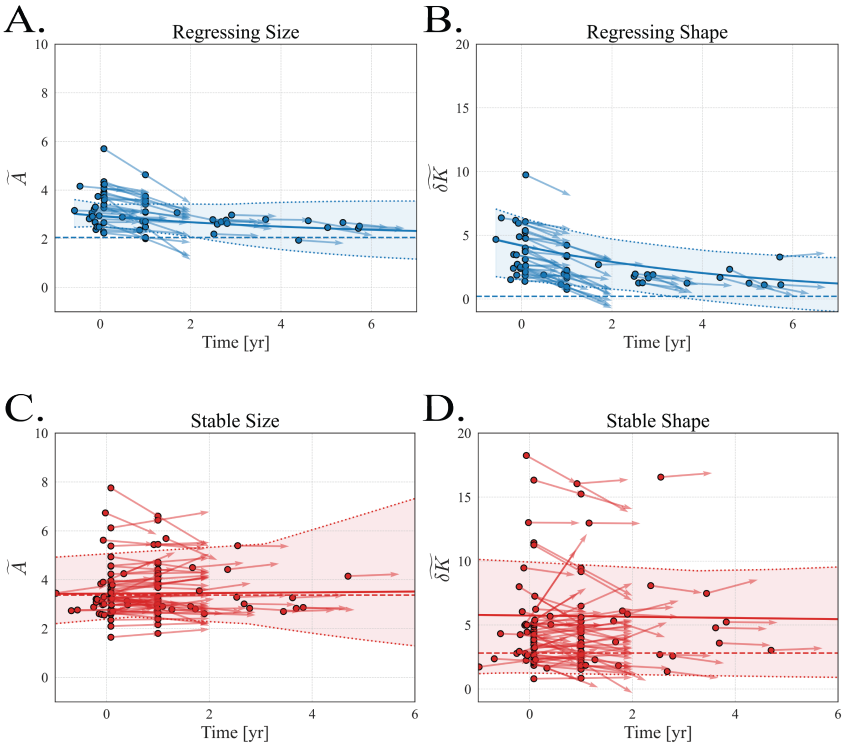


Figure 3: Z–SINDy trajectories for regressing and stable sacs in normalized size–shape space. Panels show the evolution of normalized surface area $\widetilde{A}(t)$ and normalized fluctuation in integrated Gaussian curvature $\widetilde{\delta K}(t)$ over time t (years since EVAR). **A**–**B** (blue): regressing model; **C**–**D** (red): stable model. Points are observed scan values; faint lines are individual Z–SINDy trajectory realizations obtained by integrating the learned ODEs from population-level initial conditions while sampling coefficient distributions. Bold lines are mean trajectories; shaded regions indicate ensemble spread induced by coefficient uncertainty. Dashed horizontal lines mark steady-state (fixed-point) solutions computed from the mean ODEs. Mean fixed points in $(\widetilde{A},\ \widetilde{\delta K})$ are approximately (2.1, 0.2) for the regressing model and (3.4, 2.8) for the stable model. Panel assignments: **A** $\widetilde{A}(t)$ (regressing), **B** $\widetilde{\delta K}(t)$ (regressing), **C** $\widetilde{A}(t)$ (stable), **D** $\widetilde{\delta K}(t)$ (stable).

To relate model-level dynamics to cohort behavior, we rendered population flow fields in the $(\widetilde{A},\ \widetilde{\delta K})$ plane (Fig. 4). The *Raw* panels plot finite-difference patient derivatives overlaid on the convex hull of observed states, showing individual progression vectors: regressing sacs flow down and left, while stable sacs circulate within a higher-$\widetilde{\delta K}$ region with comparatively

small changes in $\widetilde{A}$. *Interpolated* derivatives, projected onto the convex-hull grid, concentrate streamlines and motion magnitude, yielding average behavior for each cohort. The *Model* panels, generated by evaluating the mean Z–SINDy vector fields, mirror the empirical trends and demonstrate the effectiveness of simple parametric models. Stars mark fixed points predicted by each class-specific ODE; local eigenvectors (black lines) align with the principal directions of flow. The regressing model has a stable fixed point (both eigenvalues with negative real parts), whereas the stable model exhibits a saddle (eigenvalues with opposite signs). Characteristic time scales estimated from the eigenvalues of the linearized flow, via $\tau = 1/|\mathrm{Re}\,\lambda|$, indicate substantial remodeling for regressing sacs ($\sim$5.0 years, both eigenvalues with similar negative real parts, consistent with a stable node) and much slower, anisotropic dynamics for stable sacs. The estimated time scales are consistent with the clinical observations: regressing sacs demonstrate substantial remodeling over the $\sim$5 years of follow-up imaging, while stable sacs do not. Together, Figs. 3–4 show that (i) the two outcome groups are governed by distinct dynamics in a two-variable description and (ii) Z–SINDy reproduces cohort flow structure observed in the empirical trajectories while providing a unified system of equations for each outcome class. Throughout the Supplemental Information, we demonstrate that an affine function library is sufficient to capture the essential dynamics of both regressing and stable sacs. In particular, Fig. 8 compares Z–SINDy flow fields learned using second- and third-order polynomial libraries. No qualitative differences are observed: the regressing model continues to exhibit rapid descent toward nonpathological size–shape states, while the stable model retains a saddle structure at intermediate values of $(\widetilde{A}, \widetilde{\delta K})$. Higher-order terms introduce modest local curvature in the vector field but do not alter mean behavior. These results indicate that the observed cohort dynamics are low-order and structurally robust, supporting the use of minimal parametric models without loss of explanatory power.

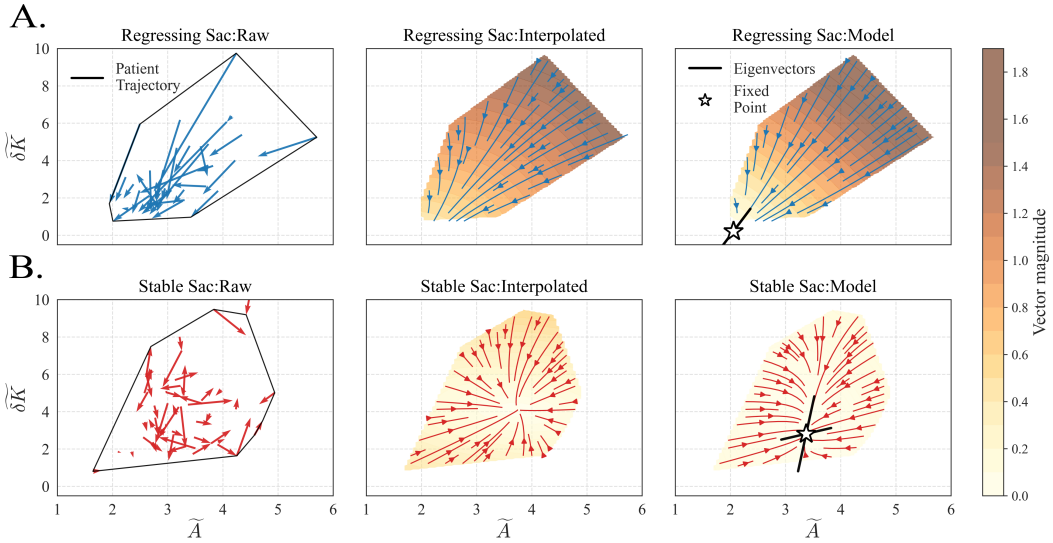


Figure 4: Cohort-level flow fields in normalized size–shape space. Panels illustrate empirical and modeled dynamics of sac remodeling in the $(\widetilde{A}, \widetilde{\delta K})$ plane for regressing (**A**, top row, blue) and stable (**B**, bottom row, red) post-EVAR sacs. *Raw:* finite-difference derivatives from longitudinal trajectories (black outlines represent the convex hull of observed states; arrows are the per-patient vectors). *Interpolated:* smoothed flow fields obtained by gridding/interpolating pooled derivatives with background shading proportional to vector magnitude. *Model:* mean Z–SINDy vector fields with fixed points (stars) and associated eigenvectors (black lines) indicating local stability structure. Eigenvalues of the linearized flow at each fixed point set local contraction/expansion rates; characteristic time scales $\tau = 1/|\mathrm{Re}\,\lambda|$ are $\sim$5.0 years along both eigen directions for the regressing model (stable node with two negative real parts), and $\sim$17.5 and 51.6 years for the stable model, reflecting much slower and direction-dependent remodeling.

Bayesian classification and model stress tests

Having established class-specific dynamics, we next asked whether the Z–SINDy models could support individualized, probabilistic classification of postoperative outcome and how early such decisions can be made under simulated real-world sampling. Figure 5A shows ensembles of 10,000 simulated trajectories generated from the two Z–SINDy models. These ensembles reflect plausible temporal evolutions under coefficient uncertainty for regressing (blue) and stable (red) sacs and already illustrate divergence over time.

Figure 5B reformats these ensembles into one- and two-dimensional distributional views across clinically meaningful reference times. Kernel-density estimates of $\widetilde{A}$ and $\widetilde{\delta K}$ marginals (left and middle) and their joint densities (right) summarize how the classes separate. Immediately after repair the two classes overlap; by two to five years their marginals and joint densities diverge substantially, with regressing sacs shifting toward lower size and shape while stable sacs maintain higher values.

Figure 5C applies the static Bayesian classifier, which evaluates the probability of being a stable sac, $p(\text{Stable} \mid \text{Statics})$, using only instantaneous coordinates in size–shape space. By contrast, Fig. 5D shows the dynamic Bayesian classifier, which evaluates likelihoods, $p(\text{Stable} \mid \text{Dynamics})$, thereby exploiting directional flow information rather than position alone. By convention, $p = 1$ predicts a stable sac, while $p = 0$ predicts a regressing sac.

Across time points, the dynamic classifier produces higher-confidence predictions earlier: by one to two years post-EVAR, a larger fraction of patients exit the "uncertain" band ($0.1 < p < 0.9$) in Fig. 5D than in Fig. 5C. This behavior is reproduced in Supplementary Fig. 11, where the dynamic classifier consistently concentrates posterior mass toward the correct class earlier than the static classifier, using both CT-only data and CT data augmented with FEA-based upsampling, and remains robust under K-fold cross-validation (Table 1).

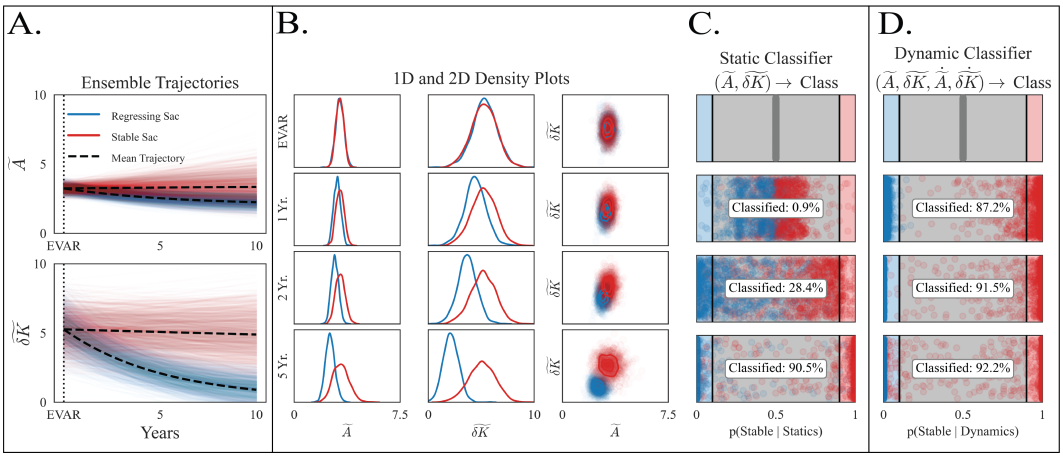


Figure 5: From Z–SINDy ensembles to static and dynamic Bayesian classification. **A** Ensembles of 10,000 simulated trajectories from the regressing (blue) and stable (red) Z–SINDy models for $\widetilde{A}(t)$ (top) and $\widetilde{\delta K}(t)$ (bottom). **B** Distributional views of the same ensembles: 1D marginals (left and middle) and joint densities (right) at EVAR, 1 yr, 2 yr, and 5 yr. The middle column and the inputs to panel C use identical point clouds. **C** Posterior probabilities from the static Bayesian classifier, pooling $p(\mathbf{x}_k \mid c, t_k)$ over observed states. **D** Posterior probabilities from the dynamic Bayesian classifier using empirical derivatives $p(\dot{\mathbf{x}}_k \mid c)$. Under the dynamic classifier, patients exits the "uncertain" region earlier than the static classifier, reflecting the added value of directional information. Importantly, the inputs to C and D use the same underlying point clouds shown in B, enabling a direct comparison between raw ensemble separation and the decision structure of the static classifier.

Finally, we quantified the robustness of classification to clinically realistic spatial and temporal noise (Fig. 6). We injected controlled spatial noise into $(\widetilde{A}, \widetilde{\delta K})$ and coarsened sampling from exactly annual (*Optimal 1.0*) to jittered (*Ideal*, 0.5–1.5 yr) and sparse/irregular (*Reality*, 1–3 yr) intervals, re-estimating derivatives and reapplying the dynamic classifier. The left panel shows the fraction of scans that meet the high-confidence threshold ($\max_c p(c \mid \cdot) \geq 0.9$); the right panel shows the accuracy among this classified subset. Three consistent trends emerged. (i) The proportion of confidently classified scans is systematically reduced by temporal irregularity rather than spatial noise. (ii) When spatial noise is modest (0–10%), accuracy for the classified subset remains high ($\gtrsim 90\%$) even under non-ideal sampling, indicating that most confident decisions are correct. (iii) Under the most aggressive corruption (25% spatial noise with sparse/irregular sampling), accuracy drifts toward the non-informative prior and highlighting the regime in which measurement error and follow-up gaps overwhelm the dynamical signal. Supplementary Fig. 13 further extends

these stress tests by incorporating real-world patient attrition (Fig. 12) and time-dependent class priors (Supplementary Section S8). Under these clinically realistic conditions, the dynamic classifier again exits the uncertain region earlier than the static classifier (well before the majority of patients are lost to follow-up), reinforcing the translational importance of modeling aortic remodeling as a dynamic system rather than static process.

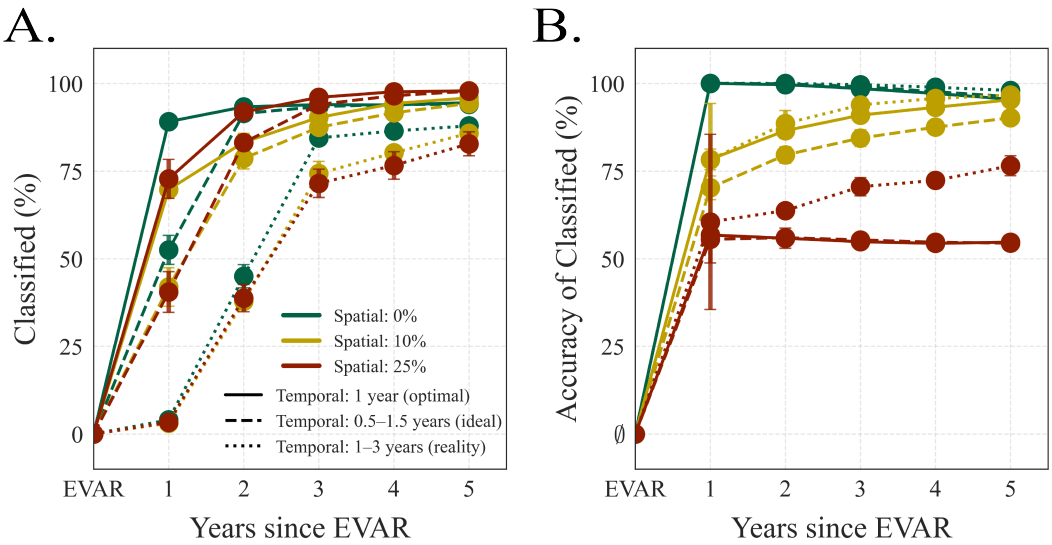


Figure 6: Robustness of the dynamic Bayesian classifier to spatial and temporal noise. **A**: fraction of scans confidently classified ($\max_c p(c \mid \cdot) \geq 0.9$) versus time since EVAR. **B**: accuracy among the classified subset. Colors encode spatial noise added independently to $\widetilde{A}$ and $\widetilde{\delta K}$ prior to classification (green: none; yellow: 10% of feature magnitude; red: 25%). Line styles encode temporal sampling: solid = exactly annual (*Optimal 1.0*); dashed = jittered intervals randomly sampled from $[0.5, 1.5]$ yr (*Ideal*); dotted = sparse/irregular intervals from $[1.0, 3.0]$ yr (*Reality*). The EVAR baseline corresponds to the non-informative 50% prior with essentially no scans classified. Error bars represent variability across 50 realizations. The classifier confidently labels an increasing fraction of scans over time, with accuracy remaining high except under the most severe combination of spatial noise and temporal sparsity.

Discussion

This work leverages a physically interpretable, anatomically grounded approach to model postoperative AAA remodeling and uses the fitted models for clinically actionable classification. Motivated by the limitations of size-only assessments and by sparse, irregular follow-up, we framed sac regression or stabilization as dynamics in a low-dimensional, interpretable state-space described by normalized size ($\widetilde{A}$) and shape ($\widetilde{\delta K}$). Two main findings emerged: (i) outcome-specific Z–SINDy models capture cohort dynamics and yield meaningful fixed points (Figs. 3–4); and (ii) these dynamics enable probabilistic classification that exploits dynamical directional information rather than position alone in size–shape space, providing earlier and more accurate decisions that can be stress-tested for robustness (Figs. 5–6).

Dynamical forecasting

A central challenge for clinically useful time-series models is balancing accuracy with interpretability. By restricting the library to linear terms, Z–SINDy identifies the simplest models consistent with the data while preserving a direct physiological mapping between features and coefficients [32]. For regressing sacs, trajectories trend toward a stable fixed point near $(\widetilde{A}, \widetilde{\delta K}) \approx [1, 1]$, i.e., the definition of a nonpathological aorta in the normalized space; the characteristic time scale ($\sim$5 years) matches typical follow-up imaging. For stable sacs, the mean field exhibits a fixed point with much longer time scales; the model predicts near-constant $\widetilde{A}$ with persistently higher $\widetilde{\delta K}$ on clinical time horizons. Importantly, Fig. 4 closes the loop between model and data: raw finite-difference vectors, interpolated flow fields, and Z–SINDy vector fields show consistent directions across the observed convex hull, indicating that the learned ODEs reflect underlying cohort behavior.

Bayesian classifiers

Translating learned dynamics into clinical utility requires principled decision-making under measurement uncertainty. Our first practical contribution is a pair of Bayesian classifiers—*static* (coordinate-based) and *dynamic* (derivative-based)—that take an individual patient's trajectory and return the posterior probability of regressing or stable outcome. Because derivatives encode *where a patient is heading* rather than *where they are now*, the dynamic classifier becomes discriminative sooner: by 1 year it confidently classifies 87% of patients with high predictive accuracy, whereas the static classifier only classifies 1% of patients (Fig. 5). This is crucial in practice, where only 2–3 scans may be available per patient and early risk stratification can guide surveillance and reintervention strategy. While static or instantaneous characterization of aortic morphology remains useful, dynamic information of anatomic trajectory is invaluable. Clinically, we need to identify non-regressing AAA sacs at one year post-EVAR, not wait until they don't follow up at year 5. The Society for Vascular Surgery's (SVS) Vascular Quality Initiative (VQI) shows loss-to-follow-up is associated with a 6.45-fold higher mortality risk underscoring the need for early risk identification and proactive surveillance [33]. We've seen similar trends with aortic dissection patients at our own institution [34]. Supplementary analyses (Supplemental Section S8) of different attrition rates of regressing and stable cohorts indicate that the mere existence of a patient's follow-up CT scan several years post-EVAR can enhance the classification regardless of the specific scan outcome.

Stress testing and study design

Our second contribution is explicit quantification of how spatial measurement error (e.g., CT resolution and segmentation variability) [35, 36, 37] and temporal irregularity (loss to follow-up, variable scheduling) [38, 39, 40, 41] influence performance of the classifier (Fig. 6). Three conclusions follow. (i) Accuracy improves monotonically with postoperative time across conditions, reflecting accumulating information and growing separation between outcomes. (ii) Spatial noise degrades performance in a predictable fashion; when perturbations reach roughly a quarter of the feature scale, performance approaches the 50% prior. (iii) Temporal irregularity induces meaningful losses relative to exactly annual imaging, emphasizing the value of timely scans for derivative estimation. These stress tests provide a quantitative lens for planning studies, enabling investigators to trade off sample size, cadence, and anticipated

noise to meet prespecified accuracy targets.

The proposed framework suggests an anatomically defined, rate-based approach to post-EVAR surveillance. The size–shape mapping is straightforward to compute from standard CT-derived segmentations, and Z–SINDy has an adaptable open-source implementation [42]. Posterior-predictive simulations generate reference distributions for both outcomes, supporting downstream analyses and visualization. The framework naturally accommodates new data as additional scans become available, enabling adaptive surveillance: patients whose dynamic posteriors favor the regressing model might maintain standard follow-up, whereas those with persistently high $\widetilde{\delta K}$ could be flagged for closer evaluation or adjunctive treatment. More broadly, summarizing cohort behavior with fixed points (and eigenvectors) aids communication by distilling complex remodeling into intuitive descriptors ("which fixed point are you approaching, and how fast?").

Future work

Our formulation rests on simplifying geometric, dynamical, and clinical assumptions that guide future extensions. *Geometry:* we deliberately restricted the feature space to $\widetilde{A}$ and $\widetilde{\delta K}$ to highlight size–shape coupling; richer variables (e.g., thrombus morphology, graft type, hemodynamics) may refine subgroup structure. The fluctuation in integrated Gaussian curvature δK is empirically informative [10] but remains a heuristic; principled shape metrics from differential geometry could be incorporated [43]. *Dynamics:* we used a linear library to balance fit and interpretability; given current data volume/noise, nonlinear extensions offer limited gains, but future datasets may warrant higher-order terms. Measurement noise chiefly affects empirical derivatives ($\dot{\widetilde{A}}, \dot{\widetilde{\delta K}}$); adaptive denoising and derivative-regularization methods could further stabilize estimates in clinical settings [44, 45, 46, 47]. *Clinical scope:* here we modeled two classes within AAA. Extending to expanding sacs (Type I endoleak or pre-rupture), thoracic aneurysms, post-TEVAR remodeling, and other morphology-driven conditions—paired with stress-test paradigms to inform surveillance design and quality standards—is a natural trajectory for further research. Heterogeneity across devices, imaging practices, and segmentation quality was embraced rather than harmonized, underscoring the need for prospective, multi-institutional validation of dynamical modeling approaches.

By reframing AAA remodeling as inference of simple, physically meaningful dynamics, we bridge mechanistic understanding and clinical decision-making. Z–SINDy provides compact models that match cohort behavior and yield interpretable flow structure; Bayesian classifiers convert those models into early, patient-specific probabilities; and stress tests reveal when and how performance degrades under real-world constraints. Together, these elements support translational progress toward reliable, transparent, and resource-aware modeling that can inform how, when, and for whom postoperative surveillance should adapt.

Methods

Clinical data and morphological feature space

We curated longitudinal CT angiography for EVAR patients, each contributing 2–5 usable time points spanning months to years. From each segmented sac we computed the sac surface area A and a shape descriptor defined as the fluctuation in integrated Gaussian curvature, δK. After scale averaging, both were normalized by cohort means to yield dimensionless

features $\widetilde{A}$ and $\widetilde{\delta K}$; the patient state is $\mathbf{x}(t) = [\widetilde{A}(t), \widetilde{\delta K}(t)]^\top$. (Operational details of scale averaging and curvature computation follow our prior work [48].)

Temporal resampling and derivative estimates

To mitigate irregular follow-up, each subject's time series was upsampled to a quasi-regular grid and finite-difference derivatives were computed on the resampled series. We enforced a minimum separation between scans used for differencing and removed clearly non-physiologic spikes in $\mathrm{d}\,\widetilde{\delta K}/\mathrm{d}t$ (Supplementary Sections S1–S2).

Sparse identification of dynamics (Z–SINDy)

We modeled the temporal evolution of $\mathbf{x}(t)$ using Z–SINDy [30]. In the main text we use an affine library

$$\mathbf{\Theta}(\mathbf{x}) = [\,1,\ \widetilde{A},\ \widetilde{\delta K}\,]^\top, \qquad \dot{\mathbf{x}}(t) \approx \mathbf{\Xi}^\top \mathbf{\Theta}(\mathbf{x}(t)),$$

with Bayesian estimation of coefficients $\mathbf{\Xi}$ and a data-matched resolution parameter ρ obtained from residual variance. Model uncertainty enters through the posterior over $\mathbf{\Xi}$, which we use to generate trajectory ensembles and error bars. The complete likelihood/prior specification, posterior formulas, and hyperparameter sweep are provided in Supplementary Section S3.

Flow fields, fixed points, and uncertainty visualization

We compared empirical finite-difference vectors with model-predicted dynamics in the $(\widetilde{A}, \widetilde{\delta K})$ plane by rendering (i) raw arrows at scan locations, (ii) an interpolated cohort field over the empirical convex hull, and (iii) the mean Z–SINDy vector field with streamline seeding proportional to data density. Fixed points of the mean ODEs were obtained by solving $\dot{\mathbf{x}}(\mathbf{x}) = \mathbf{0}$; local eigenvectors are plotted to indicate stability structure. Forecast bands reflect coefficient-posterior sampling (Supplementary Sections S4–S5).

Bayesian classification

We built two complementary classifiers: (i) a dynamic classifier that evaluates the likelihood of observed derivatives under each class-specific Z–SINDy model, and (ii) a static classifier that evaluates the likelihood of observed states against time-indexed simulated ensembles. In all analyses we used equal class priors $p(c) = 1/2$ and declared a confident decision when $\max_c p(c \mid \cdot) \geq 0.9$; the corresponding likelihoods and normalization are detailed in Supplementary Section S6.

Robustness to clinical sparsity: noise and sampling inhomogeneity stress tests

To emulate real-world data degradation, we applied controlled corruptions to the resampled trajectories before classification. First, we injected spatial Gaussian noise with a standard deviation set to a fixed percentage of each feature's magnitude (e.g., 10% and 25% of $\widetilde{A}$ and $\widetilde{\delta K}$), independently at each time point. Second, we used either regular temporal sampling (e.g., annual intervals) or random time increments within bounds (e.g., 0.5–1.5 years),

thereby reducing effective time resolution to stress the dynamic classifier. After corruption, we re-estimated derivatives where applicable, reapplied the dynamic classifier, and quantified shifts in posterior class probabilities, directly assessing sensitivity to segmentation/measurement noise and to attrition typical of clinical datasets.

Author contributions

Conceptualization: L.P., A.K., J.P., J.K. Data Curation: J.K., J.P., M.M., N.N., C.L., H.V., R.M., L.P. Formal Analysis: J.P., M.M. Funding Acquisition: L.P. Investigation: J.P. Methodology: L.P., A.K. Project Administration: L.P. Resources: L.P. Supervision: L.P., A.K. Validation: J.P. Writing - Original Draft: J.P., M.M. Writing - Review & Editing: L.P., A.K., J.P., M.M.

Funding

This study was funded by the National Institutes of Health, USA, NHLBI Grant R01-HL159205 (to L.P.)

References

[1] R. M. Greenhalgh et al. "Endovascular versus open repair of abdominal aortic aneurysm". In: *New England Journal of Medicine* 362 (2010), pp. 1863–1871.

[2] J. L. De Bruin et al. "Long-Term outcome of open or endovascular repair of abdominal aortic aneurysm". In: *New England Journal of Medicine* 362 (2010), pp. 1881–1889.

[3] N. Pandey et al. "Surveillance imaging following endovascular aneurysm repair". In: *Seminars in Interventional Radiology* 32 (2015), pp. 239–247.

[4] T. F. X. O'Donnell et al. "Aneurysm sac failure to regress after endovascular aneurysm repair is associated with lower long-term survival". In: *Journal of Vascular Surgery* 69 (2019), pp. 414–422.

[5] S. E. Deery et al. "Aneurysm sac expansion is independently associated with late mortality in patients treated with endovascular aneurysm repair". In: *Journal of Vascular Surgery* 67 (2018), pp. 157–164.

[6] G. J. Boer et al. "A stable aneurysm sac after endovascular aneurysm repair as a predictor for mortality: An in-depth analysis". In: *Journal of Vascular Surgery* 76 (2022), pp. 445–453.

[7] M. Franchin et al. "A morphovolumetric analysis of aneurysm sac evolution after elective endovascular abdominal aortic repair". In: *Journal of Vascular Surgery* 74 (2021), pp. 1423–1432.

[8] T. Smith and K. B. Quencer. "Best practice guidelines: imaging surveillance after endovascular aneurysm repair". In: *American Journal of Roentgenology* 214 (2020), pp. 1220–1230.

[9] F. Zaiem et al. "A systematic review of surveillance after endovascular aortic repair: imaging modality and frequency". In: *Journal of Vascular Surgery* 67 (2018), 320–331.e37.

[10] K. Khabaz et al. "The geometric evolution of aortic dissections: Predicting surgical success using fluctuations in integrated Gaussian curvature". In: *PLOS Computational Biology* 20 (2024), e1011815.

[11] E. L. Chaikof et al. "The Society for Vascular Surgery practice guidelines on the care of patients with an abdominal aortic aneurysm". In: *Journal of Vascular Surgery* 67 (2018), 2–77.e2.

[12] S. Ikeda et al. "One-year sac regression is associated with freedom from fatal adverse events after endovascular aneurysm repair". In: *Journal of Vascular Surgery* 77 (2023), 136–142.e2.

[13] G. Tinelli et al. "The sac evolution imaging follow-up after endovascular aneurysm repair". In: *Journal of Vascular Surgery* 80 (2024), pp. 937–945.

[14] W. Zhou et al. "Outcome and clinical significance of delayed endoleaks after endovascular aneurysm repair". In: *Journal of Vascular Surgery* 59 (2014), pp. 915–920.

[15] J. E. Jones et al. "Persistent type 2 endoleak after endovascular repair of abdominal aortic aneurysm is associated with adverse late outcomes". In: *Journal of Vascular Surgery* 46 (2007), pp. 1–8.

[16] E. Andraska et al. "Longer follow-up intervals following EVAR are safe and appropriate after marked aneurysm sac regression". In: *Journal of Vascular Surgery* 76 (2022), pp. 454–460.

[17] P. Jetty et al. "Variability in aneurysm sac regression after endovascular aneurysm repair based on a comprehensive registry of patients in Eastern Ontario". In: *Journal of Vascular Surgery* 70 (2019), pp. 1469–1478.

[18] V. Rastogi et al. "Aneurysm sac dynamics and its prognostic significance following fenestrated and branched endovascular aortic aneurysm repair". In: *European Journal of Vascular and Endovascular Surgery* 67 (2024), pp. 728–736.

[19] H. L. Marecki et al. "Characterization and management of type II and complex endoleaks after fenestrated/branched endovascular aneurysm repair". In: *Journal of Vascular Surgery* 78 (2023), pp. 29–37.

[20] Y. Rubanova, R. T. Q. Chen, and D. Duvenaud. "Latent ordinary differential equations for irregularly-sampled time series". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2019.

[21] F. Doshi-Velez and B. Kim. "Towards a rigorous science of interpretable machine learning". In: *arXiv:1702.08608* (2017).

[22] C. Rudin. "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead". In: *Nature Machine Intelligence* 1 (2019), pp. 206–215.

[23] C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning*. Cambridge, MA: MIT Press, 2006.

[24] P. Kidger et al. "Neural controlled differential equations for irregular time series". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2020.

[25] S. L. Brunton, J. L. Proctor, and J. N. Kutz. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems". In: *Proceedings of the National Academy of Sciences* 113 (2016), pp. 3932–3937.

[26] S. H. Rudy et al. "Data-driven discovery of partial differential equations". In: *Science Advances* 3 (2017), e1602614.

[27] B. M. De Silva et al. "Discovery of physics from data: Universal laws and discrepancies". In: *Frontiers in Artificial Intelligence* 3 (2020), p. 25.

[28] M. S. Schmitt et al. "Machine learning interpretable models of cell mechanics from protein images". In: *Cell* 187 (2024), pp. 481–494.

[29] R. Lyons, V. Dukic, and D. M. Bortz. "Learning structured population models from data with WSINDy". In: *PLOS Computational Biology* 21 (2025), e1013742.

[30] A. A. Klishin et al. "Statistical mechanics of dynamical system identification". In: *Physical Review Research* 7 (2025), p. 033181.

[31] K. Khabaz et al. "Temporal geometric mapping defines morphoelastic growth model of type B aortic dissection evolution". In: *Computers in Biology and Medicine* 182 (2024), p. 109194.

[32] J. N. Kutz and S. L. Brunton. "Parsimony as the ultimate regularizer for physics-informed machine learning". In: *Nonlinear Dynamics* 107 (2022), pp. 1801–1817.

[33] C. W. Hicks et al. "Endovascular aneurysm repair patients who are lost to follow-up have worse outcomes". In: *Journal of Vascular Surgery* 65.6 (2017), pp. 1625–1635.

[34] J. Schutt et al. "Social determinants of health factors and loss-to-follow-up in the field of vascular surgery". In: *Annals of Vascular Surgery* 105 (2024), pp. 316–324.

[35] P. Jaakkola et al. "Interobserver variability in measuring the dimensions of the abdominal aorta: Comparison of ultrasound and computed tomography". In: *European Journal of Vascular and Endovascular Surgery* 12 (1996), pp. 230–237.

[36] N. S. Cayne et al. "Variability of maximal abdominal aortic aneurysm diameter measurements on computed tomography: implications for clinical decision making". In: *Journal of Vascular Surgery* 39 (2004), pp. 811–815.

[37] J. R. Leach et al. "Abdominal aortic aneurysm measurement at CT/MRI: potential clinical ramifications of non-standardized measurement technique and importance of multiplanar reformation". In: *Quantitative Imaging in Medicine and Surgery* 11 (2021), p. 823.

[38] T. Garg, L. C. Baker, and M. W. Mell. "Adherence to postoperative surveillance guidelines after endovascular aortic aneurysm repair among medicare beneficiaries". In: *Journal of Vascular Surgery* 61 (2015), pp. 23–27.

[39] A. Schanzer et al. "Follow-up compliance after endovascular abdominal aortic aneurysm repair in medicare beneficiaries". In: *Journal of Vascular Surgery* 61 (2015), 16–22.e1.

[40] G. A. Antoniou et al. "Meta-analysis of compliance with endovascular aneurysm repair surveillance: The EVAR surveillance paradox". In: *European Journal of Vascular and Endovascular Surgery* 65 (2023), pp. 244–254.

[41] L. E. Newton, A. Ponukumati, and G. Zwain. "Imaging surveillance adherence after endovascular abdominal aortic aneurysm repair at VA hospitals". In: *JAMA Network Open* 8 (2025).

[42] J. Bakarji and A. A. Klishin. *ZSINDy*. 2024.

[43] T. J. Willmore. *Total curvature in riemannian geometry*. Ellis Horwood, 1982.

[44] R. Chartrand. "Numerical differentiation of noisy, nonsmooth data". In: *International Scholarly Research Notices* 2011 (2011).

[45] A. Cortiella, K.-C. Park, and A. Doostan. "A priori denoising strategies for sparse identification of nonlinear dynamical systems: A comparative study". In: *Journal of Computing and Information Science in Engineering* 23 (2023), p. 011004.

[46] J. M. Hokanson, G. Iaccarino, and A. Doostan. "Simultaneous identification and denoising of dynamical systems". In: *SIAM Journal on Scientific Computing* 45 (2023), A1413–A1437.

[47] L. Fung. "Overcoming error-in-variable problem in data-driven model discovery by orthogonal distance regression". In: *arXiv:2507.23426* (2025).

[48] J. A. Pugar et al. "Thoracic aortic ahape: A data-driven scale space approach". In: *medRxiv* (2025).

Supplemental Information

In addition to this additional information, the code used for data processing, model construction, and analysis is publicly available at https://github.com/SurgBioMech/aaa-dynamics. The anatomic feature data (per-patient, per-CT) supporting the findings of this study is available from the same repository.

S1. FEA-based up-sampling and polynomial resampling

To address irregular clinical follow-up and inherent sparsity, we upsampled trajectories using a two-step procedure. Using finite-element analysis (FEA) simulations, we first constructed dense surrogate trajectories for A and δK that respect sac mechanics (see prior work for details [31]). These sequences furnish physically plausible intermediate states between clinical scans without extrapolating beyond the postoperative ground truth $\mathbf{x}(t_f)$. We then resampled each subject's trajectory by fitting a third-order polynomial to each feature as a function of time and evaluating the fit at fixed increments of $\Delta t = 0.26$ and 0.14 years. These increments were chosen by taking the mean inter-scan intervals in the regressing (2.6 yr) and stable (1.4 yr) datasets and dividing each by 10. If this procedure generated more points than a preset per-subject limit, we instead produced exactly $N = 30$ evenly spaced samples between the first and last observed scan. This soft cap prevents oversampling of long trajectories while preserving short-timescale variation. The resulting dataset is sufficiently upsampled and evenly spaced for stable derivative estimation and dynamical modeling, while mitigating variance stemming from irregular clinical follow-up and patient attrition.

S2. Derivative estimation and outlier handling

We estimated time derivatives using pairwise finite differences between successive observations within subject. To avoid spuriously large derivatives from nearly coincident scans, we required a minimum time separation of 3 months. Let $(t_i,\, \widetilde{A}_i,\, \widetilde{\delta K}_i)_{i=1}^{N}$ denote the ordered time points and corresponding normalized features for a subject. To estimate instantaneous rates of change $(\dot{\widetilde{A}}, \dot{\widetilde{\delta K}})$, we applied second-order central differences to the compressed time

series:

$$\dot{\widetilde{A}}_i \;=\; \frac{\widetilde{A}_{i+1}-\widetilde{A}_{i-1}}{t_{i+1}-t_{i-1}}, \quad \dot{\widetilde{\delta K}}_i \;=\; \frac{\widetilde{\delta K}_{i+1}-\widetilde{\delta K}_{i-1}}{t_{i+1}-t_{i-1}}, \quad \text{for } i=2,\ldots,N-1. \tag{1}$$

We also applied light, physically motivated outlier rejection on $\mathrm{d}\,\widetilde{\delta K}/\mathrm{d}t$ rates to remove biologically implausible spikes introduced by algorithmic error (segmentation, curvature estimation, or timestamps). Specifically, if

$$\frac{\mathrm{d}\,\widetilde{\delta K}}{\mathrm{d}t} \;>\; 6\ \mathrm{yr}^{-1} \quad \text{or} \quad <\; -3\ \mathrm{yr}^{-1},$$

the offending scans were removed from the dataset, and the remaining data were then used to fit the dynamical model.

S3. Z–SINDy inference details, hyperparameters, and linear-library justification

We model the temporal evolution of the state $\mathbf{x}(t) = [\widetilde{A}(t),\, \widetilde{\delta K}(t)]^\top$ as $\dot{\mathbf{x}}(t) \approx \boldsymbol{\Xi}^\top \boldsymbol{\Theta}(\mathbf{x}(t))$, where in the main text we restrict the library to the affine form $\boldsymbol{\Theta}(\mathbf{x}) = [1, \widetilde{A}, \widetilde{\delta K}]^\top$ (alternative higher-order libraries are explored here for completeness). Parameters $\boldsymbol{\Xi}$ are inferred in a Bayesian framework with a Gaussian likelihood that penalizes the squared residuals between observed derivatives and model predictions,

$$p(\{\dot{\mathbf{x}}\} \mid \boldsymbol{\Xi}) \propto \prod_l \exp\left[-\frac{1}{2\rho^2}\sum_k \big(\dot{x}_{l,k} - \boldsymbol{\Xi}_l^\top \boldsymbol{\Theta}(\mathbf{x}_k)\big)^2\right], \tag{2}$$

and a sparsity-inducing Bernoulli–Gaussian prior over active-set indicators γ, $p(\boldsymbol{\Xi}_l) \propto \prod_i \big(\delta(\Xi_{l,i}) + w(\Xi_{l,i})\, e^{-\lambda n}\big)$, with λ controlling parsimony. Conditioning on a fixed active set γ yields a multivariate Gaussian posterior for the coefficients,

$$\boldsymbol{\Xi}_l \mid \{\dot{\mathbf{x}}\} \sim \mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}), \qquad \boldsymbol{\mu} = \mathbf{C}^{-1}\mathbf{V}, \quad \boldsymbol{\Sigma} = \rho^2\, \mathbf{C}^{-1}, \tag{3}$$

where $\mathbf{C} = \sum_k \boldsymbol{\Theta}(\mathbf{x}_k)\boldsymbol{\Theta}(\mathbf{x}_k)^\top$ and $\mathbf{V} = \sum_k \boldsymbol{\Theta}(\mathbf{x}_k)\dot{\mathbf{x}}_k^\top$; the full posterior is a Gaussian mixture $\sum_\gamma p(\boldsymbol{\Xi} \mid \gamma)\, \mathcal{Z}_\gamma$ typically dominated by a single γ. We estimate the resolution parameter from residual variance, $\rho \approx \sqrt{(dn)^{-1} \sum_{k,l} (\dot{x}_{l,k} - \boldsymbol{\mu}_l^\top \boldsymbol{\Theta}(\mathbf{x}_k))^2}$, and, where relevant, sweep the regularization weight λ to balance parsimony and fit by explained variance. This inference supports uncertainty propagation by sampling $\boldsymbol{\Xi}^{(m)} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to generate trajectory ensembles and confidence bands, and by analyzing fixed points/eigenstructure of the mean model (procedures detailed in Supplementary Methods S4–S5). Crucially, the hyperparameter sweep in Fig. 7 shows that the linear (affine) library already captures the bulk of explainable variance for both cohorts across the central range of λ, whereas quadratic/cubic libraries offer negligible gains and may reduce R^2 at extreme sparsity—motivating our Occam-optimal choice of a linear library in the main analyses. Fig. 8 demonstrates the vector fields of higher-order models that reproduce the key dynamical features of the linear models. Complementing this fit–complexity tradeoff, Fig. 9 demonstrates coefficient reproducibility under randomized train splits: split-wise means concentrate near the across-split mean with narrow one–standard–deviation bars relative to coefficient magnitudes. Together,

Figs. 7–9 justify restricting $\mathbf{\Theta}$ to linear terms and indicate that the inferred dynamical functions are well constrained by the data and robust to moderate perturbations of the training set.

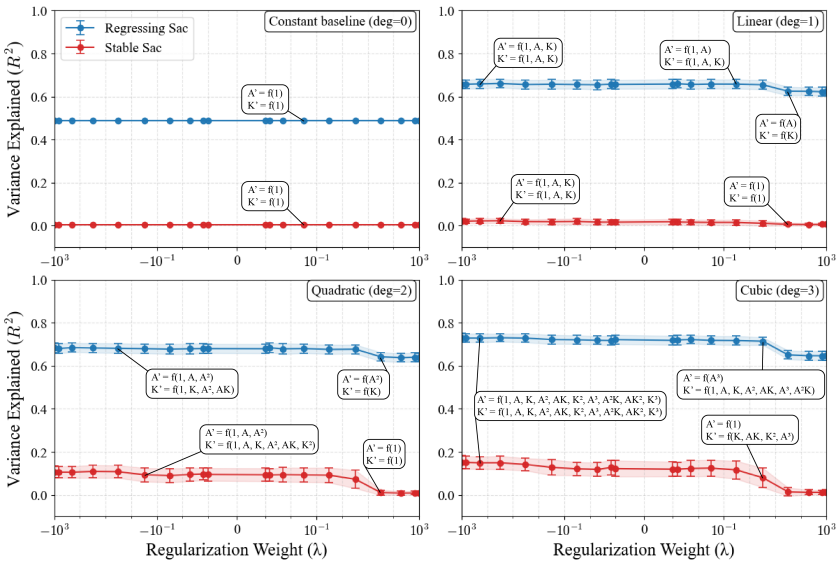


Figure 7: Hyperparameter sweep for Z–SINDy model fit. Mean variance explained (R^2) ± 1 SD versus regularization weight λ for four polynomial libraries of increasing maximal degree: constant (deg= 0), linear (deg= 1), quadratic (deg= 2), and cubic (deg= 3). Blue curves: regressing sac cohort; red curves: stable sac cohort. For each degree and λ, Z–SINDy is fit to $(\widetilde{A}, \widetilde{\delta K})$ and their time derivatives using the cohort-specific resolution parameter ρ. Uncertainty bands are obtained by sampling from the coefficient covariance of the Gaussian posterior. The x-axis uses a logarithmic scale to show the hyperparameter axis about $\lambda = 0$. Call outs denote representative active term sets in the learned right-hand sides (e.g., $d\widetilde{A}/dt = f(1, \widetilde{A}, \widetilde{\delta K})$). Relative to the constant baseline, linear models already capture most explainable variance for the regressing cohort, while higher-order terms tend to overfit the derivative training data.

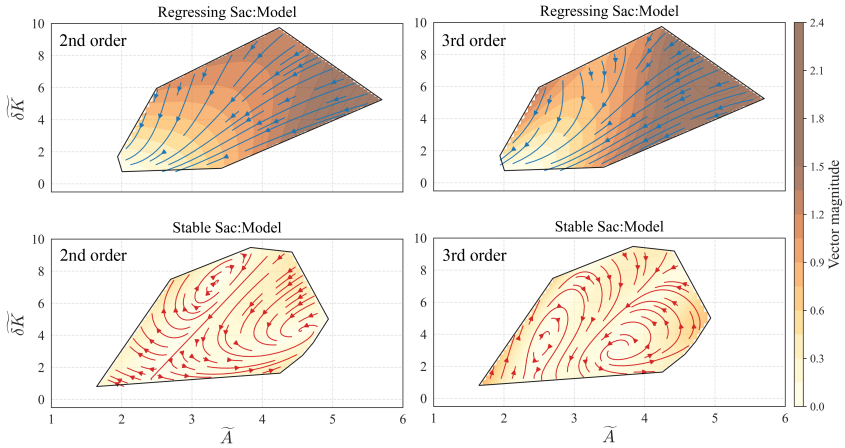


Figure 8: Z–SINDy vector fields with higher-order polynomial libraries. Mean cohort-level flow fields in the normalized size–shape space ($\widetilde{A}$, $\widetilde{\delta K}$) for regressing (top row) and stable (bottom row) sacs, analogous to the *Model* panels of Fig. 4 but using polynomial libraries up to second- and third-order. Left column: quadratic library (up to 2nd order); right column: cubic extension (up to 3rd order). Streamlines indicate characteristic trajectories and background shading encodes vector magnitude. Increasing the polynomial degree modestly warps streamlines and local speeds but preserves the qualitative flow structure observed for the affine (main-text) models—faster convergence toward a low-size/low-shape region for regressing sacs and slower recirculation within a higher-$\widetilde{\delta K}$ band for stable sacs—demonstrating that the main dynamical conclusions are robust to reasonable expansions of the function library.

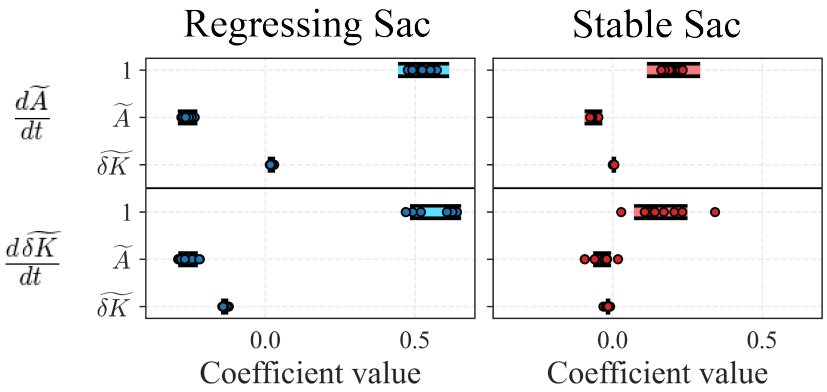


Figure 9: Coefficient distributions for the affine Z–SINDy model. Aggregated coefficient summaries for the linear (affine) library $\Theta(x) = [1, \widetilde{A}, \widetilde{\delta K}]$ in each differential equation (top: $d\widetilde{A}/dt$; bottom: $d\,\widetilde{\delta K}/dt$). Left panel: regressing sac model (blue); right panel: stable sac model (red). The dataset is repeatedly subsampled into randomized train splits; the model is re-fit on each split; and the resulting coefficients are summarized at the term level. Small circles show split-wise means; thick colored horizontal bars show mean ± one standard deviation across splits, overlaid on a thin black bar marking the same interval. This compact, uncertainty-aware view reveals which contributions (constant, $\widetilde{A}$, $\widetilde{\delta K}$) are reproducibly nonzero in each equation and how their magnitudes differ between cohorts.

S4. Uncertainty-aware simulation and fixed-point analysis

We quantified forecast uncertainty by generating ensembles via sampling model coefficients from the posterior $\boldsymbol{\Xi}^{(m)} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and initial conditions from a Gaussian around a preoperative baseline. Each sampled model was integrated forward to obtain trajectories $\mathbf{x}^{(m)}(t)$. We visualized uncertainty with shaded bands about the mean trajectory for each class (*regressing* and *stable*) for $\mathbf{x}(t)$ and for $\dot{\mathbf{x}}(t) = \boldsymbol{\Xi}^{(m)\top}\boldsymbol{\Theta}(\mathbf{x}(t))$. Fixed points $\mathbf{x}^*$ were computed from the mean model by solving $\dot{\mathbf{x}}(\mathbf{x}) = 0$ and are plotted as horizontal lines in the respective panels.

S5. Population flow-field visualization in morphological space

To compare empirical patient trajectories with model-predicted dynamics, we visualized cohort-level flow fields in the $(\widetilde{A}, \widetilde{\delta K})$ plane. For the original CT-derived data (labeled *Raw*), we computed the convex hull of all observed states and overlaid finite-difference estimates of $\left(\mathrm{d}\,\widetilde{A}/\mathrm{d}t,\ \mathrm{d}\,\widetilde{\delta K}/\mathrm{d}t\right)$ as arrows anchored at the corresponding scan locations. Next, these derivative estimates were pooled across patients and interpolated onto a regular grid spanning the convex hull, with the background magnitude $\|\dot{\mathbf{x}}\|$ rendered to indicate local dynamic speed (labeled *Interpolated*). The interpolation was performed with radial basis-functions (RBF) with the smoothing parameter tuned to the plateau of variance explained (Fig. 10). Streamlines were seeded using a kernel-density estimate over the empirical point cloud so that streamline density reflects regions with data support. Lastly, we evaluated the mean Z–SINDy vector field for each class, rendering streamlines and shading by vector magnitude (labeled *Model*). Fixed points from the mean ODEs are marked with stars, and associated eigenvectors are shown as black lines to indicate local stability structure.

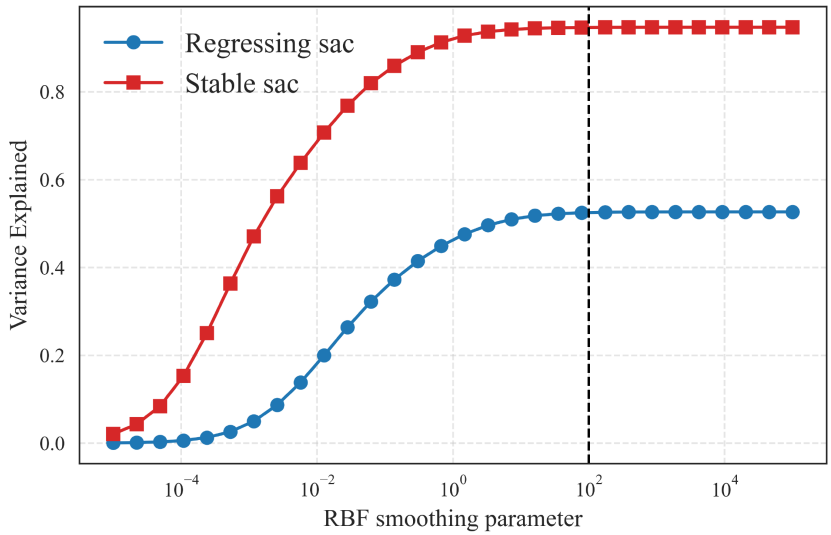


Figure 10: Tuning the smoothing scale for empirical vector fields. Normalized root-mean-square error (RMSE) between the radial basis–function (RBF) smoothed flow field and the raw finite-difference derivatives as a function of the RBF smoothing parameter for the regressing (blue circles) and stable (red squares) cohorts. Errors are normalized so that values near zero indicate close agreement with the unsmoothed empirical vectors. As smoothing increases, the RMSE rises and then plateaus; the dashed vertical line at 10^2 marks the smoothing parameter used for the interpolated flow fields in Fig. 4. Choosing this value lies at the onset of the plateau, providing visual regularization of the empirical field without appreciably degrading fidelity to the original derivative data.

S6. Bayesian classifiers: full likelihoods and normalization

The Bayesian classifiers take per-patient observations and convert them into posterior probabilities for the two outcome classes, $c \in \{\text{regressing}, \text{stable}\}$. Throughout this section we denote the state by $\mathbf{x}_k = [\widetilde{A}_k, \widetilde{\delta K}_k]^\top$ at time t_k and, where available, its finite-difference derivative $\dot{\mathbf{x}}_k$. Class-specific Z–SINDy models are summarized by coefficient matrices $\boldsymbol{\Xi}_c$ and resolution parameters ρ_c (Sec. S3). Unless otherwise noted, we use equal class priors $p(c) = 1/2$; attrition-aware variants with time-dependent priors are described in Sec. S8.

Dynamic classifier. The dynamic classifier evaluates how likely the observed derivatives $\{\dot{\mathbf{x}}_k\}$ are under each class-specific Z–SINDy model. Conditioning on class c, we assume independent Gaussian errors with variance ρ_c^2 about the model-predicted derivatives $\boldsymbol{\Xi}_c^\top \boldsymbol{\Theta}(\mathbf{x}_k)$, where $\boldsymbol{\Theta}$ is the chosen library:

$$z_c = p(\{\dot{\mathbf{x}}_k\} \mid c) = \prod_{k,\ell} \frac{1}{(2\pi\rho_c^2)^{1/2}} \exp\left[-\frac{1}{2\rho_c^2}\left(\dot{x}_{k,\ell} - \boldsymbol{\Xi}_{c,\ell}^\top \boldsymbol{\Theta}(\mathbf{x}_k)\right)^2\right].$$

Bayes' rule then yields the posterior class probability

$$p(c \mid \{\dot{\mathbf{x}}_k\}) = \frac{z_c\, p(c)}{\sum_{c'} z_{c'}\, p(c')}, \qquad p(c) = \tfrac{1}{2}.$$

In words, the dynamic classifier compares how well each class-specific ODE explains the observed motion in size–shape space; patients are assigned to the class whose model best predicts their derivatives.

Static classifier. The static classifier, by contrast, operates on coordinates rather than derivatives. For each class c we generate a large ensemble of trajectories by sampling initial conditions and coefficients from the class-specific Z–SINDy posterior and integrating forward in time (Sec. S4). These ensembles are used to estimate time-indexed densities $p(\mathbf{x} \mid c, t)$ over the state space. Given a patient's observations $\{(\mathbf{x}_k, t_k)\}$, we assume conditional independence across time and compute the class likelihood

$$z_c = \prod_k p(\mathbf{x}_k \mid c, t_k), \qquad p(c \mid \{\mathbf{x}_k, t_k\}) = \frac{z_c\, p(c)}{\sum_{c'} z_{c'}\, p(c')}.$$

The static classifier therefore reflects where a patient is relative to the simulated class ensembles at the same follow-up times, whereas the dynamic classifier reflects where the patient is according to the class-specific ODEs.

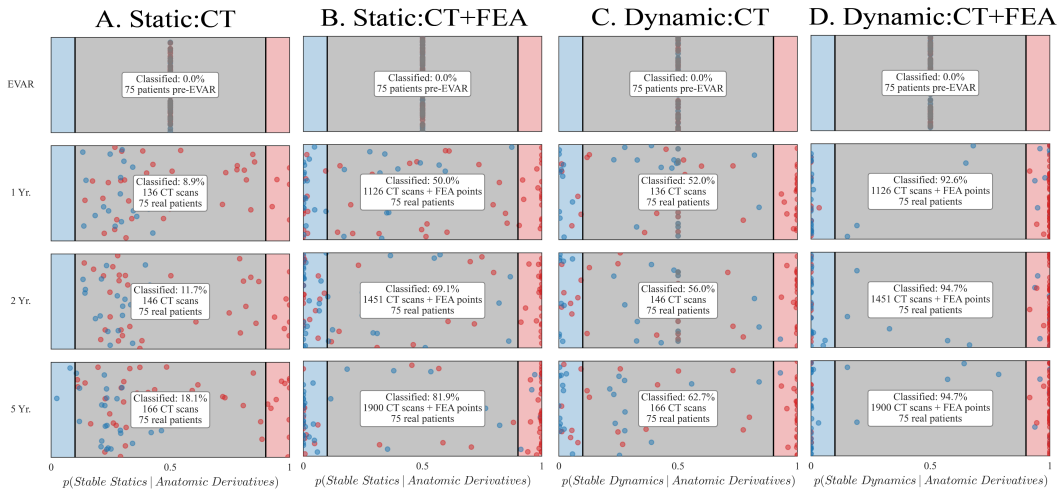


Figure 11: Posterior probability jitter plots for static and dynamic Bayesian classifiers across both real patient CT scans and synthetic data from the Z–SINDy models. Columns show (from left to right): A. Static classifier with CT data, B. Static classifier with synthetic data, C. Dynamic classifier with CT data, and D. Dynamic classifier with synthetic data. Rows correspond to EVAR (preoperative baseline) and 1, 2, and 5 years post-EVAR. The horizontal axis in each panel is the posterior probability of the stable class, either $p(\text{Stable Statics} \mid \text{anatomic derivatives})$ or $p(\text{Stable Dynamics} \mid \text{anatomic derivatives})$; blue and red points represent regressing and stable sacs, respectively. Vertical side bands at $p = [0.1, 0.9]$ indicate decision regions for each class, and the central grey band marks the "uncertain" region in which neither class exceeds the confidence threshold. In-panel boxes report the percentage of scans confidently classified, together with the number of scans and real or synthetic patients contributing to each panel.

Decision rule and performance metrics. For both classifiers we declare a confident decision when the posterior for one class exceeds a fixed probability threshold, $\max_c p(c \mid \cdot) \geq 0.9$; otherwise the scan is labeled "uncertain" and excluded from accuracy calculations.

The fraction of scans meeting this criterion is reported as "Classified (%)", and the accuracy is computed over this subset as the percentage whose predicted class matches the clinical label. Figure 11 shows the distribution of posterior probabilities for each classifier and dataset (CT only versus the FEA upsampled datasets), with the central grey band marking the uncertain region $0.1 < p(\text{Stable} \mid \cdot) < 0.9$. The cross-validated summary in Table 1 quantifies these patterns: the dynamic classifier confidently labels a much larger fraction of CT-only scans at early times than the static classifier, and the introduction of upsampled data from FEA simulations further increases both the fraction classified and the resulting accuracy, especially at 1–2 years post-EVAR.

Table 1: Cross-validated performance of dynamic and static Bayesian classifiers. Within each of five folds, Z–SINDy models are trained on random 80% of patients and classification is applied to the remaining 20%. Entries are mean±std across folds.

Classifier	Dataset	Time (yr)	Classified (%)	Accuracy (%)
Static	CT	1	10 ± 10	100 ± 0
		2	12 ± 9	100 ± 0
		5	19 ± 17	100 ± 0
	CT+FEA	1	50 ± 6	77 ± 18
		2	69 ± 13	74 ± 16
		5	82 ± 7	76 ± 11
Dynamic	CT	1	52 ± 12	88 ± 13
		2	56 ± 12	82 ± 13
		5	63 ± 11	83 ± 12
	CT+FEA	1	93 ± 5	80 ± 14
		2	95 ± 4	81 ± 13
		5	95 ± 6	82 ± 12

S7. Stress-test protocols (noise and sampling inhomogeneity)

We evaluate robustness by corrupting the resampled series prior to classification: (i) add independent Gaussian noise at each time point with SD equal to a fixed fraction of each feature's magnitude (e.g., 10%, 25% for $\widetilde{A}$, $\widetilde{\delta K}$); (ii) impose either regular sampling (e.g., annual) or random increments in $[0.5, 1.5]$ years to reduce temporal resolution. We then recompute derivatives (for the dynamic classifier), re-estimate posteriors, and quantify probability shifts.

S8. Attrition-aware Bayesian classification

To quantify how loss to follow-up influences outcome prediction, we incorporated the empirically observed patient attrition rates for each clinical cohort into both the static and dynamic Bayesian classifiers. Fig. 12 summarizes the declining proportions of stable and regressing patients remaining at each postoperative time point following EVAR. For each class, the Bayesian classifiers evaluated likelihoods using time-dependent priors equal to the linearly interpolated fraction of patients still under surveillance.

To emulate the availability of real clinical data, ensembles of simulated trajectories were additionally subjected to randomized dropout consistent with the same patient attrition

curves. Therefore, both the effective class priors and the number of trajectory realizations contributing to posterior estimates changed over time in a manner that reflects real-world surveillance behavior. Fig. 13 shows posterior probabilities for the attrition-aware classifiers. The dynamic classifier once again exits the "uncertain" region earlier than the static classifier. As shown in Fig. 5, at 5 years following EVAR (with no awareness of patient dropout), the static model classifies a similar percentage of trajectories compared to the dynamic model. However, the observed attrition rates show that by this time, there are less than 30% of the regressing cohort remaining and less than 5% of the stable cohort remaining. This highlights the significance of using morphologic dynamics to stratify patients at times earlier than static measures are able to inform clinical decision-making.

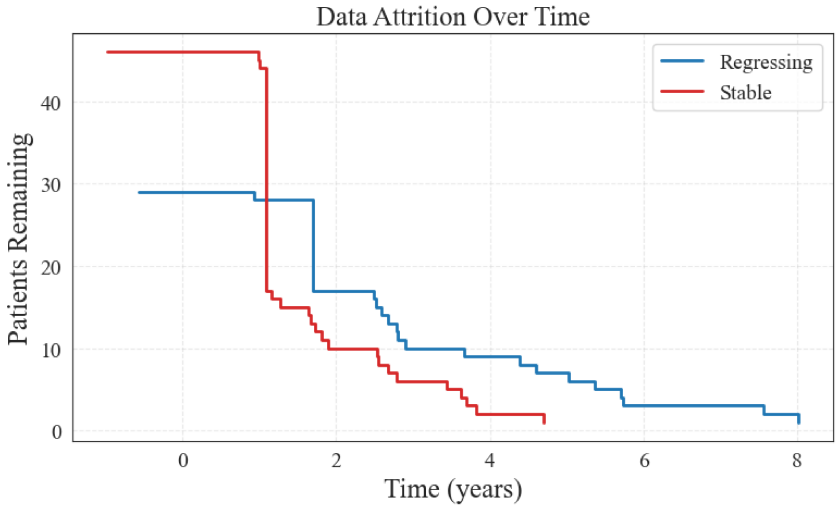


Figure 12: Quantities of regressing and stable patients in the time following EVAR (time=0). The differing rates of attrition for each class were incorporated into the stable and dynamic Bayesian classifiers, informing the models of clinically-observed patient distributions and loss to follow-up.

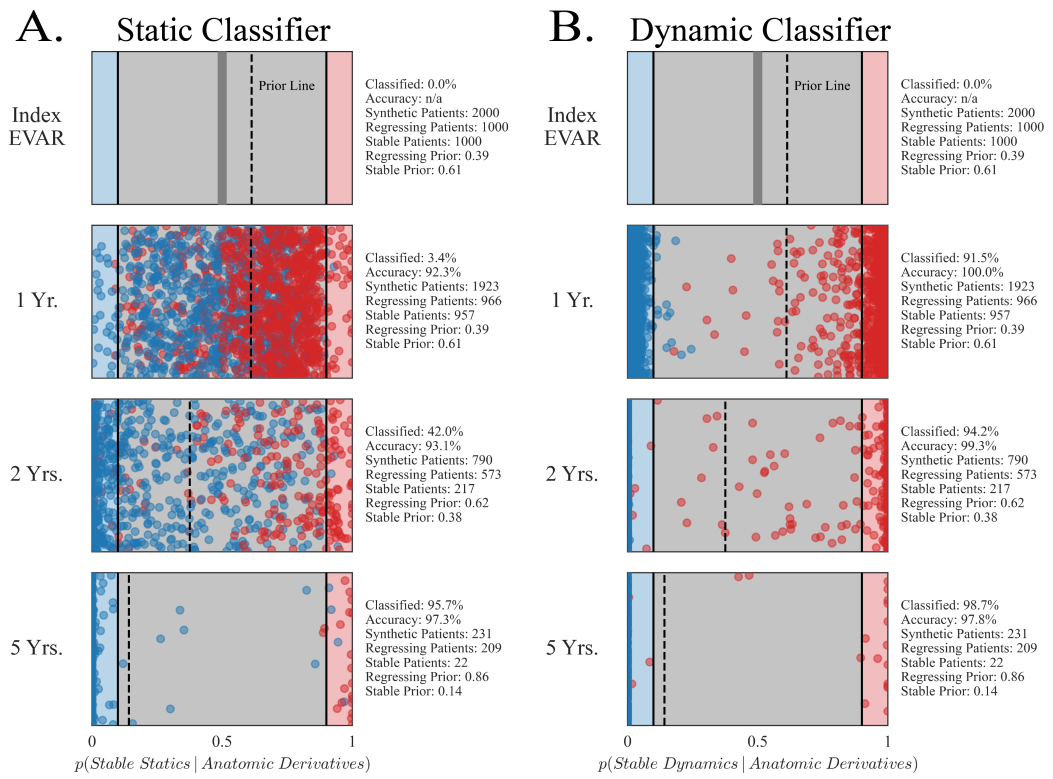


Figure 13: Posterior probabilities for static (A.) and dynamic (B.) Bayesian classifiers determined from simulated regressing and stable patient trajectories that proportionally follow the clinically-observed patient attrition rates shown observed in Fig. 12. Simulated trajectories experience randomized dropout corresponding to the fraction of patients remaining at that time point for each class, and the classifier priors are set to the proportion of remaining patients within each class. The dynamic classifier once again exits the "uncertain" region earlier than the static classifier, prior to the loss of the majority of patients, reflecting the added value of directional information in a clinically representative dataset. By maintaining this representativeness in both models' priors and the available proportions of data, we demonstrate the improved translational validity of probabilistic forecasts derived from dynamical models as compared to static models. This approach underscores how patients lost to follow-up shape evidentiary depth and should be explicitly accounted for in analyses of clinical datasets.