Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A mechanistic digital twin of the corneal endothelium separates early postoperative and long-term endothelial cell loss after keratoplasty

Benton Chuter

Benton.Chuter@hitchcock.org

Dartmouth–Hitchcock Medical Center

Fae B. Kayarian

Dartmouth–Hitchcock Medical Center

Shad Hassan

Dartmouth College

Ammar Shaadab

Dartmouth College

Soham Shah

Dartmouth College

Bailey R. Millis

Dartmouth–Hitchcock Medical Center

Monica M. Jablonski

University of Tennessee Health Science Center

William G. Gensheimer

Geisel School of Medicine at Dartmouth

Article

Keywords:

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11570113/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A mechanistic digital twin of the corneal endothelium separates early postoperative and long-term endothelial cell loss after keratoplasty

Benton Chuter, MD, MS[1]* · Fae B. Kayarian, MD[1] · Shad Hassan[2] · Ammar Shaadab[2] · Soham Shah[2] · Bailey R. Millis[1] · Monica M. Jablonski, PhD[3] · William G. Gensheimer, MD[1, 4, 5]

[1] Section of Ophthalmology, Department of Surgery, Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA
[2] Dartmouth College, Hanover, NH, USA
[3] Department of Ophthalmology, Hamilton Eye Institute, The University of Tennessee Health Science Center, Memphis, TN, USA
[4] Geisel School of Medicine at Dartmouth, Hanover, NH, USA
[5] White River Junction VA Medical Center, White River Junction, VT, USA

*Corresponding author: Benton Chuter, MD, MS, One Medical Center Drive, Lebanon, NH 03756, USA. Benton.Chuter@hitchcock.org. ORCID 0000-0002-0354-0569.

Abstract

Post-keratoplasty endothelial cell density (ECD) is modeled with phenomenological decay curves that do not account for underlying cellular mechanisms or identify potential therapeutic targets. We built a two-phase (early, then long-term loss) agent-based mechanistic digital twin, calibrated to the Cornea Preservation Time Study (CPTS; 995 subjects, post-DSAEK) and tested in the Cornea Donor Study (CDS; 469 subjects, post-PKP). ECD identified a 21% early postoperative loss and a net ECD-loss rate of 0.0064 month^{-1}, but not proliferation, migration or guttae formation. The simulator matched held-out CPTS 48-month means (residuals −26 to +21 cells/mm²; constant-rate comparator −325 to −295); in CDS, per-eye calibration outperformed only carry-forward and the CPTS-calibrated cohort posterior. Decay-twin 80% intervals covered 79.6% (CPTS) and 82.3% (CDS); the image feeder's error averaged 150 cells/mm². A bounded in-silico ROCK-inhibitor counterfactual gave mean 36-month preservation from 94 (95% interval −4 to 280) to 266 (0 to 821) cells/mm² across drug targets.

Introduction

Corneal transplantation restores vision in conditions such as Fuchs endothelial corneal dystrophy and bullous keratopathy.[1,2] How long a graft survives depends on its endothelium, the single layer of cells that keeps the cornea transparent by pumping fluid out of the stroma; in humans these cells do not meaningfully proliferate, so cells lost after surgery are not replaced.[3–5] After keratoplasty, endothelial cell loss is steep in the early postoperative period and then slows to a steady rate, and when the endothelial cell density (ECD) falls too low to keep the cornea transparent, the graft decompensates.[6–8] Two landmark studies established the natural history of endothelial cell loss after keratoplasty. They are the Cornea Donor Study (CDS) after penetrating keratoplasty (PKP) and the Cornea Preservation Time Study (CPTS) after Descemet-stripping automated endothelial keratoplasty (DSAEK).[7,9–12] Both report ECD measured by a central specular-microscopy reading center, and both are now public.

The quantitative models built on these data are almost entirely phenomenological, being biexponential or log-linear decay curves, or mixed-effects slopes, fit to the ECD time series.[7,13]

Such models summarize what the density does, but they contain no cells, no cell-loss or migration events, and no place to insert a drug. They identify factors associated with cell loss but not a cell-level mechanism, so they cannot explain why one cohort loses cells faster or predict what an intervention that slowed cell loss would do. Rho-associated kinase (ROCK) inhibitors such as ripasudil and Y-27632 promote corneal endothelial cell migration and survival; ripasudil eye drops are in clinical use and trials, whereas Y-27632 is an experimental compound that has reached patients only as an adjunct to injected cultured endothelial cells.[14–17] A virtual-trial platform that could turn a candidate cell-protective mechanism into a quantitative, uncertainty-aware ECD trajectory would let investigators triage hypotheses before committing to scarce donor tissue and patient years.

Digital twins and in-silico trials address this problem at the mechanistic level. In a mechanistic twin of the endothelial cell layer, each process that removes or adds cells has its own rate, so a drug can be simulated by changing a rate and reading out the resulting ECD over time, with its uncertainty. Such a tool needs a mechanism that is identifiable from the available measurements, a forecast with calibrated uncertainty and an intervention that is bounded and falsifiable, and these properties are usually reported separately. The corneal endothelium is an unusually clean testbed for this discipline. It is a quasi-two-dimensional monolayer, its central density is measured by a standardized reading center, and a candidate pharmacology already exists whose several actions on the endothelium (promotion of adhesion, migration, and proliferation, and anti-apoptosis) we collapse, for modeling, into a reduction of net cell loss, the quantity ECD can identify. It therefore allows a mechanistic twin to be built and tested for where it fails.

Here we build such a platform for the post-keratoplasty endothelial monolayer (Fig. 1). An agent-based monolayer simulator with an early postoperative loss of cells at surgery and four rate constants is calibrated to CPTS by simulation-based inference,[18] tested per eye on held-out CDS visits and put through a formal identifiability analysis, which leaves net cell loss identifiable as two quantities, the early postoperative loss and a long-term net ECD-loss rate. A per-cohort decay-twin with split-conformal prediction intervals supplies per-patient forecasts, and its empirical coverage is reported against the nominal target. The per-cell loss rate and the early postoperative loss are then perturbed in a bounded ROCK-inhibitor counterfactual, a mechanistic sensitivity analysis. An image-to-ECD feeder supplies an image-derived input and is benchmarked against a published specular-ECD engine with released weights and on an external eye-bank image set. A graft-survival model with 40 events is reported as exploratory.

Results

Two harmonized post-keratoplasty cohorts

We assembled multi-visit, analyzable eyes from CPTS (post-DSAEK, n = 995 subjects) and CDS (post-PKP, n = 469 subjects), each with central-reading-center ECD at scheduled visits to 60 months and beyond (Table 1). The cohorts were demographically similar (recipient age 69.7 ± 8.7 versus 69.1 ± 8.7 years, p = 0.250; female 61.5% versus 61.8%, p = 0.905) while differing in surgery, providing a within-platform contrast between two keratoplasty modalities. Donor diabetes was modestly more frequent in CPTS (25.7% versus 20.3%; p = 0.022). Harmonization aligned the two reading centers' ECD fields, visit calendars (6, 12, 24, 36, 48, and 60 months), and donor/recipient covariate encodings onto a common schema, so that the simulator calibrated

on CPTS could be tested per eye on CDS and a single forecasting pipeline evaluated in both cohorts.

Three image analysis sets are used throughout. The image-to-ECD readout uses 2,297 specular images from 584 eyes (one frame per visit) with matched reading-center endothelial cell density (Table 2); two frames duplicate other frames (one within a patient at two visits, one across two patients) and were retained. The morphometric validation uses 2,280 of these images (579 eyes), of which 2,252 tessellated at 2.0 μm per pixel. The cross-instrument benchmark attempted all 2,297 images and yielded 1,554 successful segmentations (67.7%) from 499 eyes.

A two-phase mechanistic simulator of endothelial loss

The monolayer simulator represents the endothelium as interacting cell agents (Fig. 2). A fraction of the donor cells, the early postoperative loss fraction, is removed once at surgery, and four monthly rate constants then govern per-cell loss, migration, proliferation, and guttae formation. Cells are initialized at the donor density and the simulated ECD is read out as surviving cells per effective area. We calibrated the five parameters by simulation-based inference (neural posterior estimation[19–21], 10,000 simulations per donor-age stratum of CPTS), targeting the mean of the reading-center ECD readings observed at 6, 12, 24, and 36 months. Averaged over 30 independent seeds of the estimator, the pooled posterior means were an early postoperative loss fraction of 0.211 (seed range 0.207 to 0.215), rising with donor age from 0.186 in the youngest stratum to 0.234 in the oldest, a per-cell loss rate of 0.0116 $month^{-1}$ (0.0113 to 0.0120), migration 0.105 $month^{-1}$ (0.102 to 0.108), proliferation 0.00263 $month^{-1}$ (0.00256 to 0.00272) and guttae formation 0.00264 $month^{-1}$ (0.00253 to 0.00270), from a donor density of 2,664.7 cells/mm^2, the mean of the three stratum baselines (Table 3); the third significant figure of each rate is seed-dependent.

Calibrated on the same targets without the early postoperative loss, a constant-rate comparator reproduced the scale of the cohort's ECD loss but not its shape (Supplementary Table S12). Its pooled mean ECD lay 401 cells/mm^2 above the observed mean at 6 months and 238 above it at 12 months, and 158 below it at 36 months, and its 80% posterior predictive interval contained the observed stratum mean in 19% of stratum, month and seed combinations. The two-phase model's pooled residuals at the four calibration months were +1, −18, −24 and +8 cells/mm^2. At the 48-month visit, which was not used in calibration, its residuals in the three donor-age strata were −26, −4 and +21 cells/mm^2 (129, 108 and 129 eyes), against −297, −295 and −325 for the constant-rate comparator, and the deviance information criterion, which penalizes the extra parameter, favored the two-phase model in all 30 seeds in every stratum. Its predictive intervals are wider than the sampling error of the observed means, which they contained in 98% of combinations at the nominal 80% level, and at 60 months, with 5 to 14 eyes per stratum, it under-predicted the observed means by 17 to 193 cells/mm^2.

Not all five parameters were learnable from ECD trajectories (Fig. 3). In an identifiability analysis with a fixed-reference rejection-ABC estimator, the early postoperative loss fraction and the per-cell loss rate contracted substantially from their priors (0.992 and 0.978) and were recovered from simulated data (Pearson r = 0.997 and 0.988), whereas migration, proliferation, and guttae formation remained prior-dominated (contraction 0.0038, 0.026, and 0.034). A variance-based sensitivity decomposition[22,23] of 36-month ECD attributed first-order Sobol indices of 0.084 to the early postoperative loss fraction and 0.764 to the per-cell loss rate, with

negligible contributions from the other three rates; at 6 months the early postoperative loss fraction dominated instead (first-order index 0.69). The two contracted parameters were separately constrained (posterior correlation −0.17). The per-cell loss rate correlated with proliferation (0.70) and guttae formation (0.74), both prior-dominated, so the data constrain the net rate, per-cell loss minus proliferation minus guttae formation (guttae formation is subtracted because density is counted over the area not covered by guttae, so an enlarging guttae area raises the measured density; 0.0064 $month^{-1}$, 95% interval 0.0053 to 0.0076), more sharply than the per-cell loss rate itself (95% interval 0.0074 to 0.0159 $month^{-1}$; Table 3). With proliferation and guttae formation pinned at the ends of their priors, the per-cell loss rate ranged from 0.0063 to 0.0167 $month^{-1}$ while the net rate stayed between 0.0063 and 0.0068 $month^{-1}$ (Supplementary Table S1). Simulation-based-calibration rank histograms[24] of the rejection-ABC estimator were uniform for all five parameters ($p \geq 0.175$; Supplementary Table S1), so its posteriors pass this check even where the data are uninformative. The neural posterior estimator behind the Table 3 means recovered the early postoperative loss fraction and the per-cell loss rate from simulated data ($r = 1.00$ and 0.99), but in simulation-based calibration its posteriors were too narrow for the per-cell loss rate and the three prior-dominated rates (Supplementary Table S1). Table 3 therefore gives its point estimates with seed ranges and no posterior interval, and the identifiability verdicts and the parameter uncertainty propagated into the counterfactual use the rejection-ABC posterior. Longitudinal ECD therefore identifies net cell loss as two quantities, an early postoperative loss and a net ECD-loss rate, and the remaining rates are reported as prior-dominated. To test whether measurements beyond ECD would make more of the rates identifiable, we repeated the inference with image-derived morphometric features (endothelial cell hexagonality and guttae area fraction, computed on the simulated monolayers) appended to the summary vector. This made exactly one further rate identifiable, guttae formation (prior-to-posterior contraction from 0.01 to 0.55 and recovery r from 0.06 to 0.94 under this multi-observable rejection-ABC analysis, whose 5,000-simulation reference table does not pass simulation-based calibration; Supplementary Table S1), while migration and proliferation stayed non-identifiable (contraction ≈0.00–0.01).

The per-cell loss rate also differed between the cohorts. With the two-phase model fitted to the mean of each cohort's observed readings at each visit from 6 to 60 months, starting from the donor density, the per-cell loss rate was 0.0215 $month^{-1}$ in CDS (penetrating keratoplasty, 308 eyes) and 0.0111 $month^{-1}$ in CPTS (endothelial keratoplasty, 1,002 eyes), a ratio of 1.94 (95% CI 1.64 to 2.32 over estimator seeds and eye resamples, without posterior-parameter uncertainty; Supplementary Table S12). In this CDS fit the early postoperative loss fraction was 0.008, at the lower bound of its prior (0.210 in CPTS; Supplementary Table S12), so the CDS fit did not separate an early postoperative loss from the per-cell loss rate. The ratio was 1.81 without the CPTS 60-month target (36 eyes), 1.66 in the eyes observed at every visit from 6 to 48 months and 2.00 with the start anchored at 6 months. The ratio of the identified net rate, per-cell loss minus proliferation minus guttae formation, was 2.86 (95% CI 2.13 to 4.28) and stayed between 2.56 and 2.89 with proliferation and guttae formation pinned at the ends of their priors, against 1.61 to 2.77 for the per-cell loss rate (means over 30 estimator seeds; the net-rate ratio exceeded 1 in all 30 seeds unpinned and in 29 of 30 under each pinned setting). Because the CDS donor density is an eye-bank count, we also started each cohort at its 6-month mean, after the early postoperative loss, which gave a similar contrast (Supplementary Table S12). Surgery type is confounded with cohort and era (CDS 2000 to 2012, CPTS 2012 to 2017), so this is a contrast between two cohorts rather than a procedure effect.

A decay-twin with calibrated prediction intervals

For per-patient forecasting we wrapped the cohorts in a race-blind ridge decay-twin that runs from routinely collected donor and recipient data, with split-conformal prediction intervals[25–28], using subject-grouped training, calibration and test sets (989 / 251 / 440 eyes in the first of five re-splits) that are patient-disjoint by construction with zero patient-identifier leakage (0 leaked identifiers across all five seeds; Supplementary Table S2). At the nominal 80% target, empirical coverage of held-out patients' observed visits, averaged over five patient-grouped re-splits, was 0.796 for CPTS (seed range 0.771 to 0.832, every seed inside the [0.75, 0.85] acceptance band) and 0.823 for CDS (0.773 to 0.862, four of five seeds inside the band; the fifth, 0.862, had an eye-clustered 95% CI of 0.821 to 0.901 on its 118 test eyes) (Fig. 4; Table 4). Held-out mean absolute error (MAE) on observed visits, averaged over the five re-splits, ranged from 435.2 cells/mm² at 6 months to 499.0 at 36 months for CPTS. For CDS the 6-month visit is the cohort's baseline-anchoring reading. The 6-month value is numerically identical to the seeded baseline in 464 of 469 eyes (98.9%), giving a last-value-carried-forward MAE of 1.3 cells/mm², so this horizon is an anchoring identity and is not scored; CDS forecasting MAE ran from 219.7 cells/mm² at 12 months to 397.8 at 36 months. Few CPTS test eyes were observed at 48 and 60 months (93 to 107 and 6 to 13 visits per re-split), so accuracy claims are scoped to 6 to 36 months. Under subject-grouped five-fold cross-validation over all eyes, scored on observed visits (CDS 6-month visit excluded), the twin lowered per-eye MAE relative to a biexponential decay baseline (439.1 versus 452.4 cells/mm², 3.0%; paired Wilcoxon $p = 5.7 \times 10^{-5}$), establishing the phenomenological curve as the comparator it improves upon; this gap survived patient-clustered bootstrapping (mean difference −13.4 cells/mm², 95% CI [−19.5, −7.3] over 2,000 resamples of 1,445 patients) and was clearer in CDS than in CPTS (CPTS alone $p = 0.027$). Across 50 re-splits of the same patient-grouped five-fold design the twin's MAE was lower in all 50, with $p < 0.05$ in all 50 (pooled mean difference −13.5 cells/mm², eye-clustered 95% CI −19.2 to −7.8; 3.0%), and CPTS alone reached $p < 0.05$ in 32 of 50 (Supplementary Table S3). Scored the same way, the twin had the lowest MAE of a full baseline ladder (last-value-carried-forward, constant-mean, log-polynomial mixed, linear mixed-effects random-slope and biexponential comparators) in all 7 scored cohort × horizon cells from 6 to 36 months; these are point comparisons, and the narrowest margin was 2.9 cells/mm² over last-value-carried-forward at CDS 12 months (Supplementary Table S3). On the five Table 4 re-splits the twin's gain over the biexponential was in the same direction in all five (mean 2.2%) but individually significant in one. The absolute gap (13.4 cells/mm²) is small relative to the 439.1 cells/mm² MAE, so its clinical materiality should be read against an independently sourced measurement-error floor, which we could not attribute from these data.

Beyond marginal coverage, conditional (Mondrian, by-visit and by-risk-decile) coverage held in the body of the distribution in CPTS, the lowest-risk CDS decile over-covered (0.98), and the highest-risk decile under-covered in both cohorts (CDS 0.70, seed range 0.57 to 0.80; CPTS 0.67, 0.59 to 0.77; Supplementary Table S4), and a model-based Gaussian-process interval gave concordant uncertainty (mean 80% coverage 0.82), so the calibrated-uncertainty claim does not depend on a single method (Supplementary Table S5). Conformalized quantile regression did not improve on split-conformal. It matched coverage with intervals 3–6% wider on average (Supplementary Table S5). On the observed 48- and 60-month visits alone, coverage was 0.81 and 0.69 in CPTS (6 to 13 visits per re-split at 60 months) and 0.84 and 0.90 in CDS (Supplementary Table S6).

Cross-cohort transfer degraded prediction in both directions (median degradation ratio 1.25 from CPTS to CDS and 1.07 from CDS to CPTS; mean absolute degradation 170.3 and 75.1 cells/mm²), so we scope generalization claims to within-cohort use. Error was concentrated in the decision-relevant low-density bands (≤750 cells/mm² MAE 1,088 for CPTS, 394 for CDS), partly reflecting regression toward the cohort mean (Supplementary Table S7).

To test whether the parsimonious twin leaves accuracy on the table, we evaluated three more expressive models on the identical patient-disjoint split (a hierarchical-Bayesian cross-cohort nonlinear mixed-effects model with calibrated posterior uncertainty (80% interval coverage 0.85), a mechanism-plus-machine-learning universal differential equation with a per-cell loss rate skeleton and a neural closure, and a shared-random-effect joint longitudinal-survival model), and none improved held-out forecasting accuracy or survival discrimination over the simple twin (Supplementary Table S8). The added complexity bought principled, calibrated uncertainty but no accuracy gain; a failure of more expressive models to improve does not by itself validate the simpler one, but it is at least consistent with the parsimonious identifiable twin on this corpus.

A bounded in-silico ROCK-inhibitor counterfactual

Because the early postoperative loss and the per-cell loss rate are both constrained and together set the ECD trajectory, they are the natural levers for an in-silico what-if. We modeled a ROCK inhibitor as a bounded multiplier (0.4–1.0, where 1.0 is no effect) on the per-cell loss rate (a drug started after surgery), on the early postoperative loss fraction (a drug acting at surgery only), or on both (a drug given at surgery and continued), and propagated the uncertainty in that multiplier and in the calibrated parameters, drawn from the rejection-ABC posterior, through the simulator (Fig. 5). Under a probabilistic sensitivity analysis that placed 30% prior mass on no effect (values are means over 10 independent realizations; Methods), the mean preserved ECD at 36 months was 155 cells/mm² (95% interval 0 to 504) for a drug acting on the per-cell loss rate alone and 266 cells/mm² (0 to 821) for a drug acting on both phases, and at 60 months 238 and 345 cells/mm². A drug acting on the early postoperative loss alone preserved 94 cells/mm² at 36 months (−4 to 280; no benefit in 33% of draws). For the drugs acting on the per-cell loss rate only and on both phases, both intervals start at the no-effect mass (no benefit in 31% of draws for either), the probability of a clinically meaningful preservation of at least 100 cells/mm² was 0.526 and 0.596, values close to the prior probability of a multiplier below the tipping point (0.53 and 0.59), so they mainly restate the assumed drug-effect prior, and the tipping-point multiplier at which that preservation is reached was 0.85 and 0.91 (ranges over realizations 0.84 to 0.87 and 0.89 to 0.92). Supplementary Table S9 gives the one-way sensitivity, probability of benefit and tipping point for all three targets, including a drug acting on the early postoperative loss alone. In each draw, null and treated were simulated under the same parameter draw, so the interval carries the parameter uncertainty as well as the drug-effect prior. This is a mechanistic sensitivity illustration, not an efficacy estimate. There are no treated patients in CPTS or CDS, so no treatment-outcome association is observed and the magnitude is imposed by the assumed multiplier; for the same reason an E-value[29], which measures how strong unmeasured confounding would have to be to explain away an observed association, is not applicable here, and we do not report one. The only empirical anchor for the assumed direction and rough scale is the external clinical literature. In a retrospective, observational, post-cataract series of eyes with low baseline cell density, the ROCK inhibitor ripasudil was associated with markedly less endothelial cell-density loss than in untreated controls (an endothelial cell-density loss of −4.5% ± 13.0% with ripasudil versus a 14.1% ± 10.3% loss in untreated controls at 90–120 days after

cataract surgery, P = 0.0003), a different surgical context (not post-graft), so it supports plausibility only.[30] Because that series measured loss in the first months after surgery, it bears on the early postoperative phase at least as much as on the long-term one. The counterfactual uses the simulator's two constrained levers to convert a candidate cell-protective mechanism into a quantitative, uncertainty-bounded ECD-preservation hypothesis that states when the drug must act.

To show that the same mechanism can be posed spatially, we also ran the two-phase simulator on a curved corneal cap (posterior radius 6.5 mm, 124.0 mm², 330,387 cells) coupled to a radial drug field in which peripheral limbal clearance produces a central-hot exposure. At the bulk-density level the curvature itself was negligible (the dome and flat trajectories differed by at most 1.4 cells/mm², within Monte-Carlo noise), so the geometry adds nothing to a one-number ECD forecast; the dome's simulated 6-month density, 2,022 cells/mm², lay close to the observed CPTS mean of 2,024, a calibration target. On the dome a protective field acting on the per-cell loss rate alone raised the simulated 36-month density from 1,677 to 1,941 cells/mm² (+264) and produced a modest center-versus-periphery survival gradient, illustrating how the twin can represent a spatially non-uniform intervention once exposure is calibrated. The absolute exposure is a physiological-range estimate, and the dome result is a mechanism demonstration.

Image-to-ECD input path

A twin is only usable if its inputs are obtainable in clinic, so we built and benchmarked an image-to-ECD feeder that estimates density directly from a specular-microscopy frame (Fig. 6). A five-seed ensemble[31] regressor (shared ImageNet backbone initialization; head initialization and minibatch order vary) with test-time augmentation reached an ensemble MAE of 150.1 cells/mm² (per-seed 157.6 ± 3.7) on a frozen CPTS held-out test fold of 1,744 images, with ensemble predicted-versus-measured Pearson r = 0.950. The feeder has two operating regimes. Training on raw frames gives the best within-cohort accuracy (the values above), whereas an image-cleaning pass improves transfer to a confocal external set (ensemble Pearson r 0.935 vs 0.901 on raw frames) but not to a second specular external set (r 0.793 vs 0.808), at the cost of 29.8 cells/mm² of within-cohort accuracy (ensemble MAE 179.9 vs 150.1 cells/mm²). Raw-frame accuracy is therefore used for within-cohort results, and cleaning only where cross-device portability is the goal. A frozen-feature readout over public vision backbones[32–34] reached an ensembled MAE of 256.8 cells/mm² (R^2 = 0.823) on the CDS corpus. In-domain self-supervised (masked-image) pretraining of the feeder backbone, with every frame of the held-out test patients excluded from the pretraining pool, gave no detectable gain over ImageNet initialization across 10 fine-tune seeds (per-patient-visit MAE difference −1.4 cells/mm², 95% CI over seeds −4.3 to +1.5; patient-clustered bootstrap −6.5 to +3.8). When the pool included the test-fold frames the difference was −14.9 cells/mm² (95% CI over seeds −17.5 to −12.3), and −13.5 cells/mm² of it (95% CI over seeds −16.2 to −10.8) came from the backbone having seen the test images (Supplementary Table S10).

Across frozen foundation-model backbones, each represented by its class token concatenated with its mean patch token and read out by ridge regression under subject-grouped five-fold cross-validation, mean absolute error was 306.6 cells/mm² for ImageNet-21k ViT-S/16, 304.1 for DINOv2 ViT-S/14 and 273.3 for the histopathology-pretrained Phikon ViT-B/16. A mean ensemble of the three reached 256.8 cells/mm² (95% eye-clustered bootstrap CI [241.9, 272.7]; R^2 = 0.823, Spearman ρ = 0.914). A readout from a single token per backbone (the mean patch

token for ViT-S, the class token for DINOv2 and Phikon) on the same frames and folds gave 322.5, 313.9 and 298.1 cells/mm² and an ensemble MAE of 270.9 cells/mm² [255.4, 287.6] (R^2 = 0.807, ρ = 0.907), 14.1 cells/mm² above the concatenated readout (paired difference −14.1 [−17.1, −10.9]). Two medical foundation models read out on the same frames and folds, RETFound[35] (class token concatenated with mean patch token) and BiomedCLIP[36] (projected image embedding), had MAEs higher than the single-token ensemble by 118.0 [104.0, 131.2] and 86.2 [75.4, 97.1] cells/mm².

We ran two benchmarks against stronger comparators, segmentation-based density and incremental value. First, we compared three routes from the same CDS frames to ECD. They were a leading published turnkey specular-ECD segmentation engine[37,38] (official released weights; one of the few automated specular segmentation tools with public weights[39]); a generalist cell segmenter, Cellpose-SAM[40], applied without retraining to the uncropped captures; and the in-domain frozen-feature readout above. The engine segmented 1,554 of the 2,297 frames (67.7%), and its automated density ranked against reading-center ECD at Spearman 0.711 (95% CI 0.65 to 0.77; $p = 5.9 \times 10^{-240}$), above the $|r| \geq 0.5$ transfer criterion. The segmentation is deterministic, and a 534-frame subset gave Spearman 0.744 (Supplementary Table S11). Cellpose-SAM segmented 2,012 of the 2,297 frames (87.6%), at Spearman 0.777 (95% CI 0.74 to 0.81). Both segmenters agreed with ECD in rank, but neither matched the readout in absolute density. Converted to cells/mm² with one scale constant fitted out of fold, the engine's MAE on the 1,554 frames it segmented was 522.5 cells/mm² (95% CI 471.0 to 576.3), against 241.1 for the single-token frozen-feature ensemble (the readout against which Cellpose-SAM was also scored) and 636.7 for Cellpose-SAM's predictions on those frames, 49 of which it did not segment and which were scored by the training folds' median ECD. Over all 2,297 frames, with a failed frame predicted by the training folds' median ECD, the engine's MAE was 592.5 cells/mm², Cellpose-SAM's 656.0 and the single-token ensemble's 270.9. Second, image-derived ECD from a frozen ImageNet-21k ViT-S/16 ridge readout on the CDS frames met the incremental-value criterion over measured ECD plus elapsed time (lower forecast MAE with paired Wilcoxon $p < 0.05$), but the gain is negligible (measured-only MAE 241.2 cells/mm² against 240.5 with the image-derived ECD added, a 0.76 cells/mm² difference on 1,701 visit pairs, Wilcoxon $p = 0.016$); its demonstrated value is as an image-only ECD surrogate when a reading-center measurement is unavailable.

Out-of-sample mechanistic validation

Held out from the way it was calibrated, the simulator's decay mechanism was tested per eye on real observed visits in 251 CDS eyes with at least three calibration visits at 6 to 36 months, with 616 held-out visits at 48, 60, 90 and 120 months (Supplementary Table S12). Started from each eye's first post-operative reading, after the early postoperative loss, the two-phase simulator reduces to its long-term phase with one constant rate; we call this the constant-rate simulator to distinguish it from a time-varying arm. It has the same equations as the constant-rate comparator above, calibrated per eye instead of per cohort, so migration and proliferation, which are not identifiable per eye, are fixed at the comparator's cohort estimates (Methods). Calibrated per eye and averaged over 25 realizations, the constant-rate simulator forecast the held-out visits with a per-eye MAE of 425 cells/mm², better than last-value-carried-forward (536; paired patient-clustered bootstrap difference −111 cells/mm², 95% CI −153 to −66) but worse than per-eye log-linear extrapolation (359; +67, 95% CI +43 to +92) and than a biexponential with a pooled shape and one per-eye scale (333; +93, 95% CI +48 to +141). A per-cell loss rate that relaxes toward a

slower rate (endothelial loss after penetrating keratoplasty slows with time,[10,13] and normal adult endothelium loses about 0.6% of its cells per year[41]), with its shape fitted on the other half of the eyes, did not help (463; +38 against the constant rate, 95% CI +24 to +53). The gap widened with horizon (per-eye MAE 623 and 706 cells/mm² at 90 and 120 months for the constant-rate simulator against 319 and 253 for the biexponential), and at those horizons the simulator over-predicted ECD, because it reads ECD as surviving cells per guttae-free area, the per-eye guttae formation rate is not identifiable (posterior mean equal to the prior mean), and that term alone inflates the readout by a median factor of 1.4 at 120 months. Reading ECD as cells per total area instead lowered the constant-rate per-eye MAE from 425 to 392 cells/mm² (120-month MAE 706 to 593), pinning guttae formation to zero instead gave the same result within 0.1 cells/mm² because guttae enter the simulator only through the readout area, and the simulator still trailed log-linear extrapolation (+33, 95% CI +6 to +61) and the biexponential (+59, 95% CI +20 to +101), so the readout accounts for part of the long-horizon gap and the constant net-loss mechanism, set against a trajectory that plateaus, for the rest. The simulator is therefore the identifiable-mechanism and in-silico-trial layer of the twin, and per-eye forecasting is carried by the decay-twin above.

Independently, image-measured endothelial cell hexagonality[42–44] rose with ECD in the mechanistically expected direction (Spearman 0.214, $p = 9.2 \times 10^{-25}$ on 2,252 images), in the same direction across the micron-per-pixel sensitivity range (0.169 to 0.214). Calibrated on ECD alone, the simulator reproduced the direction of two morphometric relations it was not fitted to. Under the two-phase posterior fitted to CDS, simulated cell-area variation rose and hexagonality fell with ECD loss (Spearman correlation with ECD −0.54 and +0.16; 95% range over 30 estimator seeds −0.61 to −0.46 and 0.09 to 0.26), as observed on 2,476 CDS images from 807 patients, segmented by a Voronoi pipeline that estimates cell scale from each image and including 571 donor frames (−0.15, 95% CI −0.19 to −0.11; +0.27, 95% CI 0.23 to 0.31). The simulated magnitudes differ, with the area-variation relation about 3.7 times the observed one and the hexagonality relation about 0.6 times.

As an external-validity check on net cell loss, we related each eye's fitted exponential rate of ECD loss to graft outcome. Faster loss was associated with graft failure in both cohorts (CDS Spearman $\rho = +0.147$, $p = 1.4 \times 10^{-3}$; CPTS $\rho = +0.111$, $p = 1.4 \times 10^{-4}$), weakly with higher baseline ECD (CPTS $\rho = +0.106$), and not significantly with donor age or preservation time (Supplementary Table S13), which is consistent with accelerated cell loss contributing to failure.

Per-eye agent-based per-cell loss rate posteriors, calibrated by rejection ABC on each eye's visits up to a landmark, were computed for all eligible eyes (3,495 eye-landmark pairs) and tracked the exponential rate of ECD loss fitted to the same visits (Spearman $\rho = 0.87$ in CDS and 0.79 in CPTS). In CDS they did not predict graft failure after the 36-month landmark (hazard ratio per SD 0.97, 95% CI 0.65 to 1.44, adjusted for baseline ECD, donor age and recipient age; 28 failures in 279 eyes), and changed the apparent C-index of a landmark Cox model of the most recent ECD and donor age by −0.008 (95% CI −0.052 to +0.072) when added to it. They forecast the visits after the landmark more accurately than the two-phase cohort posterior calibrated on CPTS (per-cell loss rate 0.0116 month^{-1}, applied from each eye's first post-operative reading; per-eye MAE 425 against 742 cells/mm² from the 36-month landmark, difference −317, 95% CI −378 to −255; −347 from 48 months and −262 from 60 months), while against a cohort per-cell loss rate fitted to CDS itself (0.0215 month^{-1}) they were about as accurate from 36 months (difference +36, 95% CI −17 to +92) and less accurate from 48 and 60 months (+62, 95% CI +10

to +114; +171, 95% CI +111 to +225), a sensitivity comparator that is partly in-sample because it was fitted to many of the same eyes (Supplementary Table S13). The CPTS per-cell loss rate therefore does not transfer to the penetrating-keratoplasty cohort, consistent with the cohort contrast above, and per-eye calibration absorbs the difference. In CPTS, in an exploratory analysis with 7 failures after the 24-month landmark (915 eyes), a faster per-eye per-cell loss rate was associated with failure (adjusted hazard ratio per SD 2.60, 95% CI 1.64 to 4.09); the pre-landmark rate of ECD loss alone gave an optimism-corrected C-index of 0.945, and with both in one model the per-cell loss rate was no longer significant (hazard ratio 1.74, 95% CI 0.88 to 3.46). In CPTS, where the cohort posterior was calibrated on the same cohort, per-eye forecasts were more accurate than that posterior from the 24-month landmark (252 against 310 cells/mm², difference −59, 95% CI −78 to −39) but not from the 12-month landmark (300 against 265, difference +35, 95% CI +18 to +54), and less accurate than last-value-carried-forward (231) and log-linear extrapolation (227) from 24 months, and the per-eye 90% predictive intervals covered 17.2% of held-out visits in CDS and 32.5% in CPTS (Supplementary Table S13). A partially pooled (hierarchical) per-eye ABC model (Supplementary Table S13) gave the same conclusions, except that its CDS forecasts from 48 months did not differ significantly from those of the CDS-fitted cohort rate (+50, 95% CI −1 to +100). Of the unpooled per-cell loss rate posterior means at the 36-month (CDS, 304 eyes) and 24-month (CPTS, 933 eyes) landmarks, 26.0% and 49.3% lay within 10% of the prior range of either bound. Against the same range, 27.3% and 58.8% of the hierarchical means lay within that margin or beyond the range (12.5% and 31.6% beyond it), and none lay within 10% of either bound of the hierarchical model's wider range on the log scale (Supplementary Table S13).

The slowing of endothelial loss after penetrating keratoplasty (median loss 69% at 5 years and 76% at 10 years in eyes whose graft had not failed and that received corneas from donors aged 12 to 65 years[10,13]) is what the pooled-shape biexponential captures and the per-eye simulator, with its fixed guttae-free-area readout, does not.

Exploratory graft-survival analysis

We also fit a Cox graft-failure model on CDS (469 eyes, 40 events). It is underpowered (2.667 events per variable) and is exploratory. The apparent concordance of 0.616 fell to an optimism-corrected 0.496 with a corrected calibration slope of 0.483, indicating severe overfitting, and an independent bootstrap placed the corrected concordance at 0.495 with a 95% interval of [0.25, 0.66] that straddles 0.50. The standalone Cox model is therefore at chance. Elastic-net penalization recovered a corrected concordance of 0.568, the best result of this arm. Competing-risks analysis (Fine-Gray and Aalen-Johansen, with death as the competing event) and dropout diagnostics are in Supplementary Table S14. The cumulative-incidence function modestly corrects the naive 1−Kaplan-Meier failure estimate, and administrative dropouts differed from completers mainly on Fuchs status, recipient age and baseline ECD (standardized mean differences 0.37, 0.25 and 0.24).

Discussion

We replaced the fitted curve usually used for post-keratoplasty ECD loss with a mechanistic digital twin. In it, net endothelial cell loss has two components that the data identify, an early postoperative loss of about a fifth of the donor cells in CPTS and a slower long-term net ECD-loss rate. A single constant net ECD-loss rate started from the donor density reproduces the scale

of the cohort's ECD loss but not its shape, because it averages the two; separating them recovers the trajectory, including the 48-month visit held out of calibration. The per-cell loss rate after penetrating keratoplasty is 1.94 times that after endothelial keratoplasty in these two cohorts (2.86 times for the identified net rate), a mechanistic description of how they differ in ECD loss, although surgery type is confounded with cohort and era. Two further components sit beside the simulator, a forecasting model whose prediction intervals are calibrated empirically (split-conformal) and a readout that estimates ECD from a specular image. The early postoperative loss and the per-cell loss rate are the two parameters perturbed in a bounded in-silico drug counterfactual.

Assessing identifiability before the counterfactual determines which parameters an in-silico perturbation can use. The early postoperative loss fraction and the per-cell loss rate are the parameters the data constrain (contracting 0.992 and 0.978 from their priors; first-order Sobol indices of 0.69 for the early postoperative loss at 6 months and 0.76 for the per-cell loss rate at 36 months), so they are the parameters on which an in-silico perturbation can be defended, and the migration, proliferation, and guttae rates are the ones on which it cannot. A twin that reported a confident five-parameter fit and then intervened on proliferation would produce a number with no support in the data. The two constrained levers also differ in timing. A drug that acts only after surgery reaches the per-cell loss rate alone and preserved a mean of 155 cells/mm² at 36 months in the counterfactual, against 266 for one that also reduces the early postoperative loss, so under the assumed drug effect the size of the benefit depends on whether the agent acts at surgery. Pairing simulation-based inference with a formal identifiability and sensitivity analysis makes that distinction explicit. The same logic governs the uncertainty layer. Split-conformal intervals that hold their nominal coverage on held-out patients on average (0.796 and 0.823 against a 0.80 target) give the forecast an interval, and the counterfactual's 95% intervals, which carry the posterior uncertainty of the calibrated parameters as well as the drug-effect prior, keep the "what-if" inside the bounds the evidence can bear, a bound that is mechanistic and explicitly model-imposed.

Several limitations are intrinsic. ECD trajectories identify net cell loss only, as an early postoperative loss and a net ECD-loss rate (guttae formation becomes further identifiable by recovery when endothelial cell morphometry is added; migration and proliferation remain non-identifiable on available data), so we do not interpret the remaining rates. The two-phase fit is in-sample at 6 to 36 months, and its out-of-sample support is the 48-month visit. Its predictive intervals are wider than the sampling error of the cohort means, it under-predicts the few eyes observed at 60 months, and fitted to the combined CPTS eye set over 6 to 60 months it leaves residuals of +50 and +58 cells/mm² at 36 and 48 months (Supplementary Table S12). The neural posterior estimator behind the Table 3 means is over-confident, so its posterior widths are not reported as parameter intervals; they do enter the posterior predictive intervals of the fit comparison.

The per-cell loss rate depends on the priors of two unidentified rates. With proliferation and guttae formation pinned at the ends of their priors it ranged from 0.0063 to 0.0167 month^{-1}, its cohort ratio from 1.61 to 2.77 and the counterfactual for a drug acting on the per-cell loss rate only from 84 to 242 cells/mm², while the net rate, its cohort ratio (2.56 to 2.89) and the early postoperative loss changed little. In the CDS fit from the donor density the early postoperative loss fraction was 0.008, at the lower bound of its prior (Supplementary Table S12), so that fit did not separate an early postoperative loss from the per-cell loss rate, and the cohort ratio of the per-

cell loss rate rests on it. The contrast in the per-cell loss rate between the cohorts is confounded with surgery type and era. The mechanistic simulator is calibrated on CPTS at the donor-age-stratum level; the per-patient forecasts and their conformal intervals are produced by the calibrated ridge layer, so the twin is mechanistic in its parameters and counterfactual rather than in per-patient state estimation. Tested per eye at 48 to 120 months, the simulator beats only last-value-carried-forward among per-eye baselines and is outperformed by phenomenological decay curves, partly because its guttae-free-area ECD readout inflates with a guttae rate that the data do not identify, and partly because a constant net ECD-loss rate cannot follow a trajectory that plateaus (Supplementary Table S12).

Absolute per-eye ECD error remained large (MAE 449.0 and 465.5 cells/mm² at 12 and 24 months; Table 4), and the twin's gain over the biexponential baseline was modest (439.1 versus 452.4 cells/mm², 3.0%). The drug counterfactual is entirely in-silico and yields a quantitative, bounded hypothesis whose size depends on whether the drug reaches the early postoperative loss at surgery. Cross-cohort transfer degrades, so the twin is scoped to within-cohort use. Clinical site carries a quarter of the between-eye ECD variance in CPTS from 6 months onward (ICC 0.24 to 0.25, between-site SD 326 cells/mm² at 36 months), and a site term removes 62 to 65 cells/mm² of per-eye error for eyes at sites the model has seen but nothing for an unseen site (Supplementary Table S15); site is provenance rather than an input that transfers to an unseen center, and the primary twin stays site-free so that it applies to both cohorts and to a new center.

Image-derived ECD from the frozen-feature readout adds a statistically detectable but negligible 0.76 cells/mm² over a measured density, and the image feeder is positioned as a surrogate input. On an independent eye-bank image set (CLEAR-EC,[45] 9,000 frames) the end-to-end image-to-ECD feeder did not transfer at the native field of view (MAE 504 cells/mm², Spearman −0.27), and neither did the two frozen-feature readouts. For the end-to-end feeder the failure is consistent with an acquisition difference. Matching the field of view, with the crop chosen using CLEAR-EC labels, restored rank agreement (Spearman 0.58), and fine-tuning with CLEAR-EC labels reached an MAE of 193 cells/mm² (Supplementary Table S16). The survival arm is underpowered and corrected-down sharply.

The natural next step in identifiability is to find observables beyond ECD and guttae morphometry that constrain migration and proliferation. Morphometry already makes guttae formation identifiable (contraction from 0.01 to 0.55 and recovery r 0.94, in an analysis that does not pass simulation-based calibration; Supplementary Table S1) but leaves migration and proliferation unconstrained, so separating those rates will require additional spatial or dynamic readouts not available in this corpus. Calibrating the drug exposure on the curved corneal geometry and prospectively comparing the in-silico ROCK-inhibitor hypothesis against emerging clinical data are the next experimental steps. As built here on retrospective cohort data, the platform is the mechanistic and in-silico-trial layer of a digital twin; prospective two-way updating, in which each new measurement from a patient revises that patient's twin, is the next step toward a complete twin. As public endothelial pharmacology matures, a transparent mechanistic twin offers a principled way to prioritize which interventions are worth a real trial.

Methods

Cohorts and data harmonization

We used two public longitudinal post-keratoplasty cohorts with central specular-microscopy ECD, namely CPTS (post-DSAEK) and CDS, Specular Microscopy Ancillary Study (post-PKP). CDS is registered at ClinicalTrials.gov as NCT00006411 (first submitted October 13, 2000) and CPTS as NCT01537393 (first submitted February 15, 2012). Tabular clinical data were obtained through the Jaeb Center for Health Research and the specular image sets through the Cornea Image Analysis Reading Center (CIARC). Eyes with at least two reading-center ECD readings (in CPTS, readings with a released specular image, the pre-operative reading included; in CDS, readings at 6 to 60 months) were retained, giving 995 CPTS subjects (1,211 eyes) and 469 CDS subjects (one study eye each). ECD was the central reading-center value (CPTS analyzable final cell density; CDS gradeable average cell count). Baseline tables were compared with Welch (unequal-variance) tests with a normal approximation (continuous) and χ^2 tests (categorical).

Donor tissue characteristics. Donor age and donor sex are reported for both cohorts in Table 1 (CPTS 58.1 ± 12.7 years, 64.7% male; CDS 56.9 ± 14.6 years, 66.5% male). No donor tissue was received, processed, or handled by the investigators; all analyses used de-identified, publicly released clinical-trial data, and donor variables were used only as tabular covariates. Donor death-to-preservation interval and donor cause of death are present in the public CDS release and are complete for all 1,101 released CDS donors. Death to preservation was a mean of 6.6 hours (SD 2.8; median 6.3, IQR 4.6 to 8.4; range 0.4 to 22). Cause of death spanned 28 release categories, the commonest being cardiovascular disease in 519 donors (47.1%; two release labels combined), cancer in 208 (18.9%), and cerebrovascular accident in 106 (9.6%). Tissue source (eye bank), death-to-receipt interval, and preservation method are not included in the public CDS release.

Agent-based monolayer simulator

The endothelium was modeled as an off-lattice center-dynamics agent-based monolayer (each cell a moving point agent, in the broader tradition of cell-based epithelial models[46,47]) with an early postoperative loss fraction applied once at surgery and four monthly rate constants (per-cell loss, migration, proliferation, and guttae formation), initialized at the per-stratum donor ECD; simulated ECD equals surviving cells divided by an effective area discounted by guttae coverage. Because migration only displaces agents and conserves cell number, the identifiable bulk-ECD dynamics reduce to a one-time drop set by the early postoperative loss fraction followed by near-exponential decay set by the per-cell loss rate net of proliferation, with guttae formation acting through the readout area; the agent-based representation is nonetheless retained because it supplies the mechanistic early postoperative loss and per-cell loss rate levers for the counterfactual, the spherical-cap dome geometry for the radial drug field, and the morphometric channel through which guttae formation becomes identifiable. A flat unit patch was the calibration substrate; a spherical-cap variant (posterior radius 6.5 mm, surface area 124.0 mm², 330,387 cells whole-cornea) provided three-dimensional geometry for spatial extensions. At the bulk-ECD level the curvature effect was below Monte-Carlo noise, so the dome's role is qualitative (enabling a radial drug field[48]).

Each endothelial cell is a moving point agent in continuous space, rather than a lattice site (Cellular Potts) or a polygon (vertex model). The flat patch and the dome (posterior radius R ≈

6.5 mm, ~11 mm chord) share the same early postoperative loss, four rate processes and parameters.

State. A configuration is the set of live cell coordinates, a per-cell alive flag, and a scalar guttae area fraction $g \in [0, 1]$. The flat patch initializes N_0 agents uniformly on the patch and the dome initializes them on the cap surface; N_0 is set so the initial density matches the per-stratum donor ECD.

Early postoperative loss. At surgery ($t = 0+$), before the first monthly step, each cell is removed independently with probability $f \cdot m_a$, where f is the early postoperative loss fraction and $m_a \in [0.4, 1]$ is an optional drug multiplier (1 = no effect). We call this one-time loss early postoperative endothelial cell loss. The cohorts' first post-operative reading is at 6 months, so in the data it is the part of the fall from the donor density to that reading that the rate constants below do not explain; long-term endothelial cell loss is the slower loss those rate constants set.

Forward-Euler monthly step ($\Delta t = 1$ month). Each step applies four processes to the current state. Cell loss. Each live cell is lost with probability $1 - \exp(-k_d \cdot m_d \cdot \Delta t)$, where k_d is the per-cell loss rate (the monthly rate at which each live cell is lost in the long-term phase) and $m_d \in [0.4, 1]$ is an optional local drug multiplier (1 = no effect); losses are drawn independently (Bernoulli). Migration. Surviving cells are displaced by Gaussian jitter with standard deviation proportional to $k_m \cdot \Delta t$ (on the flat patch, $\sigma = k_m \cdot \Delta t \cdot L \cdot 0.02$ for patch side L; on the dome, a geodesic step in the local tangent plane re-projected to the sphere). Migration moves agents but does not change their number. Proliferation. The expected number of new cells is $k_p \cdot \Delta t \cdot N_{alive}$; the integer count is obtained by stochastic rounding, and each daughter is placed near a randomly chosen live parent. Guttae formation. The guttae area fraction increments by $k_g \cdot \Delta t$, capped at 1.

ECD read-out. Simulated ECD = N_{alive} / A_{eff}, where the effective (countable) area discounts guttae coverage, with $A_{eff} = L^2 \cdot \max(1 - g, 0.05)$ on the flat patch and $A_{eff} = 2\pi \cdot R^2 \cdot (1 - \cos \theta_{max}) \cdot \max(1 - g, 0.05)$ on the dome.

Why only net cell loss is identifiable from longitudinal ECD. Because migration is count-conserving, the cell count is governed by the early postoperative loss at surgery and the per-cell loss process net of proliferation, and the bulk ECD trajectory additionally by the guttae area discount. The early postoperative loss sets the level from which the ECD loss starts and the net long-term loss its slope on a logarithmic scale, so visits from 6 months onward separate them; guttae formation enters only through the area discount and the morphometric observable. Within the net long-term loss the per-cell loss rate trades off against proliferation and guttae formation, so the net ECD-loss rate, defined as the per-cell loss rate minus the proliferation and guttae formation rates (the net monthly rate of ECD loss), is the quantity ECD constrains most sharply. On the grafted monolayer, "guttae formation" is a generic excrescence / area-discount term rather than recurrence of host Fuchs guttae (donor grafts do not re-form Fuchs guttae), and it is non-identifiable from ECD alone.

Calibration and identifiability

The early postoperative loss fraction and the four rates were estimated by simulation-based inference (neural posterior estimation[19–21]) with uniform priors over a bounded box, using the mean of the reading-center ECD readings observed at 6, 12, 24, and 36 months (an eye without a

reading at a month does not enter that month's mean), stratified by donor-age tertile and 10,000 simulations per stratum, with each stratum's simulation started from its mean donor ECD. Estimation was per donor-age stratum of CPTS (1,091 eyes per estimator seed, drawn at random from 1,210 eligible eyes) with the sbi package, with a weighted-rejection approximate-Bayesian-computation fallback on the same summary statistics. Priors were early postoperative loss fraction ∈ [0, 0.5], per-cell loss rate ∈ [0.005, 0.06], migration ∈ [0.01, 0.20], proliferation ∈ [0, 0.005], and guttae formation ∈ [0, 0.005] per month. Summary statistics were the simulated ECD interpolated at the cohort's specular-microscopy visit months. The neural posterior estimator supplied the posterior means in Table 3. In simulation-based calibration[24] of that estimator (3 strata × 5 training seeds, 300 simulated datasets per network, 1,000 posterior draws each), its posteriors were too narrow for the per-cell loss rate and the three prior-dominated rates (pooled rank-uniformity $p < 10^{-36}$) and calibrated for the early postoperative loss fraction ($p = 0.50$), so only its point estimates are reported. Identifiability was assessed in a separate analysis with a fixed-reference rejection-ABC estimator on the same simulator and priors, using 1,000,000 reference simulations, 250 accepted draws (observed target, the mean of the CPTS readings at 6, 12, 24 and 36 months over the 1,210 eligible calibration eyes) and 300 simulation-based-calibration runs, reporting per-parameter prior-to-posterior contraction, parameter-recovery correlation, simulation-based-calibration rank uniformity, and first-/total-order Sobol sensitivity[22,23] of 36-month ECD (2,048 base samples). Reference tables of 5,000 to 1,000,000 simulations were compared, and the 1,000,000-simulation table, at which all five parameters passed simulation-based calibration, is used for the identifiability results and the counterfactual. The multi-observable bound used 5,000 reference simulations. Because the image-to-ECD readout is a regression, mechanism completeness was assessed by parameter identifiability and variance attribution (Sobol indices). The posterior was estimated with 30 independent seeds; Table 3 reports the mean over seeds with the seed range, because the estimate is seed-dependent at the third significant figure (coefficient of variation 1.0% for the per-cell loss rate). Model adequacy was assessed by a posterior predictive check against a constant-rate comparator, the same simulator with no early postoperative loss calibrated identically (200 posterior draws per seed). For each stratum and month it reported the residual of the predicted cohort mean against the observed mean, whether the observed mean lay inside the 80% and 95% predictive intervals, a Mahalanobis discrepancy over 6 to 36 months with a patient-clustered bootstrap covariance of the observed means plus the predictive covariance, and the deviance information criterion under a Gaussian synthetic likelihood, which penalizes the extra parameter; the 48- and 60-month visits were held out of calibration. For the cohort contrast, the two-phase model was fitted separately to CPTS (1,002 eyes) and CDS (308 eyes), eyes with at least four ECD readings, from the donor density (CPTS pre-operative reading-center ECD, CDS eye-bank ECD) with targets the mean of the readings observed at each month from 6 to 60 months, over 30 estimator seeds; a sensitivity analysis used targets interpolated to the visit calendar, carrying an eye's nearest reading forward or back; the 95% CI of the rate ratio pools the seeds with eye-bootstrap resamples. In a sensitivity analysis a constant rate was fitted from each cohort's 6-month mean, after the early postoperative loss, with 1,000 eye-bootstrap resamples per seed and the 95% range of the ratio over posterior draws. Per-eye forward validation calibrated the simulator from each eye's first post-operative reading on its own 6- to 36-month visits by rejection ABC on the per-cell loss and guttae formation rates (4,000 draws, 150 accepted; migration and proliferation pinned to the constant-rate comparator's single-seed posterior means, 0.105 and 0.00304 $month^{-1}$ (the same model fitted to the CPTS cohort; neither rate is identifiable per eye)), forecast the held-out 48- to

120-month visits as the mean of 25 realizations, and compared per-eye mean absolute error against last-value-carried-forward, per-eye log-linear extrapolation, a biexponential with a pooled shape and one per-eye scale, and a time-varying per-cell loss rate whose plateau shape was fitted on the other half of the eyes; paired differences were bootstrapped by patient (Supplementary Table S12).

Decay-twin and split-conformal intervals

The forecasting twin was a per-cohort, per-horizon ridge regression over race-blind, routinely collected donor/recipient features (recipient and donor age, sex, diabetes, baseline ECD, and donor tissue variables; recipient and donor race excluded for fairness; the retained features predict non-White versus White race with an out-of-fold AUC of 0.60 for recipients and 0.62 for donors pooled, at most 0.65 in any cohort). Data were split by patient identifier into training, calibration and test sets five times (seeds 0 to 4; 989, 251 and 440 eyes in the first) with zero leakage. Symmetric split-conformal intervals[25–28] were calibrated per cohort × horizon (Mondrian) at $\alpha = 0.20$ (80% target), and coverage and MAE were scored on the held-out patients' observed visits only (excluding the CDS 6-month baseline-anchoring reading), per re-split with eye-clustered bootstrap 95% CIs (2,000 resamples), and summarized as the mean over five re-splits with the seed range. Baselines (biexponential, log-polynomial mixed, linear mixed-effects random-slope, last-value-carried-forward, constant-mean) were compared by subject-grouped cross-validation with paired Wilcoxon tests and scored on observed visits, excluding the CDS 6-month anchoring visit. The five-fold twin-versus-biexponential comparison was repeated over 50 re-splits of the patient grouping (seeds 0 to 49, seed 0 being the reported partition) and pooled with an eye-clustered bootstrap (2,000 resamples).

Image-to-ECD feeder and benchmarks

The direct feeder was an ImageNet-pretrained vision-transformer backbone[32] with a dense regression head, trained per cohort with test-time augmentation and a five-seed ensemble[31] with a shared backbone initialization, and evaluated on a fixed CPTS held-out test fold. External validation used two public sets, the Rotterdam set of 52 in vivo confocal endothelium images (scored against their manual cell density[49]) and the 324 annotated images of the Freiburg specular set[39] (scored against their annotated cell count, which tracks density because the set's field of view is fixed). A frozen-feature comparison used public vision backbones[32–34], each represented by its class token concatenated with its mean patch token, with a ridge readout under eye-grouped cross-validation and 95% eye-clustered bootstrap CIs (4,000 resamples). The published turnkey specular-ECD engine[37,38] (official released weights, with the published frequency-watershed post-processing reimplemented in Python; TensorFlow 2.15.1) was benchmarked against a cross-instrument transfer criterion of $|\text{Spearman } r| \geq 0.5$ with $p < 0.05$. Our implementation loads the released weights by layer name, and 18 CNN-Edge layers carry different names in the released weight file; a layer-name mapping, which is not part of the original authors' code, loads those layers, so every layer of both networks is loaded from the released files. Cellpose-SAM was run with cellpose 4.1.1 and its released weights. The densities of the engine and of Cellpose-SAM were converted to cells/mm² by one scale constant fitted by least squares on the segmented frames of the other four of the readout's eye-grouped folds, and a frame that failed segmentation was predicted by the training folds' median ECD. The CDS image corpus comprises 2,297 frames from 584 eyes; the morphometric validation, the incremental-value test and the frozen-feature readouts of the CLEAR-EC check use a 2,280-

frame subset (579 eyes) that lacks 17 frames from 5 patients. Incremental value was tested on that subset (2,280 frames, one per visit, from 579 eyes) by forecasting the next reading-center ECD of each eye from its current visit (1,701 consecutive visit pairs from 481 eyes). A measured-only ridge regression on current ECD and elapsed time (with their product) was compared with the same regression plus an image-derived ECD, both under eye-grouped five-fold cross-validation and scored on identical held-out pairs. The image-derived ECD is the out-of-fold prediction of a ridge readout on frozen ImageNet-21k ViT-S/16 features, computed under eye-grouped cross-validation; it is distinct from the end-to-end CPTS feeder of Table 2. The image was counted as adding value only if it lowered forecast MAE with a paired Wilcoxon signed-rank test on the per-pair absolute errors at $p < 0.05$. The in-domain pretraining comparison (Supplementary Table S10) fine-tuned both initializations with 10 seeds, paired by seed, and reports a t interval over seeds beside a patient-clustered bootstrap of the seed-averaged errors (2,000 resamples).

Drug counterfactual

A ROCK inhibitor[50] was represented as a bounded multiplier (0.4–1.0) on the calibrated per-cell loss rate, on the early postoperative loss fraction, or on both, three targets that correspond to a drug started after surgery, a drug acting at surgery only, and a drug given at surgery and continued. For each target we report one-way sensitivity, a probabilistic sensitivity analysis (2,000 draws; 30% prior mass at the null, 70% uniform over the range), a tipping point for ≥100 cells/mm² preservation, and a benefit-reversal point. Parameter uncertainty was propagated by running each probabilistic draw under one of the 250 accepted draws of the rejection-ABC posterior, which passed simulation-based calibration, with null and treated simulated under the same parameter draw and simulator seed; the one-way curves and tipping points use that posterior's mean, and every counterfactual starts from the donor density of the calibration, 2,664.7 cells/mm², the mean of the three CPTS donor-age strata. The rejection-ABC analysis, with its 1,000,000 reference simulations, and the counterfactual were repeated in 10 independent Monte Carlo realizations with disjoint random seeds; the net ECD-loss rate and the counterfactual values in the text, the Table 3 note, the Fig. 5 legend and Supplementary Table S9 are means over the 10 realizations (a 95% interval is the mean of the per-realization interval bounds; ranges across realizations are given where stated), whereas the other rejection-ABC summaries, the benefit-reversal multiplier, the pinned-rate comparison in Supplementary Table S1 and the curves and values drawn in Fig. 1 and Fig. 5 are from one realization. All quantities are simulator counterfactuals; no treated patients exist in either cohort, so no observed treatment-outcome association is available and confounding-sensitivity metrics such as the E-value[29] are inapplicable by construction. Plausibility of the assumed direction is anchored to the external ROCK-inhibitor literature[30] rather than to in-cohort treatment data.

Exploratory graft-survival analysis

A Cox graft-failure model (CDS, 469 eyes, 40 events, 15 predictors) was internally validated by bootstrap optimism correction (300 replicates; 294 valid) of the concordance index and calibration slope[51–53], with an elastic-net penalized comparator[54]. Events per variable (2.667) are reported as evidence of limited power.

Statistics and reproducibility

Group comparisons used Welch (unequal-variance) tests with a normal approximation and χ^2 tests. Forecast accuracy used MAE and RMSE under subject-grouped cross-validation; model comparisons used paired Wilcoxon signed-rank tests with significance at $p < 0.05$. All statistical tests were two-sided (the χ^2 tests of independence and of rank uniformity being omnibus upper-tail tests), bootstrap confidence intervals are two-sided equal-tailed 95% intervals, and split-conformal prediction intervals are symmetric two-sided intervals at the nominal 80% level. Prediction intervals used split-conformal calibration with eye-clustered bootstrap confidence intervals for coverage. Survival discrimination used the concordance index with bootstrap optimism correction. Of the 469 analyzed CDS eyes, 30 came from 15 donors who each supplied two recipients (all 15 pairs are among the 451 CDS eyes of the comparison on observed visits), and a donor-clustered bootstrap gave a 95% CI of [−19.6, −7.3] cells/mm² for the twin-versus-biexponential MAE difference (patient-clustered [−19.5, −7.3]). For the Cox arm, missing baseline covariates (2 of the 6 modeled columns, 7 of 469 eyes) were multiply imputed[55] (10 imputations, pooled by Rubin's rules[56]), with pooled and complete-case hazard ratios differing by at most 0.02 (largest for Fuchs diagnosis, 1.02 versus 1.00); the ridge twin used median imputation for numeric features and an explicit missing level for categorical ones; structurally missing fields were not cross-cohort imputed. All random seeds were fixed and recorded (seed counts are given with each analysis), and test folds were frozen. Reporting follows TRIPOD+AI[53], with the prediction-model risk of bias assessed against PROBAST[57] (analysis domain flagged high-risk on sample size, EPV 2.667) and model credibility documented against the ASME V&V-40 framework at a moderate-influence, decision-support, non-regulatory context of use; the key reporting items and the credibility assessment are summarised in Supplementary Table S17, and the analysis sets, with their units and inclusion rules, are listed in Supplementary Table S18. Additional analyses reported in the Results (the patient-disjoint leakage audit, simulation-based-calibration of the agent-based inference, conditional/Mondrian conformal coverage and interval widths, the full baseline ladder, the Gaussian-process and conformalized-quantile-regression uncertainty comparisons, late-horizon coverage on observed visits, the band-stratified error, the competing-risks/dropout survival diagnostics, the multi-observable identifiability bound, the posterior predictive comparison with the constant-rate model and the cohort contrast, and the more-expressive-model comparison) are detailed in the Supplementary Information. A co-author (S.H.) independently reproduced the core analyses using only the public CPTS and CDS releases and the methods described in this paper. The reproduced estimates agreed with those reported here, except for the log-polynomial mixed baseline, whose optimizer sensitivity is described with Supplementary Table S3; the segmentation-engine benchmark (Supplementary Table S11) was not independently reproduced.

Use of AI-Assisted Technologies

The authors used ChatGPT and Claude for coding, debugging, and drafting assistance. AI also suggested analytical approaches. The ideas, interpretations, and conclusions are our own. All contributions from AI were reviewed and edited by the authors. The authors take full responsibility for the content of this manuscript.

Ethics statement

This study analyzed only de-identified, publicly available clinical-trial data and did not involve any new recruitment, intervention, or contact with human participants. The data derive from two National Eye Institute-funded clinical studies, the Cornea Preservation Time Study (CPTS) and the Cornea Donor Study (CDS), obtained through their publicly released, de-identified datasets. Under Dartmouth Health Institutional Review Board procedures, this secondary analysis of fully de-identified, publicly available data is classified as non-human-subjects research (formal determination: [IRB-TBD]). Informed consent was obtained from participants in the original CPTS and CDS studies under their respective institutional review board approvals, and no additional consent was required for this secondary analysis of de-identified data.

Data Availability

The clinical and ECD data analyzed here are from the Cornea Preservation Time Study (CPTS) and the Cornea Donor Study (CDS), both publicly available through the Jaeb Center for Health Research (public.jaeb.org); the CDS is registered as NCT00006411. The corresponding public specular-microscopy image sets are available by request from the Department of Ophthalmology and Visual Sciences, Case Western Reserve University, home of the Cornea Image Analysis Reading Center (CIARC), with the CPTS image set available through its request form and the Cornea Donor Study (Specular Microscopy Ancillary Study) images through the link on the Jaeb Center's CDS page. The CLEAR-EC eye-bank image set used for the external check is available as a restricted-access record on Zenodo (https://doi.org/10.5281/zenodo.21270595) to users who register with the CLEAR-EC Challenge and accept its data license, which does not permit redistribution; no CLEAR-EC images or labels are redistributed with this work. The analyses, content, and conclusions presented herein are solely the responsibility of the authors and have not been reviewed or approved by the CPTS or CDS study groups, the Jaeb Center for Health Research, the Cornea Image Analysis Reading Center, or the National Eye Institute. Aggregate derived values underlying each figure and table are provided in the manuscript and its Supplementary Information; no patient-level data are redistributed.

Code Availability

The custom analysis code supporting this study comprises the agent-based endothelial simulator, the simulation-based-inference calibration, the identifiability and Sobol-sensitivity suite, the split-conformal decay-twin, the drug-counterfactual analysis, and the image-to-ECD feeder. This code is held in a private repository that is available to reviewers on request and will be made public, with a Zenodo archive, upon publication. Trained model weights are available from the corresponding author on reasonable request, subject to the data-use terms.

Acknowledgements

We acknowledge the investigators and participants of the Cornea Preservation Time Study and the Cornea Donor Study, the Jaeb Center for Health Research as the data coordinating center, and the Cornea Image Analysis Reading Center (CIARC), Case Western Reserve University, for the central reading-center specular-microscopy data. These studies were supported by the National Eye Institute, National Institutes of Health. The funders and study groups had no role in this analysis or its conclusions. We acknowledge the CLEAR-EC Challenge for the CLEAR-EC eye-bank image set.[45] We thank Lucy Cano-Bartolo and Zayna Usman for their assistance with this work.

Author Contributions

B.C.: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing – original draft, Writing – review & editing, Project administration.
F.B.K.: Writing – review & editing.
S.H.: Software, Validation, Data curation, Visualization, Writing – review & editing.
A.S.: Visualization, Writing – review & editing.
S.S.: Visualization.
B.R.M.: Writing – review & editing.
M.M.J.: Writing – review & editing.
W.G.G.: Supervision, Writing – review & editing.

Competing Interests

The authors declare no competing interests.

References

1. Gain, P. *et al.* Global survey of corneal transplantation and eye banking. *JAMA Ophthalmol.* 134, 167–173 (2016). https://doi.org/10.1001/jamaophthalmol.2015.4776

2. Eye Bank Association of America. *2022 Eye Banking Statistical Report*. (2023).

3. Bourne, W. M. Biology of the corneal endothelium in health and disease. *Eye* 17, 912–918 (2003). https://doi.org/10.1038/sj.eye.6700559

4. Joyce, N. Proliferative capacity of the corneal endothelium. *Prog. Retin. Eye Res.* 22, 359–389 (2003). https://doi.org/10.1016/s1350-9462(02)00065-4

5. Joyce, N. C. Proliferative capacity of corneal endothelial cells. *Exp. Eye Res.* 95, 16–23 (2012). https://doi.org/10.1016/j.exer.2011.08.014

6. Price, M. O., Mehta, J. S., Jurkunas, U. V. & Price, F. W. Corneal endothelial dysfunction: evolving understanding and treatment options. *Prog. Retin. Eye Res.* 82, 100904 (2021). https://doi.org/10.1016/j.preteyeres.2020.100904

7. Lass, J. H. *et al.* Endothelial cell density to predict endothelial graft failure after penetrating keratoplasty. *Arch. Ophthalmol.* 128, 63–69 (2010). https://doi.org/10.1001/archophthalmol.2010.128.63

8. Patel, S. V. Graft survival and endothelial outcomes in the new era of endothelial keratoplasty. *Exp. Eye Res.* 95, 40–47 (2012). https://doi.org/10.1016/j.exer.2011.05.013

9. Cornea Donor Study Investigator Group. The effect of donor age on corneal transplantation outcome: results of the Cornea Donor Study. *Ophthalmology* 115, 620–626 (2008). https://doi.org/10.1016/j.ophtha.2008.01.003

10. Lass, J. H. *et al.* Donor age and corneal endothelial cell loss five years after successful corneal transplantation: Specular Microscopy Ancillary Study results. *Ophthalmology* 115, 627–632.e8 (2008). https://doi.org/10.1016/j.ophtha.2008.01.004

11. Lass, J. H. *et al.* Corneal endothelial cell loss three years after successful Descemet stripping automated endothelial keratoplasty in the Cornea Preservation Time Study. *JAMA Ophthalmol.* 135, 1394–1400 (2017). https://doi.org/10.1001/jamaophthalmol.2017.4970

12. Mannis, M. J. *et al.* The effect of donor age on penetrating keratoplasty for endothelial disease: graft survival after 10 years in the Cornea Donor Study. *Ophthalmology* 120, 2419–2427 (2013). https://doi.org/10.1016/j.ophtha.2013.08.026

13. Lass, J. H. *et al.* Donor age and factors related to endothelial cell loss 10 years after penetrating keratoplasty: Specular Microscopy Ancillary Study. *Ophthalmology* 120, 2428–2435 (2013). https://doi.org/10.1016/j.ophtha.2013.08.044

14. Okumura, N. *et al.* Enhancement of corneal endothelium wound healing by Rho-associated kinase (ROCK) inhibitor eye drops. *Br. J. Ophthalmol.* 95, 1006–1009 (2011). https://doi.org/10.1136/bjo.2010.194571

15. Okumura, N. *et al.* Rho kinase inhibitor enables cell-based therapy for corneal endothelial dysfunction. *Sci. Rep.* 6, 26113 (2016). https://doi.org/10.1038/srep26113

16. Kinoshita, S. *et al.* Injection of cultured cells with a ROCK inhibitor for bullous keratopathy. *N. Engl. J. Med.* 378, 995–1003 (2018). https://doi.org/10.1056/nejmoa1712770

17. Okumura, N. *et al.* Effect of the Rho-associated kinase inhibitor eye drop (ripasudil) on corneal endothelial wound healing. *Invest. Ophthalmol. Vis. Sci.* 57, 1284–1292 (2016). https://doi.org/10.1167/iovs.15-18586

18. Cranmer, K., Brehmer, J. & Louppe, G. The frontier of simulation-based inference. *Proc. Natl. Acad. Sci. USA* 117, 30055–30062 (2020). https://doi.org/10.1073/pnas.1912789117

19. Papamakarios, G. & Murray, I. Fast ε-free inference of simulation models with Bayesian conditional density estimation. *Adv. Neural Inf. Process. Syst.* 29, 1028–1036 (2016). https://doi.org/10.48550/arXiv.1605.06376

20. Greenberg, D. S., Nonnenmacher, M. & Macke, J. H. Automatic posterior transformation for likelihood-free inference. *Proc. Int. Conf. Mach. Learn.* 97, 2404–2414 (2019). https://doi.org/10.48550/arXiv.1905.07488

21. Tejero-Cantero, A. *et al.* sbi: a toolkit for simulation-based inference. *J. Open Source Softw.* 5, 2505 (2020). https://doi.org/10.21105/joss.02505

22. Sobol′, I. M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* 55, 271–280 (2001). https://doi.org/10.1016/s0378-4754(00)00270-6

23. Saltelli, A. *et al.* Variance based sensitivity analysis of model output: design and estimator for the total sensitivity index. *Comput. Phys. Commun.* 181, 259–270 (2010). https://doi.org/10.1016/j.cpc.2009.09.018

24. Talts, S., Betancourt, M., Simpson, D., Vehtari, A. & Gelman, A. Validating Bayesian inference algorithms with simulation-based calibration. Preprint at https://doi.org/10.48550/arXiv.1804.06788 (2018).

25. Vovk, V., Gammerman, A. & Shafer, G. *Algorithmic Learning in a Random World*. (Springer, Cham, 2022). https://doi.org/10.1007/978-3-031-06649-8

26. Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J. & Wasserman, L. Distribution-free predictive inference for regression. *J. Am. Stat. Assoc.* 113, 1094–1111 (2018). https://doi.org/10.1080/01621459.2017.1307116

27. Angelopoulos, A. N. & Bates, S. Conformal prediction: a gentle introduction. *Found. Trends Mach. Learn.* 16, 494–591 (2023). https://doi.org/10.1561/2200000101

28. Romano, Y., Patterson, E. & Candès, E. J. Conformalized quantile regression. *Adv. Neural Inf. Process. Syst.* 32, 3543–3553 (2019). https://doi.org/10.48550/arXiv.1905.03222

29. VanderWeele, T. J. & Ding, P. Sensitivity analysis in observational research: introducing the E-value. *Ann. Intern. Med.* 167, 268–274 (2017). https://doi.org/10.7326/m16-2607

30. Fujimoto, H., Setoguchi, Y. & Kiryu, J. The ROCK inhibitor ripasudil shows an endothelial protective effect in patients with low corneal endothelial cell density after cataract surgery. *Transl. Vis. Sci. Technol.* 10, 18 (2021). https://doi.org/10.1167/tvst.10.4.18

31. Lakshminarayanan, B., Pritzel, A. & Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. *Adv. Neural Inf. Process. Syst.* 30, 6402–6413 (2017). https://doi.org/10.48550/arXiv.1612.01474

32. Dosovitskiy, A. *et al.* An image is worth 16×16 words: transformers for image recognition at scale. *Proc. Int. Conf. Learn. Represent.* (2021). https://doi.org/10.48550/arXiv.2010.11929

33. Oquab, M. *et al.* DINOv2: learning robust visual features without supervision. *Trans. Mach. Learn. Res.* (2024). https://doi.org/10.48550/arXiv.2304.07193

34. Filiot, A. *et al.* Scaling self-supervised learning for histopathology with masked image modeling. Preprint at https://doi.org/10.1101/2023.07.21.23292757 (2023).

35. Zhou, Y. *et al.* A foundation model for generalizable disease detection from retinal images. *Nature* 622, 156–163 (2023). https://doi.org/10.1038/s41586-023-06555-x

36. Zhang, S. *et al.* A multimodal biomedical foundation model trained from fifteen million image–text pairs. *NEJM AI* 2, AIoa2400640 (2025). https://doi.org/10.1056/AIoa2400640

37. Vigueras-Guillén, J. P. *et al.* Deep learning for assessing the corneal endothelium from specular microscopy images up to 1 year after ultrathin-DSAEK surgery. *Transl. Vis. Sci. Technol.* 9, 49 (2020). https://doi.org/10.1167/tvst.9.2.49

38. Vigueras-Guillén, J. P. *et al.* Fully convolutional architecture vs sliding-window convolutional network for corneal endothelium cell segmentation. *BMC Biomed. Eng.* 1, 4 (2019). https://doi.org/10.1186/s42490-019-0003-2

39. Daniel, M. C. *et al.* Automated segmentation of the corneal endothelium in a large set of specular microscopy images using the U-Net architecture. *Sci. Rep.* 9, 4752 (2019). https://doi.org/10.1038/s41598-019-41034-2

40. Pachitariu, M., Rariden, M. & Stringer, C. Cellpose-SAM: superhuman generalization for cellular segmentation. Preprint at https://doi.org/10.1101/2025.04.28.651001 (2025).

41. Bourne, W. M., Nelson, L. R. & Hodge, D. O. Central corneal endothelial cell changes over a ten-year period. *Invest. Ophthalmol. Vis. Sci.* 38, 779–782 (1997).

42. Bergmanson, J. P. G. Histopathological analysis of corneal endothelial polymegethism. *Cornea* 11, 133–142 (1992). https://doi.org/10.1097/00003226-199203000-00007

43. Doughty, M. J. Concerning the symmetry of the hexagonal cells of the corneal endothelium. *Exp. Eye Res.* 55, 145–154 (1992). https://doi.org/10.1016/0014-4835(92)90102-x

44. Bourne, W. M. & McLaren, J. W. Clinical responses of the corneal endothelium. *Exp. Eye Res.* 78, 561–572 (2004). https://doi.org/10.1016/j.exer.2003.08.002

45. Al-Aswad, L. *et al.* CLEAR-EC: Corneal Learning for Endothelial Assessment and Review using AI for Endothelial Count. Zenodo https://doi.org/10.5281/zenodo.21270595 (2026).

46. Glazier, J. A. & Graner, F. Simulation of the differential adhesion driven rearrangement of biological cells. *Phys. Rev. E* 47, 2128–2154 (1993). https://doi.org/10.1103/physreve.47.2128

47. Fletcher, A. G., Osterfield, M., Baker, R. E. & Shvartsman, S. Y. Vertex models of epithelial morphogenesis. *Biophys. J.* 106, 2291–2304 (2014). https://doi.org/10.1016/j.bpj.2013.11.4498

48. Maurice, D. M. The use of fluorescein in ophthalmological research. *Invest. Ophthalmol.* 6, 464–477 (1967).

49. Selig, B., Vermeer, K. A., Rieger, B., Hillenaar, T. & Luengo Hendriks, C. L. Fully automatic evaluation of the corneal endothelium from in vivo confocal microscopy. *BMC Med. Imaging* 15, 13 (2015). https://doi.org/10.1186/s12880-015-0054-3

50. Lin, C.-W. *et al.* Discovery and preclinical development of netarsudil, a novel ocular hypotensive agent for the treatment of glaucoma. *J. Ocul. Pharmacol. Ther.* 34, 40–51 (2018). https://doi.org/10.1089/jop.2017.0023

51. Steyerberg, E. W. *et al.* Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J. Clin. Epidemiol.* 54, 774–781 (2001). https://doi.org/10.1016/s0895-4356(01)00341-9

52. Harrell, F. E., Lee, K. L. & Mark, D. B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med.* 15, 361–387 (1996). https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

53. Collins, G. S. *et al.* TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine-learning methods. *BMJ* 385, e078378 (2024). https://doi.org/10.1136/bmj-2023-078378

54. Simon, N., Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* 39, 1–13 (2011). https://doi.org/10.18637/jss.v039.i05

55. van Buuren, S. & Groothuis-Oudshoorn, K. mice: multivariate imputation by chained equations in R. *J. Stat. Softw.* 45, 1–67 (2011). https://doi.org/10.18637/jss.v045.i03

56. Rubin, D. B. *Multiple Imputation for Nonresponse in Surveys*. (Wiley, New York, 1987). https://doi.org/10.1002/9780470316696

57. Wolff, R. F. *et al.* PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann. Intern. Med.* 170, 51–58 (2019). https://doi.org/10.7326/M18-1376

Figure Legends

Figure 1. Study overview of the mechanistic corneal endothelial twin. A real specular-microscopy frame (left) is read into an endothelial cell density (ECD) estimate by an image-to-ECD feeder. The agent-based monolayer simulator is started from the donor ECD, 2,664.7 cells/mm², the mean of the three donor-age strata, and its cells are partly lost at surgery and then die, migrate, divide, and accumulate guttae according to four rate constants. The early postoperative loss and the net rate of ECD loss, the per-cell loss rate minus the proliferation and guttae-formation rates, are identified. The simulator produces a forward ECD trajectory and a bounded drug counterfactual for a drug that halves the per-cell loss rate (multiplier 0.5), in which the natural-decay curve (green) and the simulated cell-protective curve (blue) bound the preserved density. The counterfactual is hypothesis-generating and is not evidence of drug efficacy. The specular frame is third-party material from the Cornea Donor Study Specular Microscopy Ancillary Study, obtained through the Jaeb Center for Health Research and the

Cornea Image Analysis Reading Center under the Jaeb public dataset terms of use; it is reproduced under fair use (17 U.S.C. § 107) for scholarly illustration, and it is not covered by this article's license.

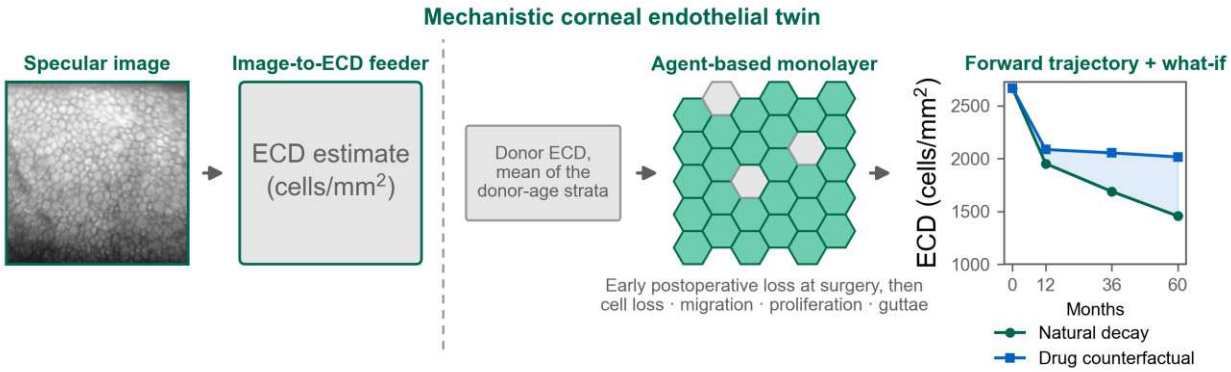


Figure 2. Twin architecture and component graph. Arrows show data flow. Specular-microscopy images and reading-center ECD feed the image-to-ECD feeder. The agent-based monolayer simulator, started from the donor ECD of each donor-age stratum, applies the early postoperative loss at surgery and then the four rate processes. Its simulations, the reading-center ECD and donor age feed calibration by simulation-based inference with the identifiability analysis, and the simulator, run under the rejection-ABC posterior from that calibration, gives the bounded drug counterfactual. The decay-twin, a ridge regression of reading-center ECD on donor and recipient clinical features with split-conformal 80% prediction intervals, gives the per-patient forecast. The graft-survival arm, a Cox model of graft-failure events on clinical features, is exploratory.

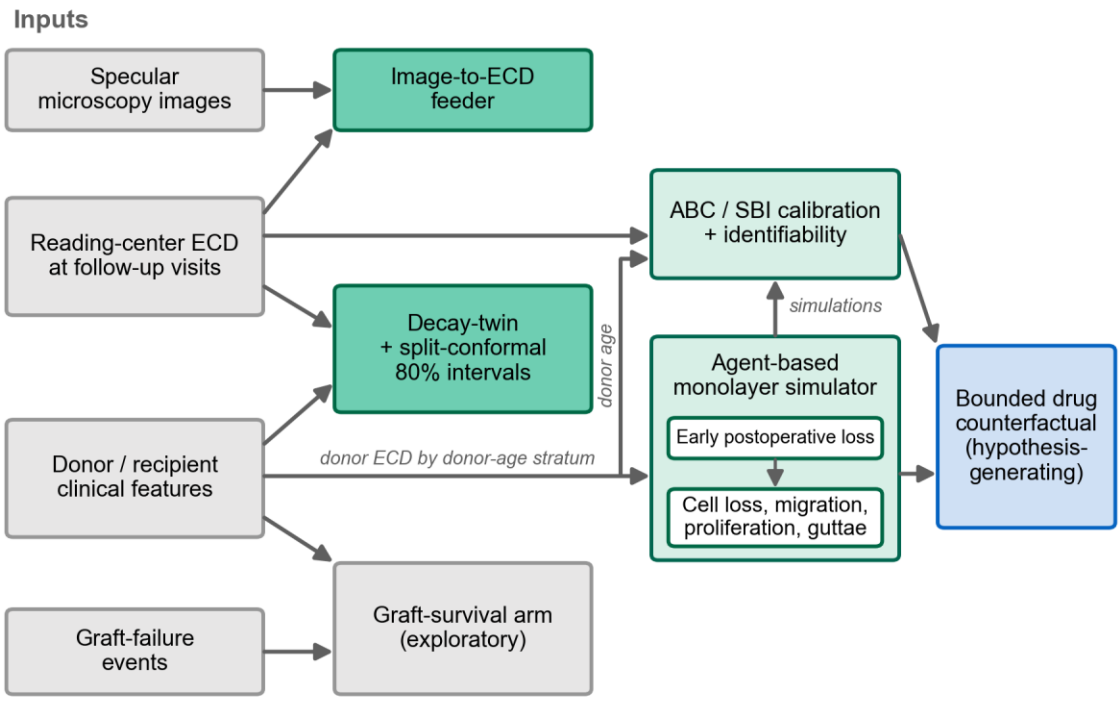


Figure 3. Calibration and identifiability of the two-phase agent-based simulator. (a) Posterior-mean parameters by donor-age stratum and pooled (log scale), with the 30-seed range, including the early postoperative loss fraction. (b) Prior-to-posterior contraction and parameter-recovery correlation per parameter under rejection ABC. By these per-parameter measures the early postoperative loss fraction and the per-cell loss rate are identifiable; migration, proliferation, and guttae formation are not. (c) First-order Sobol sensitivity of ECD at 6 and 36 months; the early postoperative loss fraction carries the early ECD loss and the per-cell loss rate the 36-month level. (d) Simulation-based-calibration rank histograms for the early postoperative loss fraction and the per-cell loss rate, uniform as expected (uniformity p annotated). Net cell loss is identifiable from ECD as the early postoperative loss and the net ECD-loss rate.

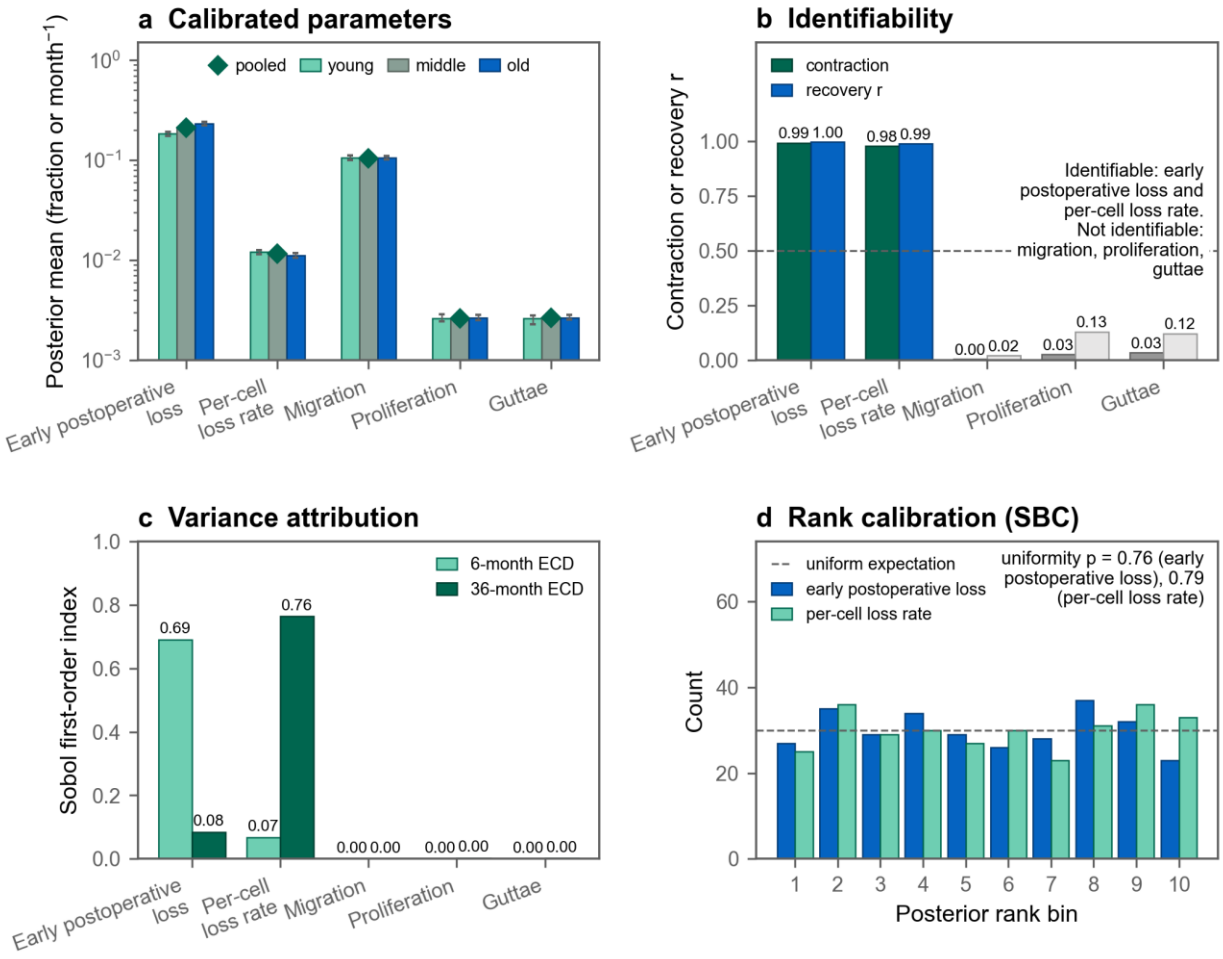


Figure 4. Decay-twin trajectories and conformal coverage. (a, b) Observed per-cohort mean ECD with the decay-twin's 80% split-conformal interval widths around it (CPTS shown over non-extrapolation horizons; CDS to 60 months). (c, d) Empirical coverage of observed visits by horizon, the mean over five patient-grouped re-splits with the seed range, against the [0.75, 0.85] acceptance band and 0.80 target; mean coverage annotated (CPTS 0.796, CDS 0.823). The CDS 6-month visit is the baseline-anchoring reading and is not scored. Coverage is lowest at 12 months in CDS and at 60 months in CPTS (6 to 13 visits per re-split).

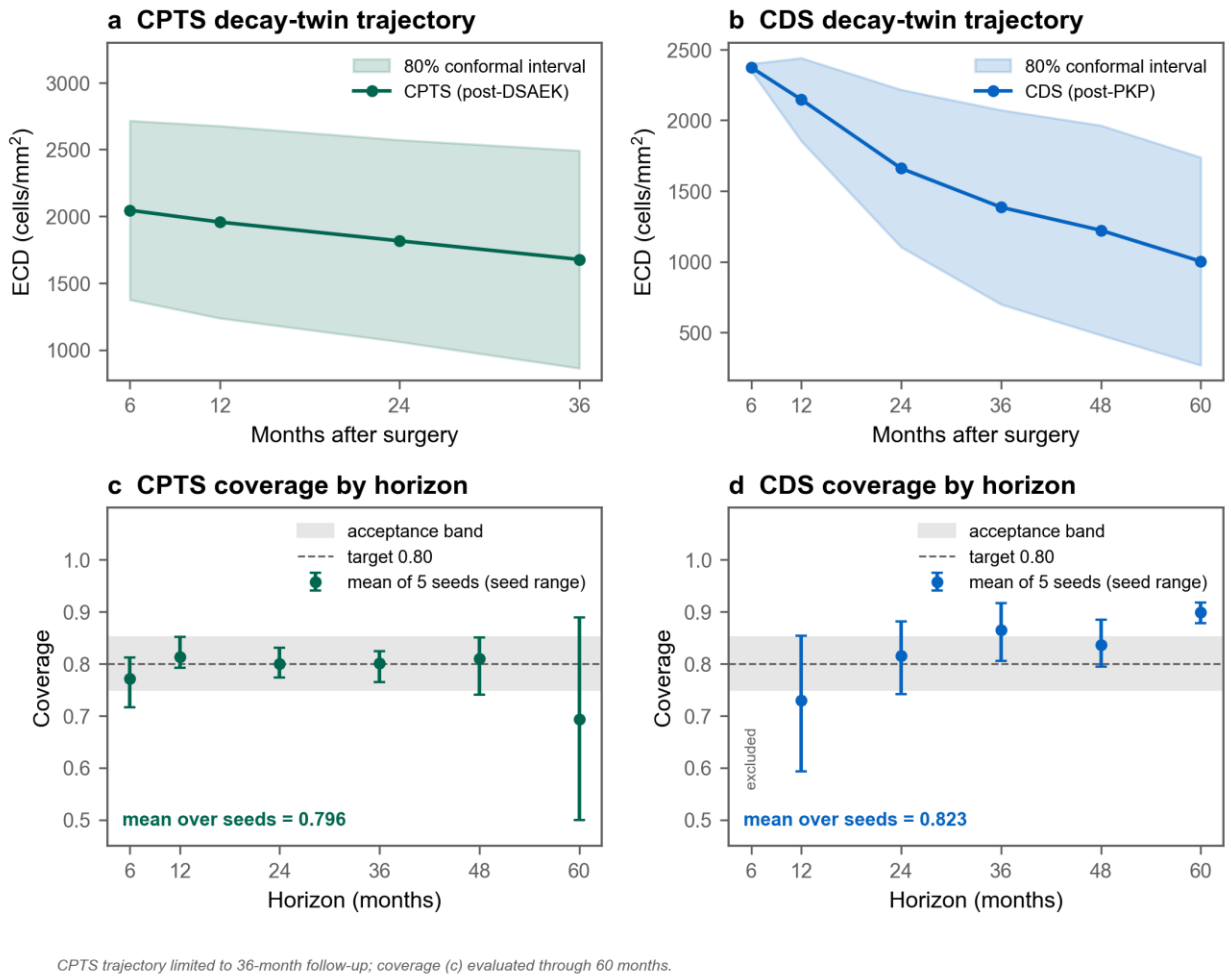


Figure 5. Bounded in-silico ROCK-inhibitor counterfactual under the two-phase model, for a drug acting on the per-cell loss rate only, on the early postoperative loss only, or on both. (a, b) One-way sensitivity, ECD preserved versus null at 36 months (a) and 60 months (b) as the drug multiplier varies from 1.0 (null) to 0.4, with the 100 cells/mm² threshold; the 36-month tipping-point multipliers of the realization drawn (0.84 per-cell loss rate only, 0.76 early postoperative loss only, 0.89 both) are marked on the threshold in (a). (c) Probabilistic sensitivity analysis with the calibrated parameters drawn from the rejection-ABC posterior. Averaged over 10 independent realizations, mean 36-month preservation was 155 cells/mm² (95% interval 0 to 504) for the drug acting on the per-cell loss rate only, 94 (−4 to 280) for the drug acting on the early postoperative loss only and 266 (0 to 821) for the drug acting on both phases, with a probability of no benefit of 0.31, 0.33 and 0.31 and of a meaningful effect (at least 100 cells/mm²) of 0.526, 0.436 and 0.596. The curves in (a, b) and the values printed in (c) are from one of these realizations; the values in this legend are 10-realization means. This analysis is a mechanistic what-if; with no treated patients there is no observed association, so it is hypothesis-generating and is not evidence of drug efficacy.

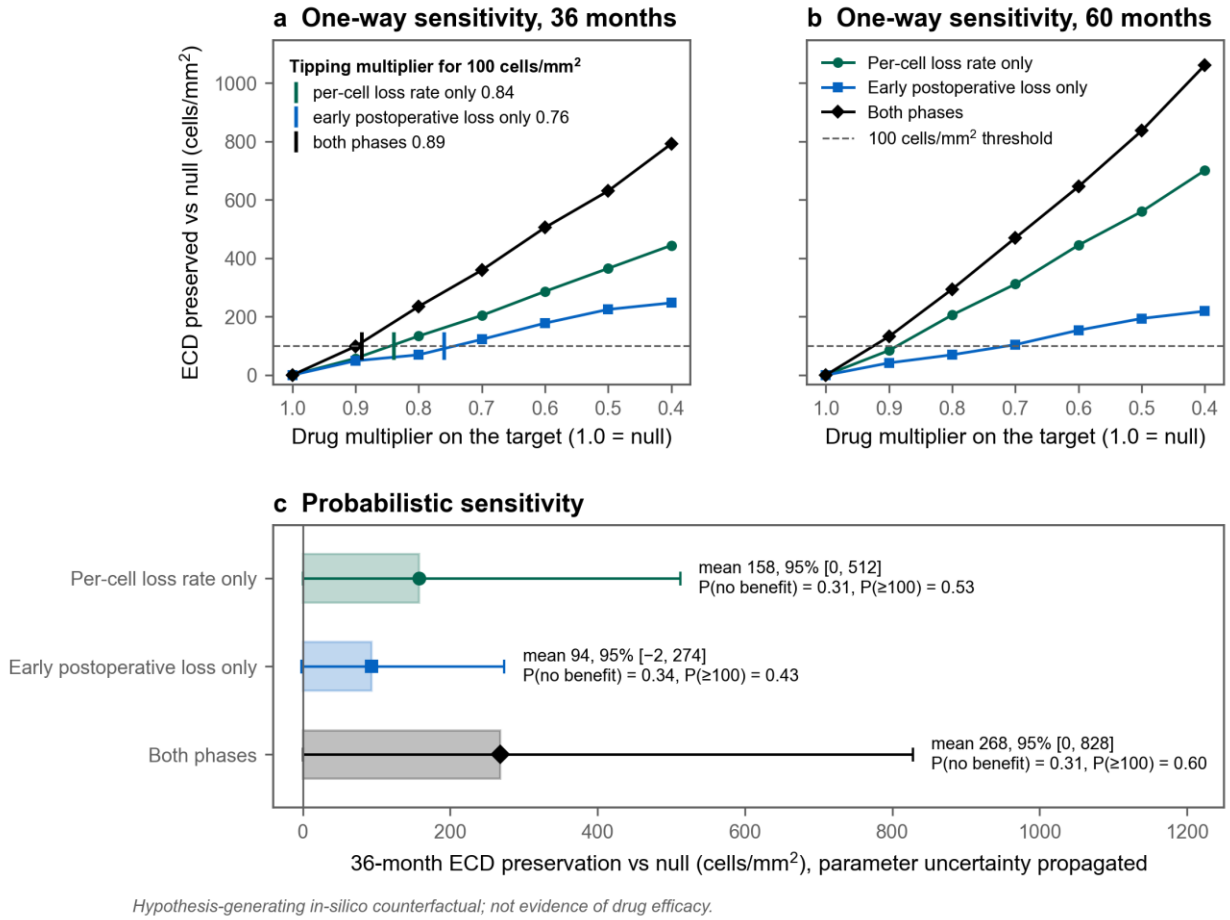
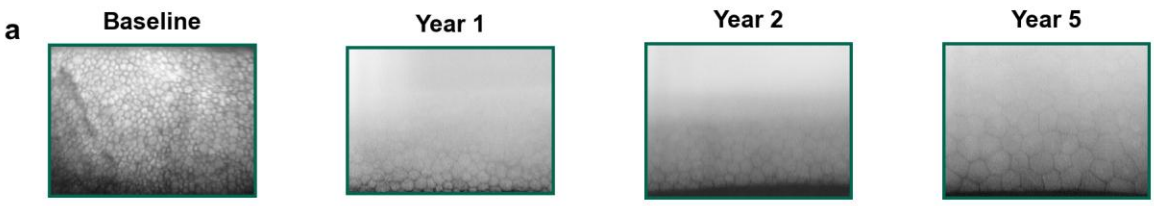
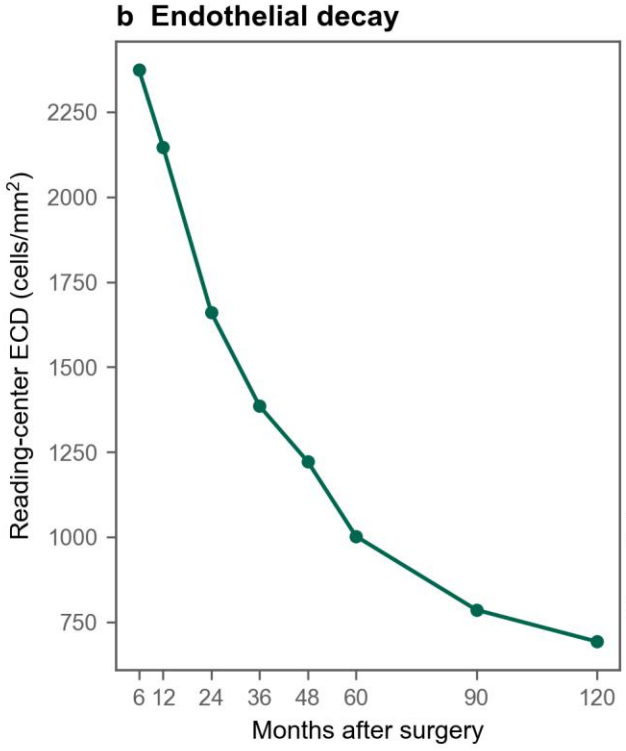


Figure 6. The image input path from real images to the decay they target, with feeder calibration. (a) Four real cleaned specular-microscopy frames of one CDS graft, the donor cornea imaged at the eye bank before surgery (Baseline) and the same graft at 1, 2 and 5 years after PKP, shown so that imaging artifacts (vignetting, overlay inpainting) are visible. (b) Mean reading-center ECD of the 2,280-frame CDS image subset at each visit, the quantity the feeder estimates. (c) Image-to-ECD feeder calibration (raw frames) on the frozen CPTS held-out test fold (1,744 images), with ensemble mean absolute error, Pearson r, Lin's concordance, RMSE, and bias. (d) Cross-device transfer (ensemble Pearson r on the confocal and specular external sets) for raw versus cleaned frames; image cleaning improves transfer to the confocal set but not to the specular set and costs within-cohort accuracy, so raw-frame accuracy is used for within-cohort results. The frames in (a) are third-party material from the Cornea Donor Study Specular Microscopy Ancillary Study, obtained through the Jaeb Center for Health Research and the Cornea Image Analysis Reading Center under the Jaeb public dataset terms of use; they are reproduced under fair use (17 U.S.C. § 107) for scholarly illustration, and they are not covered by this article's license.





c Feeder calibration (raw frames)

Ensemble test MAE	**150.1 cells/mm²**
Pearson r	**0.950**
Lin's CCC	**0.947**
RMSE	**218.4 cells/mm²**
Bias (pred − true)	**-12.6 cells/mm²**

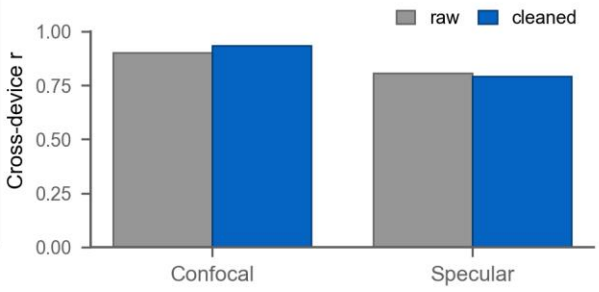


Table 1. Baseline characteristics of the two post-keratoplasty cohorts. Values are mean ± SD or n (%).

Characteristic	CPTS (post-DSAEK), n = 995	CDS (post-PKP), n = 469	p
Recipient age, years	69.7 ± 8.7	69.1 ± 8.7	0.250
Recipient female	612 (61.5%)	290 (61.8%)	0.905
Donor age, years	58.1 ± 12.7	56.9 ± 14.6	0.144
Donor male	644 (64.7%)	312 (66.5%)	0.499
Donor diabetes	256 (25.7%)	95 (20.3%)	0.022
Recipient diabetes	not recorded (CPTS)	58 (12.4%)	-

DSAEK, Descemet-stripping automated endothelial keratoplasty; PKP, penetrating keratoplasty. Recipient diabetes was not captured in the CPTS analysis tables. Comparisons: Welch (unequal-variance) test with a normal approximation (age) and χ^2 test (categorical).

Table 2. Image-to-ECD feeder performance and external benchmarks. Direct end-to-end image-to-ECD regressor (five-seed ensemble with a shared backbone initialization, test-time augmentation) on the frozen CPTS held-out test fold (1,744 images), with cross-device and external-engine comparisons. CI = 95% eye-clustered bootstrap CI (4,000 resamples). Raw-frame training gives the best within-cohort accuracy; an image-cleaning pass trades 29.8 cells/mm² of within-cohort accuracy for better transfer to the confocal external set. Pearson r values are those of the five-seed ensemble.

Metric	Value	Notes
Ensemble CPTS-test MAE (raw frames)	150.1 cells/mm²	5-member ensemble, frozen fold; within-cohort result
Per-seed MAE (raw frames)	157.6 ± 3.7 cells/mm²	training stochasticity
CPTS calibration Pearson r (raw frames)	0.950	ensemble prediction vs reading-center ECD
Within-cohort MAE, cleaned frames	179.9 cells/mm²	image cleaning regresses within-cohort accuracy
Cross-device Pearson r, confocal (raw / cleaned)	0.901 / 0.935	Rotterdam external set; cleaning aids transfer
Cross-device Pearson r, specular (raw / cleaned)	0.808 / 0.793	Freiburg external set; cleaning does not aid transfer
Frozen-feature best single backbone MAE	273.3 cells/mm²	CDS, eye-grouped CV; Phikon, class token concatenated with mean patch token
Frozen-feature ensemble MAE (R^2)	256.8 [241.9, 272.7] (0.823)	CDS, eye-grouped CV; single-token readout 270.9 [255.4, 287.6] (0.807)
Vigueras-Guillén engine, Spearman vs ECD	0.711	above $\|r\| \geq 0.5$ transfer threshold (Supplementary Table S11)
Vigueras-Guillén segmentation yield	0.677	1,554 of 2,297 images

The published turnkey specular-ECD segmentation engine transfers in rank to the older Cornea Donor Study specular corpus (Spearman 0.711, above the $|r| \geq 0.5$ transfer criterion), but with one out-of-fold scale constant its MAE over all frames is 592.5 cells/mm²; an in-domain frozen-feature readout predicts the same reading-center ECD at R^2 0.823 (three-backbone ensemble; 0.77 to 0.80 per backbone, Supplementary Table S11) on the identical images.

Table 3. Calibrated two-phase agent-based monolayer parameters and their identifiability. Posterior means from simulation-based inference (neural posterior estimation, 10,000 simulations per donor-age stratum of CPTS, targeting the mean of the readings observed at 6, 12, 24 and 36 months), pooled over the three strata and per stratum, each given as the mean over 30 independent seeds of the estimator with the 30-seed range in parentheses (the intervals are minimum-to-maximum ranges over 30 Monte Carlo seeds of the estimator, not confidence intervals); the third significant figure is seed-dependent. Identifiability and 95% posterior intervals from a separate rejection-ABC analysis (1,000,000 reference simulations, 250 accepted draws at the pooled CPTS target, 300 simulation-based-calibration runs).

Parameter	Pooled mean (30-seed range)	Youngest stratum	Middle stratum	Oldest stratum	Prior-to-posterior contraction	Recovery Pearson r	Sobol S_1 (36-month ECD)	Verdict
Early postoperative loss fraction at surgery	0.211 (0.207 to 0.215)	0.186 (0.176 to 0.192)	0.214 (0.200 to 0.223)	0.234 (0.222 to 0.241)	0.992	0.997	0.084	identifiable
Per-cell loss rate ($month^{-1}$)	0.0116 (0.0113 to 0.0120)	0.0121 (0.0114 to 0.0126)	0.0117 (0.0108 to 0.0124)	0.0111 (0.0106 to 0.0117)	0.978	0.988	0.764	identifiable
Migration ($month^{-1}$)	0.105 (0.102 to 0.108)	0.106 (0.0998 to 0.112)	0.104 (0.0984 to 0.109)	0.106 (0.102 to 0.111)	0.0038	0.020	0.000	prior-dominated
Proliferation ($month^{-1}$)	0.00263 (0.00256 to 0.00272)	0.00263 (0.00245 to 0.00288)	0.00263 (0.00247 to 0.00272)	0.00264 (0.00252 to 0.00285)	0.026	0.129	0.003	prior-dominated
Guttae formation ($month^{-1}$)	0.00264 (0.00253 to 0.00270)	0.00261 (0.00228 to 0.00280)	0.00265 (0.00254 to 0.00275)	0.00266 (0.00253 to 0.00285)	0.034	0.121	0.002	prior-dominated

Donor-density start, pooled over the three strata = 2,664.7 cells/mm². Simulation-based-calibration rank-uniformity $p > 0.05$ for all five parameters under the rejection-ABC estimator; the neural posterior estimator's posteriors are too narrow for the per-cell loss rate and the three prior-dominated rates (Supplementary Table S1), so only its point estimates are given.

Longitudinal ECD identifies net cell loss as two quantities, the early postoperative loss and the net ECD-loss rate; migration, proliferation and guttae formation are not identifiable and are reported as prior-dominated. The per-cell loss rate correlates with proliferation (0.70) and guttae formation (0.74) in the rejection-ABC posterior; the net rate, per-cell loss minus proliferation minus guttae formation, is 0.0064 month^{-1} (0.0053 to 0.0076; mean and 95% interval averaged over 10 independent realizations of the rejection-ABC analysis, the other rejection-ABC values here being from one realization). Rejection-ABC posterior means (95% intervals) were early postoperative loss fraction 0.209 (0.190 to 0.229), per-cell loss rate 0.0113 (0.0074 to 0.0159), migration 0.106 (0.012 to 0.197), proliferation 0.00251 (0.00017 to 0.00489) and guttae formation 0.00230 (0.00014 to 0.00474) month^{-1}.

Table 4. Decay-twin held-out accuracy and split-conformal coverage by horizon. Per-cohort ridge twin, five subject-grouped re-splits of all 1,680 eyes into training, calibration and test sets (989 / 251 / 440 eyes in the first; no patient identifier in more than one partition). Mean absolute error (MAE) in cells/mm² and per-visit split-conformal coverage at the nominal 80% target, both on held-out test eyes, scored on observed visits only and averaged over the five re-splits.

Horizon	CPTS MAE	CPTS coverage (5 re-splits)	CDS MAE	CDS coverage (5 re-splits)
Month 6	435.2	0.77	not scored[a]	not scored[a]
Month 12	449.0	0.81	219.7	0.73
Month 24	465.5	0.80	379.5	0.82
Month 36	499.0	0.80	397.8	0.86
Month 48*	529.0	0.81	451.7	0.84
Month 60*	628.4	0.69	405.4	0.90
Mean coverage (seed range)		0.796 (0.771 to 0.832)		0.823 (0.773 to 0.862)

*Few CPTS test eyes were observed at months 48 and 60 (93 to 107 and 6 to 13 visits per re-split); accuracy claims are scoped to months 6–36. Coverage by horizon is per observed visit, averaged over the 5 re-splits; the mean coverage row gives the range over re-splits. The mean of each cohort lies inside the [0.75, 0.85] acceptance band; every CPTS re-split does, and four of five CDS re-splits do (the fifth is 0.862, CI 0.821 to 0.901; CI = 95% eye-clustered bootstrap CI, 2,000 resamples). Under subject-grouped 5-fold cross-validation over all eyes, scored on observed visits, the twin lowers per-eye MAE versus a biexponential decay baseline (439.1 vs 452.4 cells/mm²; paired Wilcoxon $p = 5.7 \times 10^{-5}$). On the five re-splits of this table the twin's MAE is lower in all five (mean 2.2%) but individually significant in one.

[a] For CDS the 6-month visit is the baseline-anchoring reading (measured ECD coincides with the seeded baseline; last-value-carried-forward MAE 1.3 cells/mm²), not a forecast; neither its MAE nor its coverage is scored. CPTS month 6 is a genuine forecast (its baseline differs from the 6-month reading).