International Journal of Artificial Intelligence, Data Science, and Machine Learning
Grace Horizon Publication | Volume 7, Issue 3, 95-111, 2026
ISSN: 3050-9262 | https://doi.org/10.63282/3050-9262.IJAIDSML-V7I3P112

Original Article

Digital Twin Modelling Of Patient Journeys in Healthcare Relationship Management Systems

Brahmananda Naidu Dabbara
Independent Researcher, USA.

Received On: 21/06/2026 ***Revised On: 23/07/2026*** ***Accepted On: 01/08/2026*** ***Published On: 08/08/2026***

Abstract *- Healthcare relationship management systems record what has already happened to a patient. They hold appointments, milestones, care gaps, service cases and outreach history, and they act on that record through rules and schedules. What they do not do well is anticipate what happens next, which leaves engagement decisions reactive: contact after a missed appointment rather than before it, channel chosen by predefined rule rather than by predicted responsiveness, and care gaps addressed once already overdue. This paper describes a digital twin of the patient journey built inside a relationship management platform, and reports a research evaluation of it. The twin models the journey as a dynamic personalised process rather than as a copy of the medical record. It holds where the patient sits in the care and engagement lifecycle, how prior interactions produced that position, and the likely next steps under different engagement strategies. It produces three kinds of output: predictions and risk scores, what-if simulations comparing candidate interventions, and an engagement recommendation carrying a next best action, timing, channel, priority and the factors behind it. Validation used a temporal holdout on real reconstructed journeys, with the twin initialised at an earlier point and asked to simulate forward without seeing later events, compared on next-state accuracy, sequence similarity, timing accuracy, risk calibration, subgroup performance, and clinical and operational review of unusual transitions. Messy real-world behaviour was deliberately retained. The evaluation covered about 50,000 patient journeys across five service lines, with an architecture built to support 8 to 10 facilities on a shared platform. Against the existing rule-based engagement process, engagement and response rate improved by about 18 to 25 per cent, appointment adherence by about 12 to 18 per cent, care-gap and follow-up completion by about 15 to 22 per cent, engagement response time for prioritised cases fell by about 25 to 35 per cent, unnecessary outreach fell by about 15 to 20 per cent, and engagement-related operational efficiency improved by about 10 to 18 per cent. Every figure in this paper is a research-evaluation result. None is an audited production outcome, and every range is reported as a range because no point estimate was substantiated. The evaluation also produced findings that changed engagement strategy: timing mattered more than message count, repeated contact on one channel reduced responsiveness, and journey friction was routinely misread as disengagement. Sometimes the best action was no immediate outreach. A three-layer separation of patient identity, clinical data and simulation lets the twin operate on representations rather than on identified records, and simulation output is treated as sensitive in its own right. Throughout, the system is decision support and not autonomous clinical decision-making.*

Keywords - *Digital Twin, Patient Journey, Healthcare Relationship Management, Journey State Modelling, What-If Simulation, Next Best Action, Patient Engagement, Care-Gap Management, Risk Calibration, Privacy-Preserving Simulation, Clinical Decision Support.*

1. Introduction

A healthcare relationship management system is, at its core, an event log with a workflow attached. It knows that an appointment was scheduled, that a reminder went out, that the patient confirmed and then rescheduled twice, that the visit was eventually completed, and that a follow-up is now pending. It knows which service cases were opened and whether they were resolved. It knows what outreach was attempted and, sometimes, whether it landed. What it does not hold is a view of where that patient is heading.

That gap has a practical shape. Engagement decisions in such systems are typically driven by rigid rules and fixed schedules applied to historical or static data, which produces six recurring failure areas. The first is intervention timing: outreach is triggered after a missed appointment rather than before one becomes likely. The second is patient prioritisation, where a queue is worked in the order the records arrive rather than in the order that would do the most good. The third is channel selection, chosen by predefined rule rather than by predicted responsiveness on the channel. The fourth is journey sequencing, where the order of contacts is fixed by campaign design rather than by the patient's actual position in their care pathway. The fifth is care-gap management, where a gap is addressed only once it is already overdue. The sixth is scenario evaluation: changing an engagement strategy meant testing it on real patients, because there was no way to try it anywhere else. Stated plainly, traditional healthcare relationship management

recorded past events but offered little ability to predict what might happen next.

Mapping the patient journey is a well established discipline in health services improvement, and process mining takes the idea further by discovering the real pathway from event logs rather than from a documented ideal [19]. Healthcare has been a productive setting for it, first as a case study run against a real hospital's event logs [10] and later as a formalised methodology [14]. Those methods describe the population. They tell an operations team what pathways exist and where the delays sit. They do not, on their own, tell a coordinator what to do about one specific patient this week.

The digital twin concept offers a different framing. A twin is a simulated representation of an individual instance, kept current with that instance, and used to ask forward-looking questions about it [8]. Engineering practice has taken the idea a long way, from product lifecycle representation to operational monitoring and predictive maintenance [18], and health applications are a fast-growing and much-debated area [20]. Most health digital twin work concentrates on physiology: a model of an organ, a disease process, or a therapeutic response. Comparatively little of it models the patient's journey through a care and engagement lifecycle as the thing being twinned.

This paper describes such a twin and reports a research evaluation of it. The unit being modelled is the journey, not the body. The twin holds journey state, transition history and derived behavioural signals, and it is used to answer three questions that the underlying relationship management system could not answer: what is likely to happen next for this patient, what would happen differently under each candidate intervention, and which single next engagement is worth making now. The output is presented to a person. Clinically significant decisions stay with qualified professionals, in line with long-standing expectations for clinical decision support [17].

The contribution is fourfold. First, a journey state representation that preserves sequence and elapsed time rather than collapsing history into counts, with the derived behavioural features that make forward simulation possible. Second, a what-if simulation and next-best-engagement layer that compares candidate interventions from the same starting state and returns a probability and a confidence for each rather than a single verdict. Third, a validation design that runs a temporal holdout on real reconstructed journeys, retains messy behaviour instead of cleaning it away, and tests six distinct properties rather than one aggregate score. Fourth, a privacy architecture that separates patient identity, clinical data and simulation into distinct layers, and that treats simulation output as sensitive in its own right, because a predicted disengagement risk discloses something about a patient even when the source record is hidden.

Section 2 states precisely what was built and evaluated and what remains proposed architecture, because the distinction matters for every claim that follows.

2. Scope and Evidentiary Basis

This section exists so that no reader has to guess which parts of this paper are completed work, which are measured results, and which are design proposals and publication guidance. The distinction is maintained deliberately throughout, and where the source material was phrased conditionally it has not been converted into a completed action.

- What was built and evaluated: A journey state representation, an input feature set derived from patient and relationship-management signals, a forward-simulation component, a what-if comparison component, and a next-best-engagement output. These were evaluated using a temporal holdout on real reconstructed historical journeys, compared against the simpler rule-based engagement logic already in place, across six validation dimensions, at a scale of about 50,000 patient journeys spanning five service lines.
- What the numbers are: Every improvement figure reported in this paper, and specifically every figure in the results table, is a research-evaluation result: None of them is an audited production outcome. They are reported as ranges because that is the form in which they were substantiated; no point estimate within any range was substantiated, so no midpoint is given anywhere in this paper. The evaluation scale is likewise a research-evaluation scale and not a verified deployed footprint.
- What the scale figures mean: About 50,000 patient journeys and five service lines describe the evaluation, not a production deployment. The figure of 8 to 10 facilities is explicitly what the architecture was built to support on a shared platform. It is a design capacity, not a count of facilities running the system.
- What is proposed rather than completed: The multi-facility shared-platform arrangement is described as an architectural capability. The horizontal extension mechanism, by which new journey states, events and transition rules are defined for additional service lines, is described as a design property. The staged adoption sequence in Section 12 is a recommendation drawn from the work, not a report of an adoption programme that has run to completion. Several privacy and governance controls in Section 9 are stated as requirements of the design; where a control is a requirement rather than a measured outcome, the text says so.
- What is not claimed: No causal claim is made from any simulated result. A what-if outcome is a scenario estimate, and it is never treated as evidence that an intervention caused an improvement. Where historical comparison groups existed, predicted effects were compared with observed outcomes among similar patients; where they did not, the output was labelled a scenario estimate requiring further validation rather than a causal clinical conclusion. No patient-level record appears anywhere in this paper. No patient vignette is

presented as a real case. The single illustrative journey state used in Section 6 and Figure 4 is a constructed example, labelled as such on the figure itself.

- What is deliberately absent: No performance curve, receiver operating characteristic curve or calibration plot appears in this paper, because the underlying series for such charts do not exist in the evaluation record. Where a chart would have required data that was not produced, the mechanism is drawn instead and the figure is labelled a schematic.

3. Related Work

- Digital twins: The digital twin as a concept predates its current popularity and was originally framed around mitigating unpredictable emergent behaviour in complex systems by maintaining a virtual counterpart of a physical instance and interrogating it [8]. Industrial practice has since matured the idea considerably, and a survey of that practice identifies the recurring elements: a persistent link to the real instance, a model kept current with it, and simulation used for prediction and decision support rather than for description alone [18]. The health literature has taken up the term energetically. A review of health digital twins as tools for precision medicine sets out the computational, implementation and regulatory considerations, and it is candid that much of the field is aspirational and that the term covers a wide range of maturity [20]. That review is a useful corrective. It is also a reminder that most health twin work models physiology rather than process.
- Patient-journey and care-pathway modelling: Mapping the patient journey is standard improvement practice, and it remains the most direct way to surface where a pathway actually breaks. Process mining discovers the real process from event logs, which is precisely the situation a relationship management system creates, since it produces a timestamped log as a side effect of operating [19]. Applied to healthcare, it was demonstrated early against a hospital's own event logs, where the discovered pathway diverged from the documented one [10], and it has since been formalised into a methodology that handles the sequence variability and case heterogeneity that clinical settings produce [14]. This body of work establishes that journeys can be reconstructed from operational logs and that the reconstructed journeys differ substantially from documented pathways. What it does not do is simulate a specific patient forward and compare candidate interventions for them, which is the step this paper takes.
- State and time-to-event models of patient trajectories: Representing a patient as occupying one of a defined set of states, with probabilistic transitions between them, is an old and durable idea in medical decision making, and the practical guidance on building such models is long established [16]. That tradition supplies the vocabulary this paper uses for journey state and transition, and it also supplies the discipline of defining states so that they are mutually exclusive and operationally meaningful. It is worth being explicit about where the present work departs from it. Classical state models are typically population models used for decision analysis over cohorts. The twin described here is instantiated per patient and is initialised from that patient's own recorded history.
- Sequence models for clinical event prediction: Learning directly from sequences of clinical events, rather than from a flattened feature vector, has been shown to work at scale on electronic health record data across multiple prediction tasks [13]. That result matters here for one specific reason: it supports the design decision to preserve order and elapsed time in the journey representation rather than to reduce a history to counts and recency buckets. The engagement setting differs from the clinical prediction setting in the events it holds, which are largely interactional rather than physiological, but the argument for keeping the sequence intact carries across.
- Appointment adherence and no-show prediction: Predicting whether a patient will attend has a substantial literature of its own, and a systematic review of no-shows in appointment scheduling catalogues both the predictors used and the operational responses built on them [5]. That work is the closest existing analogue to one narrow slice of what the twin does. The difference is scope. No-show prediction answers a question about a single appointment. The twin is asked about a trajectory, of which an appointment is one node.
- Engagement, timing and behaviour change: The just-in-time adaptive intervention framework in mobile health formalises the idea that an intervention's effect depends on delivering the right support at the right moment, in the right amount, and that more contact is not automatically better [11]. That framework anticipates, in a behavioural-science register, one of the findings reported in Section 8. It provides the conceptual language for treating timing and dose as first-class decision variables rather than as campaign settings.
- Next-best-action and targeting effectiveness: In the customer relationship literature, the assumption that the highest-risk cases are the right cases to target has been directly challenged: targeting high-risk customers can be ineffective because risk and responsiveness to intervention are different quantities [1]. This is a load-bearing result for the present work. It is the formal statement of why a disengagement risk score is not by itself an engagement instruction, and why the twin returns a comparison across candidate actions rather than a ranked risk list.
- Simulation in healthcare operations: Discrete-event simulation has a long record in healthcare

performance modelling, and the review literature sets out both what it does well and the persistent gap between simulation studies and implemented change [9]. The what-if layer described in Section 6 borrows the framing of comparing candidate policies under a common model, while operating at the level of an individual journey rather than a service.

- Calibration of predicted probabilities: If a risk score is going to be used to decide whether to contact a patient, it has to mean what it says. The behaviour of predicted probabilities under different learning methods, and the correction of systematically distorted probabilities, is well characterised [12]. That work underpins the decision to make risk calibration an explicit validation dimension in its own right rather than a diagnostic checked in passing.
- Distribution shift in deployed health models: A model that behaved well on historical data can degrade when the population, the process or the recording practice moves underneath it, and the clinical implications of dataset shift have been set out directly for practising clinicians [7]. This is why Section 10 treats drift monitoring per service line and per cohort as part of the design rather than as an operational afterthought.
- Re-identification and the privacy of model outputs: The empirical record on re-identification attacks against health data is the reason the architecture in Section 9 removes direct identifiers before data reaches the modelling environment rather than relying on access control alone [6]. A second and less commonly addressed exposure is that a trained model and its outputs can themselves leak information about the records behind them, as membership inference attacks demonstrate [15]. Together these two lines of work motivate the position taken in Section 9: the simulation output needs protection comparable to the data it came from.
- Human oversight of decision support: The benefits and risks of clinical decision support systems, and the strategies that make them work in practice, are well surveyed [17]. The relevant lesson for this work is that a recommendation is a prompt for a person and not a substitute for one.
- Positioning within the author's prior work: The twin described here generates journey-state predictions and engagement recommendations. It does not deliver them. Delivery through a conversational interface [3] and real-time care recommendation delivery via agent-to-user interaction [4] are separate concerns addressed separately, and the governance framing that any such system operates within has also been treated on its own terms [2]. Those are consumers and constraints of the output described in this paper, not the contribution of it. The contribution here is the journey-state model, the forward simulation and the what-if evaluation.

4. What the Twin Models

The twin is a model of the patient journey as a dynamic personalised process. It is not a copy of the medical record, and the distinction is not cosmetic. A record copy would duplicate clinical content in a second system, inheriting all of that content's sensitivity and none of its clinical governance. What the twin holds instead is three things: where the patient currently is in the care and engagement lifecycle, how their past interactions produced that position, and what is likely to happen next under different engagement strategies.

The inputs it draws on span profile and demographic attributes, appointment history, care milestones, care gaps, medication events, relationship-management interaction history, service cases, patient portal activity, communication preferences and prior outreach responses. To those it adds a set of behavioural indicators, which are derived rather than recorded: missed appointments, slow responses, frequent rescheduling, declining engagement over time, and unresolved service issues. Clinical data is limited to what is necessary and authorised for the specific use case. That constraint is architectural, not advisory, and Section 9 describes how it is enforced.

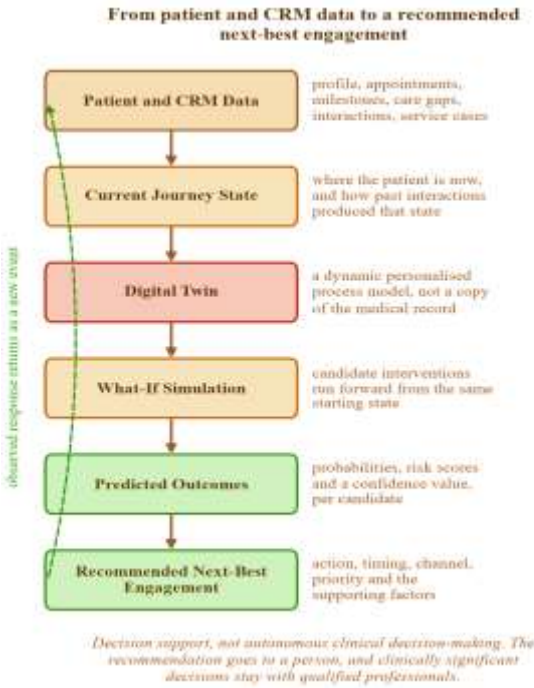

Fig 1: The Pipeline from Patient and Relationship-Management Data through Current Journey State, the Twin, What-If Simulation and Predicted Outcomes to a Recommended Next-Best Engagement. The Observed Response to Any Engagement Returns as a New Event, Which Updates the Journey State and Therefore the Twin

Sequence and context are preserved rather than summarised. A patient who was scheduled, then received a reminder, then confirmed, then rescheduled, then completed a visit, then had a follow-up marked as required and then completed it, has a journey with a shape. Reducing that

shape to a set of counts, such as one reschedule and two reminders, discards the information that matters most for predicting what comes next. The representation therefore keeps the ordered transition history and the elapsed time between transitions, which is consistent with the argument from sequence modelling on clinical event data that order carries signal [13].

Three kinds of output come out of the twin. The first is predictions and risk scores: the probability of the next state, the probability of disengagement, the probability that a care gap will remain open. The second is what-if simulations, which compare candidate interventions from the same starting state. The third is an engagement recommendation, which is not simply the highest-scoring action but a package: the next best action, when to take it, on which channel, at what priority, and which factors in the journey support that conclusion. The supporting factors are part of the output rather than an optional explanation, because a coordinator who cannot see why a recommendation was made cannot sensibly override it.

Figure 1 shows the pipeline. Two things about it are worth stating explicitly. First, the loop is real: the observed response to an engagement becomes a new event, updates the journey state, and therefore changes the twin. The twin is not a periodic batch score. Second, the terminal box is a recommendation, and it goes to a person. The system is decision support, not autonomous clinical decision-making. That boundary is a design constraint and it is not negotiable at the level of individual features.

5. Journey State Representation and Inputs

5.1. States and Transitions

The journey is represented as a set of defined states with transitions between them. The reference sequence for an appointment-centred journey runs Scheduled, then Reminder Sent, then Confirmed, then Rescheduled, then Visit Completed, then Follow-up Required, then Follow-up Completed. Real journeys deviate from that spine constantly, and the deviations are the interesting part.

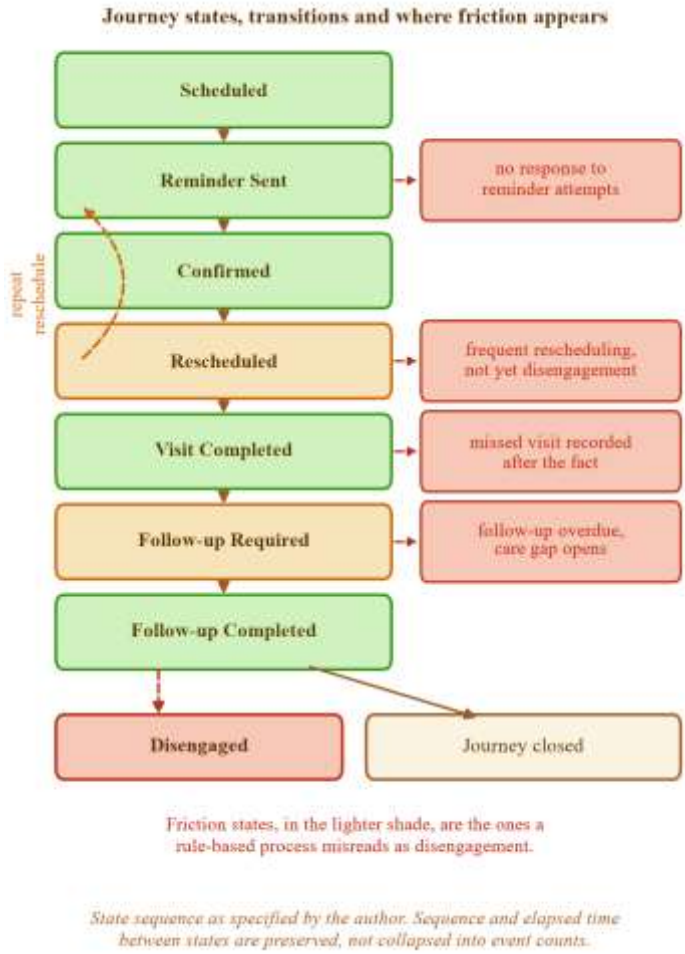

Fig 2: Journey States, Transitions And The Points At Which Friction Appears. Friction States, Shown In The Lighter Shade, Are The Ones A Rule-Based Process Tends To Misread As Disengagement. The State Sequence Is As Specified For This Work; Sequence And Elapsed Time Between States Are Preserved Rather Than Collapsed Into Event Counts

Figure 2 shows the spine together with the friction points. Four of them recur. A reminder can be sent and produce no response at all, once or repeatedly. A patient can reschedule, and then reschedule again, cycling back into the

reminder state without ever reaching a visit. A visit can be missed, which the system learns only after the fact, at which point the useful window for intervention has closed. And a follow-up can be marked as required and then sit, opening a care gap that nobody has yet acted on.

Defining states this way follows the tradition of state-transition modelling in medical decision making, where the requirement is that states be mutually exclusive, operationally meaningful and populated by observable events [16]. It differs from that tradition in instantiation: the model here is initialised per patient from that patient's own history rather than being run over a cohort.

Two terminal conditions matter for engagement. A journey can close normally, when the follow-up completes and no further action is required. Or a patient can reach a disengaged state, which is not a clinical status but an engagement status: the pattern of missed contacts, ignored outreach and abandoned follow-ups has crossed a threshold at which the existing engagement approach has stopped working. The distinction between reaching that state and merely passing through a friction state is exactly what Section 8 reports the twin got wrong at first.

5.2. Input Signal Domains and Derived Features

Table 1 sets out the input signal domains and the journey features derived from each. The right-hand column is where the modelling work sits. Raw signals are not fed forward as they arrive; they are converted into features that describe position, trajectory and responsiveness.

Table 1: Input Signal Domains And The Journey Features Derived From Each. Clinical Data Is Limited To What Is Necessary And Authorised For The Specific Use Case, And Derived Features Are Used In Preference To Raw Clinical Content Wherever A Derived Feature Is Sufficient

Input Signal Domain	Derived Journey Features
Profile and demographics	Cohort membership, service-line assignment, expected journey template, baseline contactability
Appointment history	Current journey state, transition history, attendance ratio, reschedule frequency, interval between scheduled and actual attendance
Care milestones	Position within the expected care sequence, milestones reached against milestones due, elapsed time since the last milestone
Care gaps	Open gap count, days overdue per gap, gap recurrence, whether a gap has previously been closed after outreach
Medication events	Refill regularity as an adherence proxy, gaps between expected and observed events, direction of change over recent periods
Interaction history	Contact frequency by channel, response rate by channel, time to respond, consecutive non-responses, trend in responsiveness
Service cases	Open case count, unresolved case age, whether an unresolved case coincides with a missed appointment or an abandoned follow-up
Portal activity	Recency and frequency of self-service actions, whether the patient self-serves in preference to being contacted, drop in activity
Communication preferences	Stated channel preference, stated contact window, consent state for each channel, divergence between stated and observed preference
Prior outreach responses	Outcome of each previous engagement, which channel and timing produced a response, which produced none, decay in response over repeated attempts
Behavioural indicators	Missed-appointment pattern, slow-response pattern, frequent-rescheduling pattern, declining-engagement trend, unresolved-issue flag

The last row deserves comment. Those five indicators are not events in the source systems. They are patterns computed across events, and they are the features that carry most of the predictive weight in the journey model. They are also the features that make the twin usable without identified clinical content, because a pattern such as declining engagement with a follow-up pending is a complete description of the situation for engagement purposes and reveals nothing about who the patient is. Section 9 returns to this point, because it is the mechanism that lets the twin operate on representations rather than on records.

5.3. Horizontal Extension

The representation is extended horizontally by defining new journey states, new events and new transition rules. Adding a service line, in other words, is a modelling exercise against the same framework rather than a new system. This is a design property of the representation. It is described here as such and not as a report of extension work already completed across all candidate service lines.

6. What-If Simulation and Next-Best Engagement

6.1. Running Candidate Interventions from a Common State

The what-if layer takes a journey state and a set of candidate interventions, runs each forward through the twin from that same starting state, and returns the predicted outcome distribution for each. The framing borrows from

discrete-event simulation practice in healthcare operations, where comparing candidate policies under a common model is the standard method and the standard caution is that the model's realism bounds the value of the comparison [9]. The difference is granularity: the comparison here is for one patient's journey rather than for a service's throughput.

The candidate set is deliberately small and operationally real. It consists of the interventions a coordinator could actually take: send a reminder on the current channel, send on a different channel, change the timing, escalate priority, route to a care coordinator for a human conversation, or take no immediate action and continue monitoring. That last option is a genuine candidate and not a null case, for reasons Section 8 explains.

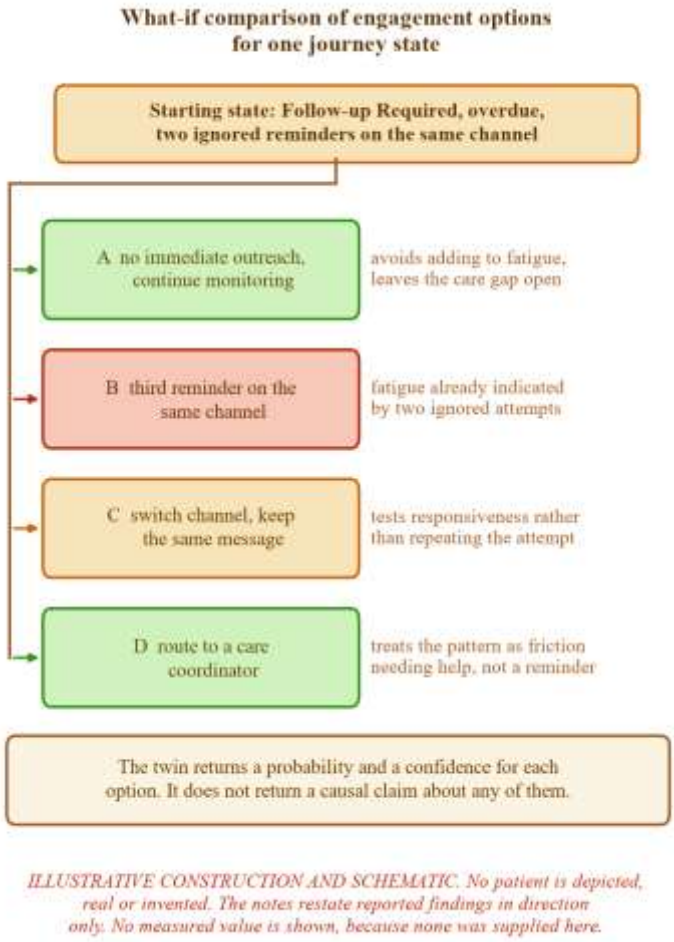

Fig 3: What-If Comparison of Engagement Options for One Journey State. This Is An Illustrative Construction And A Schematic. No Patient Is Depicted, Real Or Invented, And No Measured Value Is Shown, Because No Measured Comparison For This Specific Construction Was Produced. The Notes Restate Reported Findings In Direction Only

Figure 3 shows the structure of such a comparison for a constructed journey state: a follow-up required and overdue, with two reminders already ignored on the same channel. The figure is an illustrative construction and is labelled as one. It carries no numbers, because producing numbers for a constructed example would be fabrication. What it does show is the shape of the decision, which is the point: four candidate actions, each with a different mechanism, evaluated against the same state.

6.2. What the Output Is, and What It Is Not

For each candidate the twin returns a probability and a confidence. It does not return a causal claim. This is the single most important constraint on the what-if layer and it is worth stating without hedging: a simulated what-if outcome is never treated as proof that an intervention caused an improvement.

The reason is that the twin learns associations from recorded journeys, and recorded journeys carry the selection behaviour of whoever chose the interventions in them. A simulation that predicts a better outcome under a channel switch is predicting what tended to follow channel switches among comparable patients. That is useful and it is not the same as a causal effect. The customer relationship literature has made the equivalent point sharply in a different domain: high risk and high responsiveness are different quantities, and targeting on the first can be ineffective precisely because it is not the second [1].

Two treatments follow from that. Where historical comparison groups existed, predicted effects were compared against observed outcomes among similar patients, which gives an empirical check on whether the simulated direction held. Where no such comparison group existed, the output was labelled a scenario estimate requiring further validation, and it was not presented as a causal clinical conclusion. Both treatments are applied at the point of output rather than in documentation, so a coordinator sees which kind of statement they are looking at.

6.3. From Prediction to a Recommendation

A prediction is not an instruction. Turning one into the other requires deciding whether to engage at all, when, on which channel, at what priority and with how much human involvement. The just-in-time adaptive intervention framework treats timing and dose as decision variables in exactly this way, and its central premise, that support has to arrive at the right moment and in the right amount, matches what the evaluation found [11].

The recommendation therefore carries five fields rather than one. The action, which may be no action. The timing, expressed as a window rather than an instant. The channel, selected on predicted responsiveness rather than on a default. The priority, which positions this patient against everyone else in the queue. And the supporting factors, which name the journey features that drove the recommendation.

Generating a recommendation is where this paper stops. Delivering it, whether through a conversational interface [3] or through real-time recommendation delivery at the point of interaction [4], is a separate layer with its own design problems, and the governance framework any such system operates within is a separate matter again [2]. The twin produces the decision content. Those layers consume it.

7. Validation against Real Journeys

7.1. Temporal Holdout on Reconstructed Journeys

Validation used a temporal holdout on real historical journeys rather than on idealised pathways. Journeys were reconstructed chronologically from the event record. The twin was then initialised at an earlier point in a journey and asked to simulate forward without seeing later events. The simulated sequence was compared against the recorded sequence.

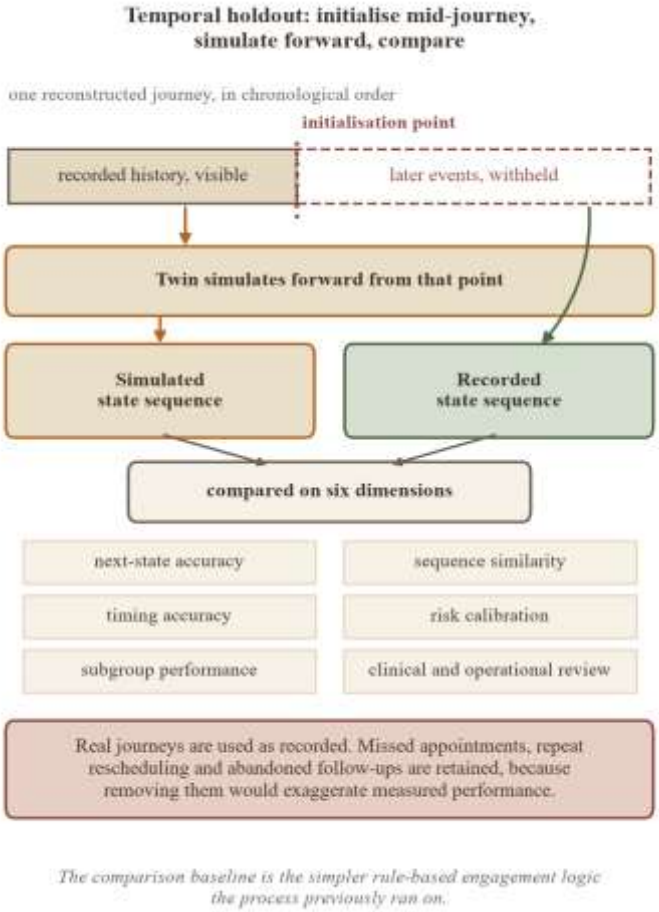

Fig 4: The Temporal-Holdout Design. The Twin Is Initialised At A Point Partway Through A Reconstructed Journey, Simulates Forward, And Its Simulated State Sequence Is Compared Against the Recorded Sequence on Six Dimensions. The Comparison Baseline Is The Simpler Rule-Based Engagement Logic The Process Previously Ran On

Figure 4 shows the design. The structure is deliberately unforgiving in two respects. First, the twin is initialised mid-journey rather than at the start, so it has to work from a partial history exactly as it would in operation. Second, everything after the initialisation point is withheld, including events that would have made the forward prediction easy.

One methodological decision shaped the results more than any other: messy real-world behaviour was deliberately retained. Missed appointments, repeat rescheduling, abandoned follow-ups and unresolved service cases were left in the evaluation set rather than filtered out as noise. Removing them would have exaggerated performance, because the journeys that are hard to predict are precisely the journeys that motivated the work. A model validated only on journeys that went to plan would report a number that told nobody anything useful.

7.2. The Six Validation Dimensions

A single aggregate accuracy score would have hidden more than it revealed, so the comparison was made on six dimensions, each testing a different property. Table 2 sets them out.

Table 2: The Six Validation Dimensions and What Each One Tests.

Validation Dimension	What it tests
Next-state accuracy	Whether the twin identifies the correct immediate next state from a given journey position, which is the narrowest and most local test of the transition model
Sequence similarity	Whether the simulated trajectory as a whole resembles the recorded trajectory, which catches models that are locally right and globally wrong
Timing accuracy	Whether transitions are predicted to occur at approximately the right time, not merely in the right order, since an engagement recommendation with the wrong timing is not actionable
Risk calibration	Whether stated probabilities match observed rates, so that patients assigned a 70 per cent disengagement probability disengage at about that rate rather than at some other rate
Subgroup performance	Whether accuracy holds across service lines and cohorts rather than being carried by the largest and best-recorded group
Clinical and operational review	Whether unusual transitions and outright failures make sense to people who know the pathway, which is the only dimension that can catch an error the metrics are blind to

Risk calibration warrants emphasis because it is the dimension that determines whether the output can be used at all. A score that ranks patients correctly but states probabilities that are systematically too high will cause a coordinator to treat moderate risk as urgent, and the whole prioritisation argument collapses. The behaviour of predicted probabilities across learning methods, and the need to check and correct them explicitly, is well documented [12], and calibration was treated here as a first-class result rather than as a diagnostic.

Subgroup performance was checked for a related reason. An aggregate that holds up because one high-volume service line dominates the evaluation set is not evidence that the model works for the others. Section 10 returns to this, because the subgroups where performance was weakest are also the subgroups where the model's limitations are structural rather than fixable by more training data.

7.3. Baseline

Performance was compared against the simpler rule-based baselines that the engagement process previously ran on: fixed reminder schedules, threshold-driven care-gap alerts and default channel selection. That is the right comparison, because it is the alternative the organisation actually had. It is also a comparison that should be read with care, since a rule-based baseline is easy to beat on any metric that rewards responsiveness to individual context, and the interesting question is not whether the twin beat it but on which dimensions and by how much. Section 11 reports the answer, as ranges.

8. Findings That Changed Engagement Strategy

The evaluation produced three findings that were not anticipated, and together they changed how engagement was approached.

The through-line is uncomfortable: more engagement did not always improve outcomes.

8.1. Timing Mattered More Than Message Count

A single well-timed message at a key journey transition outperformed multiple generic reminders. That result is easy to state and hard to act on inside a campaign-driven system, because campaign infrastructure is built to schedule volume, not to detect moments.

The mechanism is that a journey transition is a point at which a patient's disposition changes. The interval just after a follow-up is marked as required, or just before a scheduled visit that a patient has already rescheduled once, is a moment where a contact carries information the patient can use. A reminder sent on a fixed schedule may land at such a moment by coincidence. It usually does not. The just-in-time adaptive intervention literature had already established the general principle that the moment and the amount of support are decision variables in their own right [11]; what the

evaluation added was a way to detect the moment from journey state rather than from a schedule.

8.2. Engagement Fatigue

Repeated messages on the same channel reduced responsiveness, and the effect was strongest after several ignored attempts. Each additional contact on a channel that had already failed was not neutral. It made the next contact on that channel less likely to work.

This finding is the reason channel selection moved from a rule to a prediction. If responsiveness on a channel decays with unsuccessful attempts, then the correct response to two ignored reminders is not a third reminder on the same channel. It is either a different channel, a different kind of contact, or nothing at all for now. It also explains why a raw disengagement risk score is not an engagement instruction. A high score can mean a patient who needs contact, or it can mean a patient who has already been contacted past the point of usefulness. The targeting literature makes the general form of this argument directly: risk and responsiveness are different quantities and treating the first as a proxy for the second produces ineffective targeting [1].

8.3. Journey Friction Misread As Disengagement

The third finding was the one with the clearest operational consequence. Patterns that looked like disengagement were often friction. A patient who reschedules frequently, or who has an unresolved service issue sitting behind a missed appointment, is usually not withdrawing from care. They are encountering an obstacle. The appropriate response is help, not another reminder.

Rule-based logic cannot distinguish these cases, because at the level of event counts they look identical: contacts attempted, appointments not attended, follow-ups not completed. The journey representation can distinguish them, because friction has a shape. Repeated rescheduling with confirmations in between is a different pattern from silence. An unresolved service case coinciding with a missed appointment is a different pattern from a missed appointment on its own. This is the clearest single argument for preserving sequence and context in the representation rather than collapsing history into counts.

The no-show prediction literature captures part of this problem, in that predictors of non-attendance are well catalogued and the operational responses built on them are well described [5]. What a per-appointment framing cannot easily express is that the same non-attendance means different things depending on the trajectory it sits in.

8.4. The Strategic Shift, And the Case for No Outreach

Taken together, the findings moved engagement from campaign-driven outreach to journey-state-driven intervention. The operative question changed from which patients are due for a message to which patients are at a moment where an intervention would matter.

One consequence is that sometimes the best action was no immediate outreach. That is a legitimate recommendation, and treating it as such required a change of posture, because a system that measures engagement by volume of contact will always penalise it. The broader point is that personalisation is not only about message content. It is about whether to engage at all, when to engage, which channel to use, and how much human involvement the situation warrants.

9. Privacy Architecture for Simulated Patients

9.1. Minimum Content, Not a Record Copy

The twin holds the minimum needed to model the journey, not a copy of the record. Direct identifiers are removed or replaced with pseudonymous keys before data enters the modelling environment, and what the twin then works with is mainly derived journey features rather than raw clinical data. HIPAA and applicable privacy, security, governance and consent requirements apply throughout.

The empirical case for stripping identifiers before the modelling boundary rather than relying on access control inside it is well established: the record of re-identification attacks on health data shows that access control on an identified dataset is a single point of failure [6]. Removing the identifiers moves the problem, and the mapping that could reverse the pseudonymisation is held outside the modelling environment.

9.2. Three-Layer Separation

The key architectural decision was to separate the patient identity layer, the clinical data layer and the simulation layer.

Figure 5 shows the arrangement. The identity layer holds names, contact details, record numbers and the re-identification mapping. The clinical data layer holds authorised clinical detail, constrained to what the use case requires. The simulation layer holds journey states, derived behavioural features, transition history and simulation runs.

The effect of the separation is that the twin operates on representations rather than on people. A journey state for engagement purposes can be complete while consisting of nothing more than a pseudonymous token, a missed-appointment pattern, a declining-engagement trend and a pending follow-up. That representation supports the full set of predictions and simulations described in this paper. It does not identify anybody. The separation is what makes the earlier claim about derived features load-bearing rather than decorative: the features were designed so that the simulation layer would not need identity or raw clinical content to do its work.

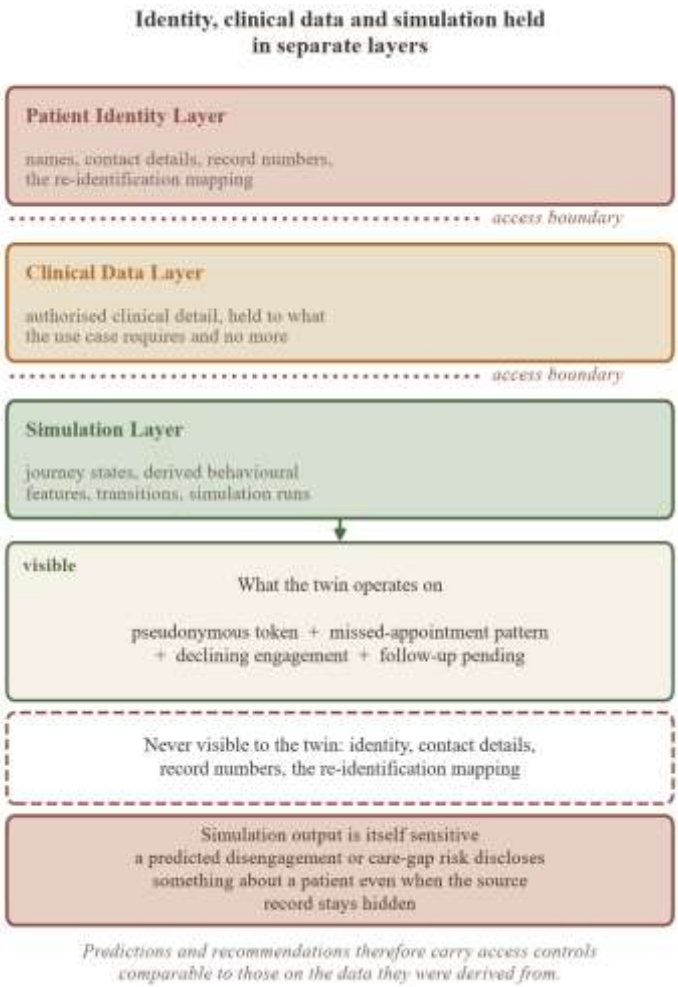

Fig 5: Identity, Clinical Data and Simulation Held in Separate Layers, With Access Boundaries between Them. The Twin Operates On A Representation Such As A Pseudonymous Token Together With A Missed-Appointment Pattern, A Declining-Engagement Trend And A Pending Follow-Up. Simulation Output Is Treated As Sensitive In Its Own Right, And Predictions And Recommendations Therefore Carry Access Controls Comparable To Those On The Data They Were Derived From

9.3. Simulation Output Is Itself Sensitive

One point is easy to miss and is worth foregrounding. The simulation output is sensitive in its own right. A predicted disengagement probability, or a predicted care-gap risk, reveals information about a patient even when the underlying record stays hidden. An inference is a disclosure.

This is not a hypothetical concern. Model outputs can leak information about the data behind them, and membership inference attacks are the clearest demonstration that a model's responses carry recoverable information about its training records [15]. The consequence adopted here is direct: predictions and recommendations are given access controls comparable to those on their source data. A prediction table is not treated as derived, low-sensitivity output merely because it contains no names.

The design principle behind the whole arrangement is worth stating in the form it was arrived at: simulate the patient journey without unnecessarily duplicating the patient.

9.4. Controls and What Each Mitigates

Table 3 lists the controls and what each one addresses. Several of these are requirements of the design rather than measured outcomes, and the table is a control map rather than an audit result.

Table 3: Privacy And Security Controls, And The Specific Exposure Each One Addresses. This Is A Control Map From The Design, Not An Audit Result

Control	What it mitigates
Data minimisation	Holding clinical content the engagement use case does not need, which creates exposure with no corresponding benefit
Pseudonymisation and tokenisation	Direct identification of a patient from the modelling environment, and the treatment of access control as the only barrier

Re-identification mapping held outside the modelling environment	Compromise of the modelling environment yielding both the data and the means to attribute it
Encryption in transit and at rest	Interception in movement between layers and exposure of data at rest in any layer
Role-based least-privilege access	Broad standing access to identity or clinical layers by roles whose work only requires the simulation layer
Environment isolation, with synthetic or de-identified data in lower environments	Real patient data reaching development, test or demonstration environments, which are the least controlled
Auditability of access and of simulation runs	Inability to answer who saw what, and which simulation produced a given recommendation
Retention limits	Indefinite accumulation of journey history and simulation output beyond the period it is useful for
Access controls on simulation output	Treating a predicted disengagement or care-gap risk as low-sensitivity output because it contains no identifiers, when the inference itself is a disclosure

10. Where the Model Fails and How Uncertainty Is Surfaced

10.1. What the Twin Cannot See

The twin is weak wherever outcomes depend on events that were never captured. Transport problems, work or family responsibilities, financial concerns, unexpected health events and changing personal preferences all shape whether a patient attends and responds, and none of them appears in the record. In the data they are indistinguishable from ordinary disengagement.

This is a structural limitation and not a data quality problem. No amount of additional operational data closes it, because the information does not exist in the operational systems at all. It sets a ceiling on achievable accuracy that is independent of model choice, and it means a proportion of predicted disengagement will always be misattributed. A patient with a transport barrier and a patient who has decided to stop engaging produce the same pattern of missed contacts.

10.2. Where Reliability Drops

Four situations degrade reliability in identifiable ways. New patients with little history give the twin too little to initialise from, and journey-state modelling is history-dependent by construction. Rare journey patterns are underrepresented in training, so the transitions the twin has learned least well are exactly the ones that occur least often and are most likely to matter clinically when they do. Patients moving between health systems produce journeys with gaps that look like inactivity but are activity recorded elsewhere. And small-data service lines cannot support the same confidence as high-volume ones, which is why subgroup performance was made a distinct validation dimension in Section 7 rather than folded into an aggregate.

Long-horizon simulation degrades in a way worth separating from the above. Uncertainty compounds at every transition, so a simulated journey many steps ahead is a distribution over increasingly divergent futures rather than a forecast. Long simulated journeys are therefore not presented as deterministic forecasts, and the further the horizon, the more the output functions as a scenario sketch than as a prediction.

The related exposure is that a model validated on historical journeys can drift as the population, the process or the recording practice moves. The clinical consequences of dataset shift are well set out, and the practical implication is that monitoring has to be continuous rather than a launch activity [7].

10.3. Compensations

Six mechanisms respond to these limits. Output is expressed as probabilities and confidence rather than as absolutes. Low-confidence cases are routed to care coordinators instead of triggering automatic intervention. The twin updates continuously as new events arrive, rather than being scored on a fixed cycle. Fallback business rules apply where data is insufficient, so a thin history produces the old behaviour rather than an unreliable new one. Monitoring runs by service line and by cohort, so degradation in a small service line is visible instead of averaged away. Drift checks run continuously against the assumption that the historical relationship still holds.

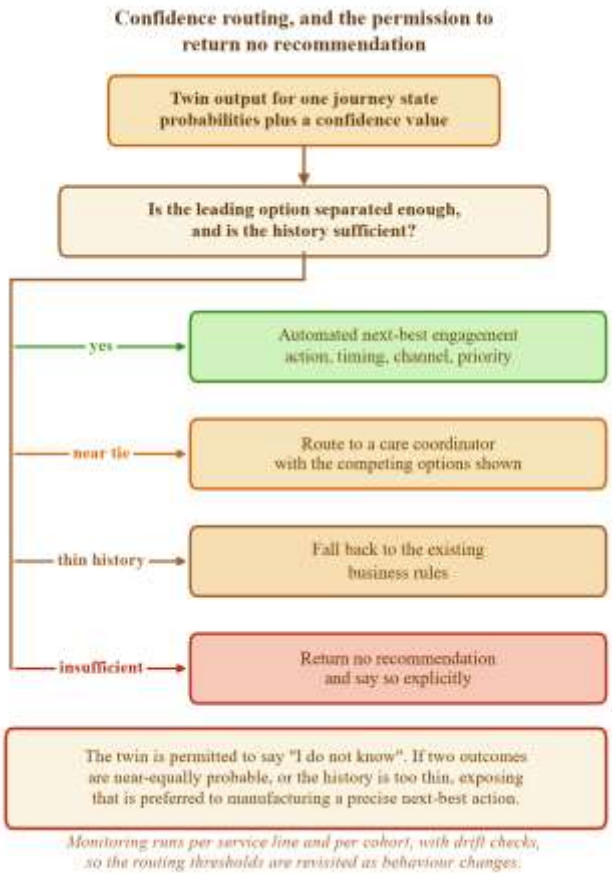

Fig 6: Confidence Routing. Separation Between The Leading Option And The Alternatives, Together With Sufficiency Of History, Determines Whether A Recommendation Is Issued Automatically, Routed To A Care Coordinator With The Competing Options Shown, Handed To Fallback Business Rules, Or Withheld With The Uncertainty Stated Explicitly

10.4. The Twin Must Be Able To Say That It Does Not Know

Figure 6 shows the routing rule, and it encodes a design principle that matters more than any individual accuracy figure: the twin must be able to say that it does not know.

If two outcomes are near-equally probable, or the available history is insufficient, the correct output is an explicit statement of that uncertainty rather than a manufactured next-best action. A system that always returns a precise recommendation trains its users to trust recommendations that do not deserve trust, and the erosion is invisible until something goes wrong. Exposing the uncertainty is preferable, even though it produces an output that looks less useful.

Four routes follow from the confidence check. A clearly separated leading option with sufficient history produces an automated next-best engagement. A near tie routes to a care coordinator with the competing options shown, so the human sees the ambiguity rather than a false resolution of it. A thin history falls back to existing business rules. Insufficient basis returns no recommendation and says so. Keeping a person in the loop for the ambiguous cases is consistent with the accumulated guidance on making decision support work in practice, which repeatedly finds that oversight is what determines whether such systems help [17].

11. Evaluation Results

Table 4 reports the results of the research evaluation against the existing rule-based engagement process. Before reading it, three statements about its status:

- These are research-evaluation results: They are not audited production outcomes, and they should not be read or cited as production outcomes.
- Every figure is a range and is reported as a range: No midpoint is given for any of them, because no point estimate was substantiated. A reader who reduces any of these ranges to a single number is asserting something this paper does not.
- The evaluation scale is a research-evaluation scale: About 50,000 patient journeys across five service lines, with an architecture built to support 8 to 10 facilities on a shared platform. That is the scale of the evaluation and the design capacity of the architecture, not a verified deployed footprint.

Table 4: Research-Evaluation Results, Compared Against The Existing Rule-Based Engagement Process. These Are Research-Evaluation Results And Not Audited Production Outcomes. Every Value Is Reported As A Range Because No Point Estimate Was Substantiated

Measure, Against The Existing Rule-Based Engagement Process	Change Observed In The Research Evaluation
Engagement and response rate	Improved by about 18 to 25 per cent
Appointment adherence	Improved by about 12 to 18 per cent
Care-gap and follow-up completion	Improved by about 15 to 22 per cent
Engagement response time, prioritised cases	Reduced by about 25 to 35 per cent
Unnecessary outreach	Reduced by about 15 to 20 per cent
Engagement-related operational efficiency	Improved by about 10 to 18 per cent

11.1. Reading the Results

The pattern across the six measures is more informative than any one of them. Engagement and response rate improved by about 18 to 25 per cent while unnecessary outreach fell by about 15 to 20 per cent. Those two move in opposite directions under a volume strategy: contacting more patients raises absolute response counts and raises unnecessary contact at the same time. Moving together is the signature of a different mechanism.

The benefit came from fewer, more relevant interventions rather than from contacting more patients. That is the single most important interpretive statement about Table 4, and it follows directly from the findings in Section 8. If timing matters more than count, if repeated contact on one channel decays in effectiveness, and if a proportion of apparent disengagement is friction requiring help rather than a reminder, then the way to improve response is to send less and send it better.

The reduction of about 25 to 35 per cent in engagement response time applies to prioritised cases specifically, and the qualifier is load-bearing. It is not a claim that all engagement got faster. It is a claim that the cases the twin flagged as needing attention were reached faster, which is what prioritisation is for. The improvement of about 10 to 18 per cent in engagement-related operational efficiency is the narrowest of the six in range and is also the one most dependent on how a given operation was working beforehand.

11.2. The Scale Figures

About 50,000 patient journeys were evaluated across five service lines: primary care, cardiology, diabetes and chronic care management, diagnostic and imaging, and post-treatment follow-up. Those five were chosen because they have genuinely distinct engagement patterns, journey lengths, appointment frequencies and disengagement risks. A twin that works only where journeys are short and regular would not have been tested by a homogeneous selection.

The architecture was built to support 8 to 10 facilities on a shared platform, with each facility keeping its own operational features while sharing the twin framework. That is a statement about design capacity. The horizontal scaling mechanism, by which new journey states, events and transition rules define an additional service line, is likewise a design property of the representation described in Section 5.

11.3. What These Numbers Do Not Establish

They do not establish that the twin caused the improvements in any specific clinical sense. They are a comparison against a rule-based baseline in a research evaluation, and the appropriate reading is that the journey-state approach outperformed fixed-schedule logic on these six measures at this evaluation scale. They do not establish that the ranges would reproduce in another organisation, whose baseline process, data completeness and service mix would differ. And they do not establish anything about long-horizon accuracy, for the reasons given in Section 10.

12. Preconditions for Adoption

The evaluation surfaced a set of conditions without which this approach does not work. They are stated here as preconditions drawn from the work rather than as a report of an adoption programme.

- A unified longitudinal view: Patient, appointment, interaction and outcome data has to be linkable by a consistent patient identifier, with reliable timestamps, in enough completeness to reconstruct what happened and when. The requirement is reconstructability, not perfection. Journeys with gaps can still be modelled; journeys that cannot be ordered in time cannot be modelled at all, because sequence is the substrate. This is the same prerequisite that process mining on operational logs depends on [19].
- Clearly defined journey states and measurable outcomes: The states have to be defined before the modelling starts, and the outcome the work is trying to move has to be measurable. The practical form of this is to begin from a specific operational problem rather than from a decision to build a digital twin. A project that starts from the technology will produce a model without a decision attached to it.
- Enough historical volume and variation, including unsuccessful journeys: Training on ideal pathways defeats the purpose. The journeys that the twin exists to handle are the ones that went wrong: the reschedules, the abandoned follow-ups, the ignored outreach. A training set curated for cleanliness produces a model that performs well on the cases nobody needed help with. This is the same argument that governed the validation design in Section 7, applied one stage earlier.
- Privacy, security and governance in place before modelling: These are preconditions and not parallel

workstreams. The three-layer separation described in Section 9 is difficult to introduce into a modelling environment that has already been built on identified data, because by then the identified data has spread. Establishing the boundaries first is cheaper and safer than retrofitting them.

- A connected, actionable workflow: Predicting risk is worthless if nobody can act on the prediction and record what happened next. Two things are needed: a route from recommendation to action, and a route from action back to the event record. The second is what closes the loop shown in Figure 1, and without it the twin cannot learn from its own recommendations.
- Start narrow: One high-volume journey, one measurable problem, a small set of candidate interventions. Validate against historical results, then against prospective results, and only then expand. The staged sequence matters because the failure mode of an ambitious first deployment is not a wrong prediction; it is an unusable system that nobody trusts and nobody can debug.

13. Limitations

- The improvement figures are research-evaluation results and not audited production outcomes: Every value in Table 4, and every improvement figure stated anywhere in this paper including in the abstract, comes from a research evaluation. None has been audited as a production outcome, and none should be cited as one.
- Ranges are reported as ranges because point estimates were not substantiated: No midpoint, narrowed interval or single-value summary of any range in this paper is supported by the underlying work. Where a reader wants a point estimate, the honest answer is that the evidence does not supply one.
- The evaluation scale is not a deployed footprint: About 50,000 patient journeys across five service lines describes the evaluation. The figure of 8 to 10 facilities describes what the architecture was built to support on a shared platform. Neither figure is a verified count of production usage.
- Simulated what-if outcomes are scenario estimates, not causal evidence: A simulated improvement under a candidate intervention is not proof that the intervention causes the improvement. Where historical comparison groups existed, predicted effects were checked against observed outcomes among similar patients. Where they did not, the output is a scenario estimate requiring further validation.
- Unobserved life events limit prediction and are indistinguishable from disengagement: Transport, work and family constraints, financial pressure, unexpected health events and changing preferences are not in the data. They produce the same observable pattern as a decision to disengage, and no modelling choice resolves the ambiguity.
- Long-horizon simulation degrades: Uncertainty compounds at each transition. Multi-step simulated journeys are distributions over diverging futures and are not presented as deterministic forecasts.
- New patients, rare patterns, cross-system movement and small service lines are weaker: Thin history gives the twin little to initialise from. Rare transitions are the least well learned. Care delivered in another health system appears as an unexplained gap. Small service lines cannot support the confidence that high-volume ones can.
- Deployment-time drift is a live risk rather than a settled one: The relationships learned from historical journeys can move as the population, the process or the recording practice changes. Monitoring by service line and cohort and continuous drift checking are the mitigations; they reduce the risk and do not remove it.
- Baseline dependence: The comparison is against the rule-based engagement logic previously in place. Organisations with a more sophisticated baseline should expect smaller differences, and the ranges in Table 4 should not be read as transferable effect sizes.
- No patient-level evidence is presented: No patient record, case history or vignette appears in this paper. The single journey state used illustratively in Section 6 is a construction and is labelled as one on Figure 3. Nothing in this paper should be read as a clinical case report.
- Clinically significant decisions remain with qualified professionals: The system is decision support. It recommends engagement actions to people, and it is not a substitute for clinical judgement on any matter of care.

14. Conclusion

Healthcare relationship management systems are good at recording what happened and poor at anticipating what happens next, and that asymmetry is what makes engagement reactive. This paper described a digital twin that models the patient journey as a dynamic personalised process, holding journey state, transition history and derived behavioural signals rather than duplicating the medical record, and using forward simulation to compare candidate interventions before any of them is taken.

The research evaluation, on about 50,000 reconstructed patient journeys across five service lines with an architecture built to support 8 to 10 facilities, found improvements against the existing rule-based process across six measures, each reported as a range: engagement and response rate by about 18 to 25 per cent, appointment adherence by about 12 to 18 per cent, care-gap and follow-up completion by about 15 to 22 per cent, engagement response time for prioritised cases reduced by about 25 to 35 per cent, unnecessary outreach reduced by about 15 to 20 per cent, and engagement-related operational efficiency by about 10 to 18 per cent. These are research-evaluation results, not audited

production outcomes, and the ranges are the finding rather than an approximation of one.

The findings that changed practice were not the numbers. They were that timing mattered more than message count, that repeated contact on a single channel eroded responsiveness, and that journey friction was routinely misread as disengagement so that patients who needed help received another reminder instead. Together those moved engagement from campaign-driven outreach to journey-state-driven intervention, and made no immediate outreach a legitimate recommendation.

Two design commitments carry the rest of the work. The first is the separation of patient identity, clinical data and simulation into distinct layers, so that the twin operates on a representation such as a token, a missed-appointment pattern, a declining-engagement trend and a pending follow-up, without knowing who the patient is; and the recognition that simulation output is itself sensitive, since a predicted risk discloses something about a patient even when the record stays hidden. The principle is to simulate the patient journey without unnecessarily duplicating the patient. The second is that the twin must be able to say that it does not know. When two outcomes are near-equally probable or the history is too thin, surfacing that uncertainty is better than manufacturing a precise next-best action, and the case goes to a person.

The wider argument is that personalisation in healthcare engagement is not primarily a content problem. It is a set of decisions about whether to engage at all, when, on which channel, and with how much human involvement. A journey twin is useful because it makes those decisions inspectable before they are taken, and because it can decline to make them.

References

[1] Ascarza, E. (2018). Retention Futility: Targeting High-Risk Customers Might be Ineffective. *Journal of Marketing Research*, 55(1), 80-98. doi:10.1509/jmr.16.0163

[2] Dabbara, B. N. (2026). Ethical AI Governance in Healthcare CRM Systems. *International Journal on Science and Technology*, 17(2). doi:10.71097/ijsat.v17.i2.11021

[3] Dabbara, B. N. (2026). Integrating Conversational AI into Healthcare CRMs for Patient Engagement Optimization. *International Journal of Emerging Trends in Computer Science and Information Technology*, *7*(1), 16-18. https://doi.org/10.63282/3050-9246.IJETCSIT-V7I1P103

[4] Dabbara, B. N. (2026). Real-Time Care Recommendations in Healthcare Using Agent-to-User Interaction (A2UI) on the Salesforce Ecosystem. *International Journal of Innovative Research and Creative Technology*, 12(3). doi:10.62970/ijirct.v12.i3.2606012

[5] Dantas, L. F., Fleck, J. L., Cyrino Oliveira, F. L., and Hamacher, S. (2018). No-shows in appointment scheduling – a systematic literature review. *Health Policy*, 122(4), 412-421. doi:10.1016/j.healthpol.2018.02.002

[6] El Emam, K., Jonker, E., Arbuckle, L., and Malin, B. (2011). A Systematic Review of Re-Identification Attacks on Health Data. *PLoS ONE*, 6(12), e28071. doi:10.1371/journal.pone.0028071

[7] Finlayson, S. G., Subbaswamy, A., Singh, K., Bowers, J., Kupke, A., Zittrain, J., Kohane, I. S., and Saria, S. (2021). The Clinician and Dataset Shift in Artificial Intelligence. *New England Journal of Medicine*, 385(3), 283-286. doi:10.1056/NEJMc2104626

[8] Grieves, M., and Vickers, J. (2017). Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In *Transdisciplinary Perspectives on Complex Systems*, pages 85-113. Springer International Publishing. doi:10.1007/978-3-319-38756-7_4

[9] Günal, M. M., and Pidd, M. (2010). Discrete event simulation for performance modelling in health care: a review of the literature. *Journal of Simulation*, 4(1), 42-51. doi:10.1057/jos.2009.25

[10] Mans, R. S., Schonenberg, M. H., Song, M., van der Aalst, W. M. P., and Bakker, P. J. M. (2008). Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital. In *Biomedical Engineering Systems and Technologies*, Communications in Computer and Information Science, pages 425-438. Springer Berlin Heidelberg. doi:10.1007/978-3-540-92219-3_32

[11] Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., and Murphy, S. A. (2018). Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Annals of Behavioral Medicine*, 52(6), 446-462. doi:10.1007/s12160-016-9830-8

[12] Niculescu-Mizil, A., and Caruana, R. (2005). Predicting good probabilities with supervised learning. In *Proceedings of the 22nd international conference on Machine learning - ICML '05*, pages 625-632. ACM Press. doi:10.1145/1102351.1102430

[13] Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., et al. (2018). Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*, 1(1). doi:10.1038/s41746-018-0029-1

[14] Rebuge, Á., and Ferreira, D. R. (2012). Business process analysis in healthcare environments: A methodology based on process mining. *Information Systems*, 37(2), 99-116. doi:10.1016/j.is.2011.01.003

[15] Shokri, R., Stronati, M., Song, C., and Shmatikov, V. (2017). Membership Inference Attacks Against Machine Learning Models. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 3-18. IEEE. doi:10.1109/SP.2017.41

[16] Sonnenberg, F. A., and Beck, J. R. (1993). Markov Models in Medical Decision Making: A Practical Guide. *Medical Decision Making*, 13(4), 322-338. doi:10.1177/0272989X9301300409

[17] Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., and Kroeker, K. I. (2020). An overview of clinical decision support systems: benefits, risks, and strategies for success. *npj Digital Medicine*, 3(1). doi:10.1038/s41746-020-0221-y

[18] Tao, F., Zhang, H., Liu, A., and Nee, A. Y. C. (2019). Digital Twin in Industry: State-of-the-Art. *IEEE Transactions on Industrial Informatics*, 15(4), 2405-2415. doi:10.1109/TII.2018.2873186

[19] van der Aalst, W. (2016). *Process Mining*. Springer Berlin Heidelberg. doi:10.1007/978-3-662-49851-4

[20] Venkatesh, K. P., Raza, M. M., and Kvedar, J. C. (2022). Health digital twins as tools for precision medicine: Considerations for computation, implementation, and regulation. *npj Digital Medicine*, 5(1). doi:10.1038/s41746-022-00694-7.