



RESEARCH ARTICLE

AI-Driven Healthcare Management and Service Quality: Examining the Effects of Digital Transformation on Patient Satisfaction — A Simulation-Based Framework Linking Clinic Operations to SERVQUAL Perceptions

*Dr. Sulabh Narayan

*Assistant Professor, Department of Management, Sumandeep Vidyapeeth Deemed to be University, Vadodara, Gujarat.
narayansulabh.12@gmail.com

ABSTRACT

Hospitals are increasingly turning to AI for appointment scheduling, patient triage and flow and are anticipating more positive outcomes for satisfaction. Will this happen will depend on the impact that changes in operation have on the perceived service quality by patients. This paper proposes and implements a two-layer model, which connects them. The operational layer is a discrete-event simulation of an outpatient clinic executed as a 2^3 leveled factorial experiment consisting of 500 simulated days (per scenario) using common random numbers. It assesses 3 different AI interventions: no show prediction and targeted reminders (NS), AI enabled triage (TR), and digital check in with pre-visit intake (DC). The simulated waiting times are transformed into top-box satisfaction through a response model with Probabilistic Sensitivity Analysis. The perceptual layer is a SERVLOAD based survey model with PLS estimated and consistent PLS (PLSc). All its inference properties are based on the Monte Carlo simulation of synthetic respondents; no patient data are used or claimed. NS increased mean waiting time by 5.5 min while delivering 10.0 additional patients per session, which was partially mitigated by the NS×DC interaction (−1.32 min), and by DC (−4.5 min). Consequently, the combined system's satisfaction level was located close to zero (+0.3 percentage points; 95% interval between minus 2.1 and plus 2.8), and DC alone increased the satisfaction level of 98% of the draws. TR was significant only at 115% load where it reduced wait time for urgent patients by 2.5 min. Group-wise reminder quota and assisted check-in closed the gap to patients' satisfaction at +1.2 points, but a less accurate model for patients with low digital literacy did so. Within the survey layer, PLS was 9% off the truth for the estimates of the paths, while PLSc was almost unbiased. These small hypothesized effects ($\beta \neq 0.10$) required several hundred or more than 1,000 respondents to reach 80% power, far more than are commonly believed to be necessary by the rules of thumb. The study concludes with design guidance for AI based clinics and a patient survey analysis protocol which is ready for use in future surveys.

Keywords: Artificial intelligence, Healthcare management, Digital transformation, Patient satisfaction, Service quality, SERVQUAL, Discrete-event simulation, No-show prediction, PLS-SEM, Consistent PLS, Statistical power, Health equity.

INTRODUCTION

Healthcare organisations are going through a digital transformation that goes well beyond electronic records [1], [2]. Today, Machine-learning model predicts patients who are at risk of missed appointments [3,4], prioritises patients at triage [5], identifies patient high-risk groups [6] and clinical decisions [7-9]. There are typically two arguments given to justify these investments: A more efficient hospital and more satisfied patients. The first one has to do with operations and the second one has to do with perception. A connection between them is simply taken for granted more so than investigated.

It's not a given that that's a connection. In service-quality theory, satisfaction is generated by the service that patients perceive, broken down into dimensions: reliability, responsiveness, assurance, empathy and tangibles [10], [11] that have been adapted to hospitals [12].

*Assistant Professor, Department of Management, Sumandeep Vidyapeeth Deemed to be University, Vadodara, Gujarat.

Email-id: narayansulabh.12@gmail.com

Corresponding Author: Dr. Sulabh Narayan Research Assistant Professor, Department of Management, Sumandeep Vidyapeeth Deemed to be University, Vadodara, Gujarat.

DOI: <https://doi.org/10.63856/ijis/v2i10/02>

How to cite this article: Narayan, S., (2026). AI-Driven Healthcare Management and Service Quality: Examining the Effects of Digital Transformation on Patient Satisfaction — A Simulation-Based Framework Linking Clinic Operations to SERVQUAL Perceptions. *International Journal of Integrative Studies*, 2(10), 29-46.

Source of support: Nil

Conflict of interest: None.

Received: 15/09/2026 **Revised:** 25/09/2026 **Accepted:** 05/10/2026 **Published:** 15/10/2026

A change in another operational indicator due to an AI system's improvement of another operational indicator. For instance, higher throughput through overbooking and based on the prediction of no-shows can increase waiting times and waiting times are strongly correlated with patient satisfaction [13]. Digital tools can also have a two-fold impact on patients with poor digital literacy, and inequities can be coded in algorithms [14], [15]. Lastly, there is predominantly cross-sectional survey data, analysed using a structural equation model. The credibility of such studies depends on quality of measurement, sample size and the choice of estimator and these are rarely considered prior to data collection.

This paper will try to resolve both the gaps and provide a reproducible framework, which is transparent and simulation based. Any data used in this study have been created by the open code of the authors under clearly defined assumptions. There is no mention of any actual patient records or of any survey respondent; no claim of having an actual effect on real patients. The framework should be calibrated and tested with actual hospital and survey data and the code is developed with that in mind.

Research questions are:

- **RQ1:** What is the impact of AI no-show management, AI-assisted triage and digital check-in (alone and in combination) on waiting time, urgent access,

throughput and overtime in an outpatient clinic?

- **RQ2:** What impact do those operational changes have, under explicit and under uncertain assumptions, on patient satisfaction and who benefits/damages?
- **RQ3:** How big should be the sample size and what is the suitable estimator in order to test a SERVQUAL-based model of digital transformation and satisfaction using sufficient power and low bias?
- **RQ4:** What is the robustness of those inferences to Likert-type measurement, to ceiling effects and to the common method variance?

These contributions are (i) the integration of operations and perception in simple and complex causal and satisfaction models (Fig. 1), including a SERVQUAL structural model with discrete-event simulation; (ii) an extensive factorial analysis of three AI interventions, quantifying main effects and interactions, with extraction of an efficient frontier for risk-based overbooking; (iii) a probabilistic sensitivity analysis of the satisfaction consequences and an equity analysis by digital literacy; and (iv) the development of a 34-item survey instrument and release of a new, original numpy implementation to conduct PLS-SEM and PLS_c, together with Monte Carlo evidence on bias, power, coverage and robustness, providing concrete sample-size guidance for future patient surveys.

Operational layer (discrete-event simulation)

Perceptual layer (PLS-SEM survey model)

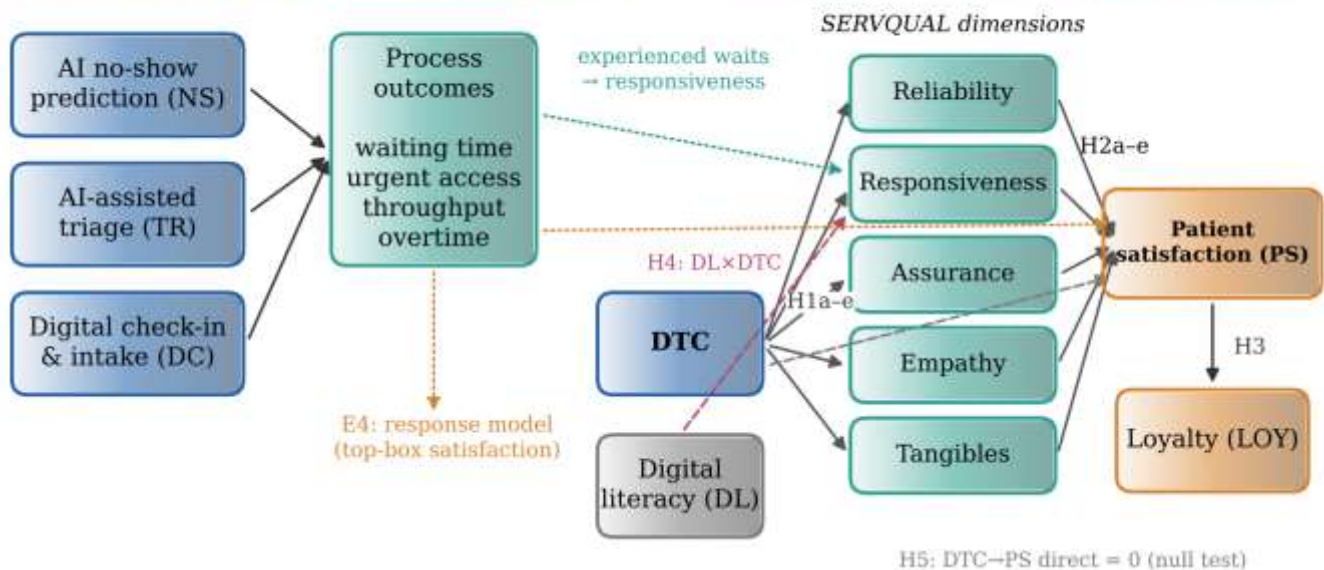


Fig. 1. Conceptual framework. The operational layer (left) is tested using Discrete Events Simulation, while the perceptual layer (right) is based on SERVQUAL and the hypotheses H1, H2, H3, H4 and H5 are tested. The dotted lines depict the relationship between simulated process outcomes (e.g., statements, objectives, and methods) and "responsiveness perceptions" (RP), and the satisfaction response model (SR model) (Section III-C).

2. Background and Hypotheses

2.1 AI and Digital Transformation in Healthcare Management

Digital transformation is when digital technologies are used as the catalyst to make a massive change on how a company

creates value [18]. In hospitals, these can span across the use of electronic record-keeping, patient portals, or even predictive analytics in hospital operations [1], [2], [6]. For service management, the most pertinent examples of AI are the prediction of "no-show" by means of machine learning

models (which have reported a useful discrimination [3]), and the separation of patients by means of "electronic triage" (which can do the separation of patients by clinical risk better than conventional scales [5]). Systems for making appointments have been a topic of studied focus in operations research for a while [19] and [20]. Short message reminders can enhance attendance and there are many cases of no show and its high cost [4]. But, adoption is socio-technical: Theories like TAM [22] and UTAUT [23] emphasize the ability of users and the compatibility of the organisation and the framework like NASSS [24] is more applicable for the introduction of technology from prototypes to clinical impact and there are known difficulties for the uptake of AI systems [25].

2.2 Hospital Patients Satisfaction and the Roles of Services

SERVQUAL considers the five dimensions of service quality, namely- reliability, responsiveness, assurance, empathy, tangibles [10]. Its perception-only form (SERVPERF) sometimes is found to perform as well or better [11]. Length measurements align with the Donabedian's structure–process–outcome (S–P–O) perspective of quality in alignment with hospital adaptations [12]. Perceived quality ultimately reflects satisfaction and

behavioural intentions like loyalty, recommendation etc. [27]. There have been many settings where telehealth/digital services have been proven to be correlated with satisfaction, and the impact of these services is contingent upon patients' ability to use them. eHealth literacy is an amenable construct [29].

2.3 Hypotheses

Digital-transformation capability (DTC) is the ability to notice how well a hospital's digital and AI-driven processes contribute to the visit, as experienced by the patient. DTC is expected to make improvement on all five SERVQUAL dimensions, with the greatest improvements in responsiveness (shorter wait, faster check-in) and the digital interface (tangibles), and the least improvements in empathy, which is more related to human interaction. The dimensions should breed satisfaction which in turn leads to customer loyalty. Digital Literacy (DL) should benefit the health and care of more patients, since those with higher literacy should be better able to use DTC. If the five dimensions provide a full picture of the mechanism, then there is no direct effect of DTC on satisfaction. The hypotheses along with the population values in the simulation study, are given in Table I. In the SMR1 it is not an estimation, but a design value.

Table 1: Hypotheses and Population Values Used For Simulation

Hyp.	Relationship	β_{pop}	Rationale
H1a-e	DTC $\rightarrow$ REL, RESP, ASR, EMP, TAN	.35, .40, .25, .10, .45	process digitisation
H2a-e	REL, RESP, ASR, EMP, TAN $\rightarrow$ PS	.25, .20, .15, .20, .10	SERVQUAL $\rightarrow$ satisfaction
H3	PS $\rightarrow$ LOY	.60	behavioural intentions
H4	DL $\times$ DTC $\rightarrow$ RESP	.125 (std.)	literacy moderation
H5	DTC $\rightarrow$ PS (direct)	0 (null)	full mediation
H6	DTC $\rightarrow$ SQ $\rightarrow$ PS (indirect)	0.270 total	mediation via SERVQUAL
Control	DL $\rightarrow$ RESP	.15	—

Population SERVQUAL disturbances correlate at 0.30; $\text{corr}(\text{DTC}, \text{DL}) = 0.30$. Values follow Cohen's small/medium conventions and serve to evaluate designs, not to describe any hospital.

3. Methodology

The methodology has six stages (Fig. 2). Stages S1–S3 address RQ1–RQ2 through operational simulation; stages S4–S6 address RQ3–RQ4 through measurement design and statistical simulation.

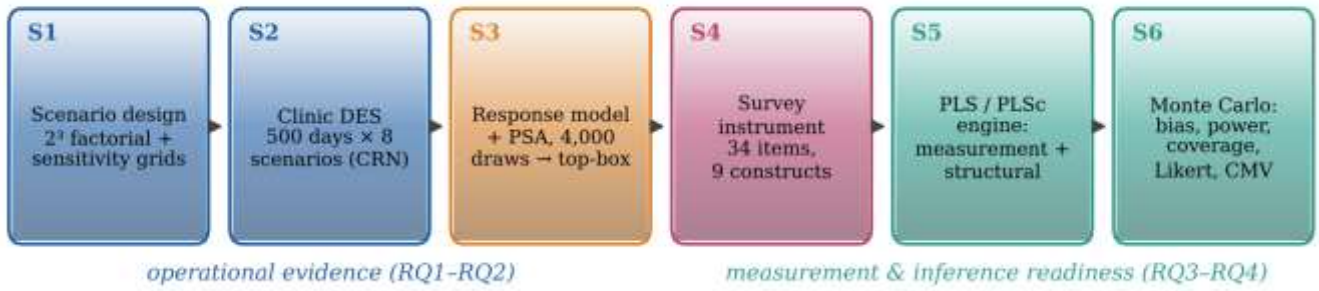


Fig. 2. Research pipeline. All numbers reported in this paper are computed from the code released (des.py, pls.py, popmodel.py, experiments.py).

3.1 Clinic Discrete-Event Simulation

The DES is one outpatient session of $T = 480$ min, 3 pooled physicians, 2 registration clerks and 96 appointments per session (one per physician per 15 min). Walk-ins have a Poisson arrival process with average of 25% urgent cases and 1.25 cases per hour. There are on average 8 minutes (+/- 10) early registered patients (scheduled) and 5% are registered with urgent needs. The registration and consultation times are lognormal (Table 2). Every patient i has a no show probability of $r_i \sim N(0,1)$ and a low-digital-literacy indicator ℓ_i (35% prevalence):

$$p_i = \sigma(\mu + s \cdot r_i + \delta \cdot \ell_i), \quad (1)$$

with $\mu = -2.55$, $s = 1.6$ and $\delta = 0.5$, which gives a baseline no-show rate of 15.9%, within the range reported in the literature [4]. The AI no-show model is a noisy score given by the model: $\hat{s}_i = r_i + \eta_{ei}$, with $\eta \sim N(0,1)$. The noise η is determined by bisection in order to reach a desired score corresponding to a given ROC AUC, which is defaulted to 0.78, the desired score corresponding to an oracle AUC is 0.83.

NS policy. 30% of the highest-scoring patients are targeted with reminders, lowering their risk of not showing up by 30% (relative). The clinic then oversubscribes K more patients in the time slots with the maximum expected no show probability $\hat{p}_i$:

$$K = \text{round}(\kappa \cdot \sum_i \hat{p}_i), \quad \kappa = 0.6. \quad (2)$$

TR policy. Patients with a flag are given priority in the

physicians queue - non-pre-emptive. Conventional triage nurse sensitivity is 0.7, false positive rate is 0.04, with a reported improvement in electronic triage of 0.9 and 0.06. DC policy. 85% of the high-literacy patients and 35% of the low-literacy patients use digital check-in. Adopters avoid having to pass the desk (1 min) and they also have consultations 0.75 min shorter, as their history is gathered beforehand. Time waiting is defined as the time from the later of arrival time and appointment time:

$$W_i = S_i - \max(A_i, \tau_i), \quad (3)$$

where S_i is the time of the consultation start, A_i the arrival time and τ_i is the appointment time ($\tau_i = A_i$ for walk-ins). The 8 (2^3) different scenarios are each simulated over 500 days. Common random numbers are used: Each stochastic quantity of a day is sampled from streams seeded only by the day index, and scenario contrasts are paired and variance is small. The factorial effects are estimated daily using the ± 1 coding,

$$\hat{E}_j = \frac{1}{4} \sum_{s=18} \text{cs}_j \cdot y_{s,d}, \quad (4)$$

and are presented as average $\pm$ 95% confidence intervals (CIs). The two sensitivity grids are: (i) $\text{AUC} \in [0.55, 0.83]$ and (ii) overbooking intensity $\kappa \in [0, 1.2]$ with the load being multiplied by the consultation time.

The main des parameters (modelling assumptions) are presented below:

Parameter	Value	Basis
Session, physicians, clerks	480 min, 3, 2	typical clinic [19]
Consultation (routine / urgent)	LN(13.5, CV 0.45) / 20 min	assumption
Registration (desk / digital)	LN(5, CV 0.4) / 1 min	assumption
Baseline no-show rate	15.9%	range in [4]
No-show model AUC	0.78 (0.55–0.83)	range in [3]
Reminder effect	–30% (top 30%)	[21]
Overbooking intensity κ	0.6 (0–1.2)	policy lever
Triage sens./FPR (conv. / AI)	0.7/0.04 vs 0.9/0.06	[5]
DC adoption (high / low DL)	85% / 35%	assumption
Walk-ins	1.25/h, 25% urgent	assumption

LN(m,CV) = Log Normal of mean m and coefficient of variation CV. Values are examples and should be substituted for local values.

3.2 Equity Scenario

For populations that do not have a large proportion in the dataset, the AI model might yield less accuracy. To do this, the noise in the no-show model is doubled and this results in a reduction of the within group AUC from 0.78 to 0.70 (compared to an AUC of 0.79 for high-literacy). Comparison is among three setups: conventional operations, full AI without fairness control and full AI with mitigation: The top 30% of the groups are assigned reminder quota; the check-in assistance from staff brings the low-literacy adoption rate near to 70%. The Satisfaction Response Model and Probabilistic Sensitivity Analysis are introduced here. To translate the process outcomes into perceived quality, whilst avoiding overstating precision, a probability functions are used to model the probability of a top-box satisfaction rating given each served patient.

$$\Pr(Y_i = 1) = \sigma(\alpha + \beta_w m_i \tilde{W}_i + \beta_{dc} D_i), \quad (5)$$

where W_i is the wait in 10-minute units, $\tilde{W}_i = \min(W_i, 120)/10$, $m_i = 1$ if patient is urgent (assuming more sensitive to delay) and $m = 1$ otherwise and D_i is the Digital check-in. The slope provides a summary of the negative relationship

between waiting and satisfaction that is well-known from the literature and β_{dc} represents the net experiential impact of digital contact, that is, convenience minus depersonalisation. As for a specific hospital, none of these parameters is known, a probabilistic sensitivity analysis (PSA) draws 4,000 sets of parameters vectors uniformly and independently from the intervals: $\pi_0 \in [0.60, 0.80]$ (conventional top-box rate, to which α is calibrated by a root searching method), $\beta_w \in [-0.40, -0.10]$ and $\beta_{dc} \in [-0.10, +0.25]$, and $\kappa_u \in$. Results are presented in the form of the distribution of the top-box satisfaction change from conventional operations, partial rank correlation coefficients (PRCC) and one-way tornado bounds.

3.4 Survey Instrument

The instrument which consisted of 34 items is shown in Table 3. The items for this study were created by the authors and modified (not copy and pasted) the constructs of SERVQUAL, eHEALS and behavioural-intention scales. All items are rated on a five-point agreements scale. The instrument is designed to take place post an outpatient visit with the instrument's administration time stamping alongside it, to allow perceptions to be matched with the DES outcome measures.

Table 3: Measurement Instrument (Original Wording; 5-Point Agreement Scale)

Construct	Code	Item
Digital transformation capability (DTC)	DTC1–4	The hospital's app, portal or kiosk made my visit easier to organise. / The hospital's systems already had my information, so I did not have to repeat it. / Automated reminders and scheduling were accurate and useful. / Technology helped staff decide quickly who needed care first.
Digital literacy (DL)	DL1–3	I am confident using smartphone apps to manage appointments. / I can find and understand health information on websites or portals. / I can complete online forms without help.
Reliability (REL)	REL1–4	My appointment started close to the scheduled time. / The hospital did things right the first time. / My records and test results were accurate and available when needed. / Staff delivered the services they promised.
Responsiveness (RESP)	RESP1–4	I did not wait long to be seen. / Staff responded promptly to my questions. / Registration and check-in were quick. / I received timely updates about delays or next steps.
Assurance (ASR)	ASR1–4	I trusted the competence of the clinicians. / I felt my personal and health data were handled securely. / Staff explained my diagnosis and treatment clearly. / I felt safe receiving care here.

Empathy (EMP)	EMP1–4	Staff gave me individual attention. / The clinician listened without rushing me. / Staff understood my specific needs. / I was treated with dignity, including when using digital services.
Digital tangibles (TAN)	TAN1–4	The app, portal and kiosks were easy to use. / Signage, screens and displays were clear. / Facilities and equipment looked modern. / Digital messages (SMS, e-mail, portal) were clear and professional.
Patient satisfaction (PS)	PS1–4	Overall, I am satisfied with this visit. / The care I received met my expectations. / This visit was a good experience. / I am pleased with how the hospital managed my visit.
Loyalty (LOY)	LOY1–3	I would recommend this hospital to family and friends. / I would choose this hospital again. / I would say positive things about this hospital.

Constructed Definitions are adapted from [10] [12] [29] [27]; wording of the items is original to this study.

3.5 PLS-SEM and PLSc estimation

This model is estimated by an original numpy implementation of Lohmöller's PLS algorithm with the path-weighting inner scheme and outer weights as mode A. At the end of each iteration the outer weight of each block j is updated as

$$w_j \propto \text{Cov}(x_j, z_j), \quad z_j = \sum_k e_{jk} Y_k, \quad (6)$$

where, e_{jk} is the regression coefficient (for predecessors) or correlation (for successors) between latent-variable proxies Y_j and Y_k . This means the estimates of the paths are deflated due to the measurement errors in the PLS composites. This is corrected in the consistent PLS (PLSc) with the reliability

$$\rho_A = (\hat{w}'\hat{w})^2 \cdot \hat{w}'(S - \text{diag } S)\hat{w} / \hat{w}'(\hat{w}\hat{w}' - \text{diag } \hat{w}\hat{w}')\hat{w}, \quad (7)$$

$$\tilde{r}_{jk} = r_{jk} / \sqrt{(\rho_{A,j} \rho_{A,k})}, \quad (8)$$

This correlation matrix is corrected and then, the structural model is re-estimated using ordinary least squares. It is an algorithm that uses the correlation matrix of the indicators, thus requiring only one matrix computation per $p \times p$ bootstrap re-sample. Some commonly used measures of measurement quality are loadings, Cronbach's α , ρ_A , composite reliability ρ_C , average variance extracted (AVE) and the heterotrait–monotrait ratio,

$$\text{HTMT}_{jk} = \tilde{r}_{jk} / \sqrt{(\tilde{r}_{jj} \tilde{r}_{kk})}, \quad (9)$$

where $\tilde{r}_{jk}$ are mean heterotrait correlation and $\tilde{r}_{jj}$ are mean monotrait correlation. Current reporting guidelines are followed to assess the structural model, which is done by using R^2 , inner VIF, effect size $f^2 = (R^2_{\text{incl}} - R^2_{\text{excl}})/(1 - R^2_{\text{incl}})$, and percentile-bootstrap CIs (2,000 resamples) and PLSpredict Q^2_{predict} with a linear-model benchmark. Moderation effect (H4) is performed based on two-stage approach. Common-method bias is penalized using the full-collinearity VIF (if $\text{VIF} > 3.3$, then it is flagged) and Harman's single-factor test.

3.6 Monte Carlo Design

The population model (Table 1) is used for the synthetic respondents. The standardised loadings of the 34 indicators range from 0.70 to 0.87. Following established practice for SEM sample-size planning, 300 replications are run at each $n \in \{100, 150, 200, 300, 400, 600, 800, 1000\}$. For each replication, both PLS and PLSc and sum-score regression is compared and each effect is tested with a 95% percentile-bootstrap interval with 100 resamples. The findings of this study include relative bias, root-mean-square error (RMSE) of the bias, coverage of CI, power, and Type-I error of the true null path between DTC and PS. The Monte Carlo sample sizes (80% power) are compared with the inverse-square-root (ISR) rule,

$$n_{\text{min}} > (2.486 / |\beta_{\text{min}}|)^2, \quad (10)$$

In both 5-point Likert discretisation cases, both with symmetric thresholds and with a ceiling-skewed distribution (thresholds $-2.0, -1.4, -0.8, 0.1$ will result in $\sim 79\%$ of the responses being collapsed in the top two categories, which is common practice in satisfaction surveys), robustness (RQ4) is explored for $n = 300$ and 800 . In addition, it is tested under a popular method factor loading $\lambda_m = 0.2/0.3/0.4$ for each indicator.

4. Results

Based on the three learning phases of the AA model, they address the operational effects of AI interventions (RQ1) in the following ways:

The eight scenarios are summarised in Table 4 and in Fig.3. The number of patients processed per clinic, and conventional mean waiting time in the clinic (6.98 min, 90th percentile > 16.8 min) and overtime (9.7 min) were 89.9 patients per session. NS ran with 12.6% No shows (vs 15.6%) allowing for 8.3 overbooked patients per day to achieve a throughput of 99.9 patients (10.0 patients/session). This meant that the mean wait time increased to 13.81 min and the overtime wait time to 21.4 min, and the three were linked to a difference in orientation: DC had the opposite profile. It did not involve any reduction in the number of patients served, but it was able to reduce the waiting time to 3.80 min and the time in clinic from 26.5 to 21.2 min. Therefore, the full AI configuration provided NS's throughput of 99.9 while providing a mean wait of 8.10 min which is 1.12 min less than conventional, and they also reduced the mean wait time of urgent-patients from 5.71 min to 3.44 min.

Table 4: Scenario Outcomes (Mean Of 500 Simulated Days) And Psa Change In Top-Box Satisfaction

Scenario	Mean wait (min)	P90 wait (min)	Urgent wait (min)	Served / day	Overtime (min)	Utilisation	Time in clinic (min)	Δ top-box [95% interval]	(pp) PSA	P($\Delta > 0$)
Conventional	6.98	16.8	5.71	89.9	9.7	0.852	26.5	reference	—	—
NS	13.81	28.1	7.82	99.9	21.4	0.924	34.5	-3.96 [-6.3, -1.6]	0.00	0.00
TR	7.03	17.1	5.19	89.9	9.7	0.852	26.5	-0.02 [-0.0, -0.0]	0.00	0.00
DC	3.80	11.2	3.10	89.9	6.0	0.827	21.2	+2.73 [+0.1, +5.6]	0.98	0.98
NS+TR	13.94	28.9	6.33	99.9	21.4	0.924	34.5	-3.98 [-6.3, -1.6]	0.00	0.00
NS+DC	7.99	19.1	4.37	99.9	12.5	0.907	26.9	+0.34 [-2.0, +2.8]	0.57	0.57
TR+DC	3.85	11.4	2.76	89.9	6.0	0.827	21.3	+2.72 [+0.0, +5.6]	0.98	0.98
Full AI	8.10	19.7	3.44	99.9	12.4	0.907	26.9	+0.31 [-2.1, +2.8]	0.56	0.56

NS = AI no-show prediction with targeted reminders and overbooking ($\kappa = 0.6$); TR = AI-assisted triage; DC = digital check-in and intake. Half-widths of the 95% CIs over days

are at most 0.86 min for the mean wait and 0.41 patients for throughput. Highlighted: the best balance of throughput and waiting (see Fig. 7a for the overbooking frontier).

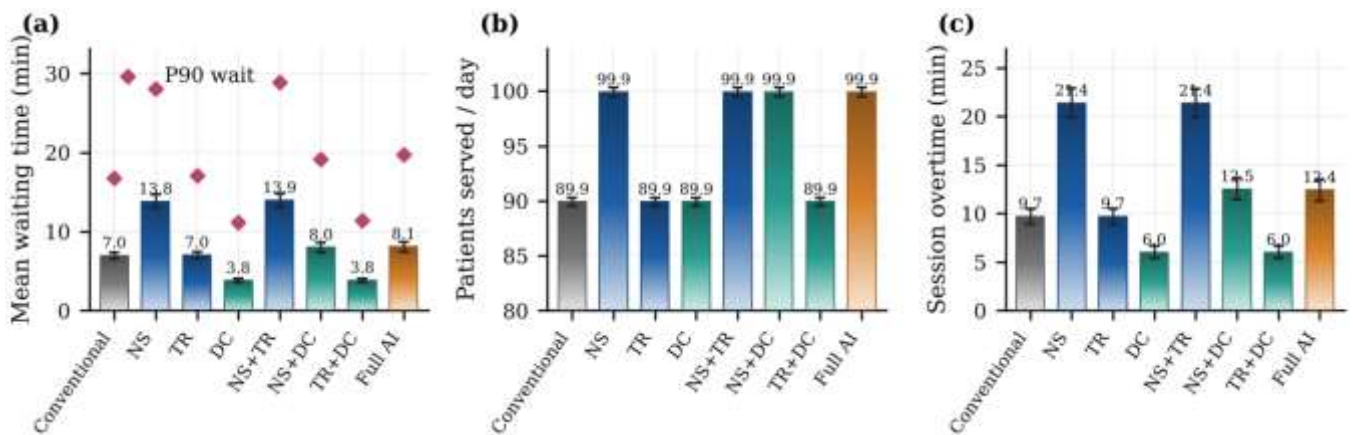


Fig. 3. Operational outcomes by scenario: (a) mean waiting time (bars, 95% CI) and 90th percentile wait (diamonds); (b) patients served/ session; (c) overtime per session.

The factorial effects (Fig. 4) separate the contributions. For mean waiting time the NS main effect was +5.55 min (95% CI +5.06 to +6.04) and the DC effect -4.50 min. The NS×DC interaction is negative: digital check-in reduces some of the overbooking congestion, as it takes away the congestion in the registration process at exactly the time of congestion in the arrivals in the system. There was no

significant impact of TR on the average wait (+0.08 min), but urgent-patient waits were reduced (-0.82 min) and where the queue is longer, prioritisation is more important and coloured positively by NS (-0.39 min). There were negligible three-way interactions (all $|\text{effect}| \leq 0.10$ min). NS added +9.1 min and DC removed 6.3 min for overtime.

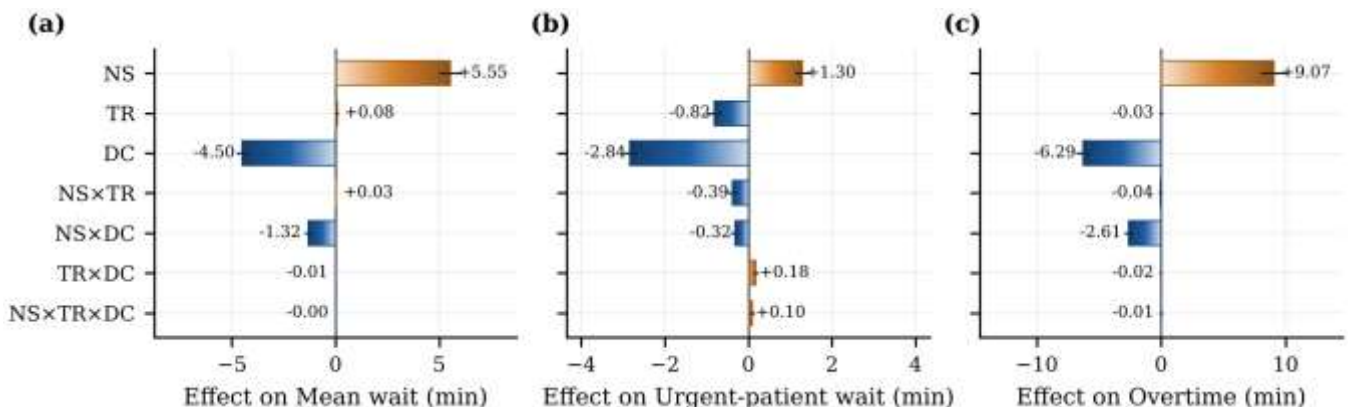


Fig. 4. Main effects and interactions for the factorial design (± 1 coding, mean over 500 paired days with 95% CI). Orange: up, blue: down.

Sensitivity to accuracy and demand of the model. The more

discrimination made for the no-shows (higher AUC), the

higher throughput (monotonically increasing, from +7.9 patients per session for AUC 0.55 to +10.9 patients per session for AUC 0.83), and the lower the residual no-show rate (from 13.9% to 12.1%). The fixed overbooking, on the other hand, results in a higher number of additional wait times at higher loads: at nominal load, the additional wait increases from 4.7 to 7.3 min. The same policy results in an

additional 14.1-20.5 minutes of waiting time and 30-43 minutes of overtime at 1.15, versus approximately 1.1 and 1.2 minutes at 0.85. This results in the gains of a more accurate model being translated into volume, unless the overbook rule is adjusted to capacity.

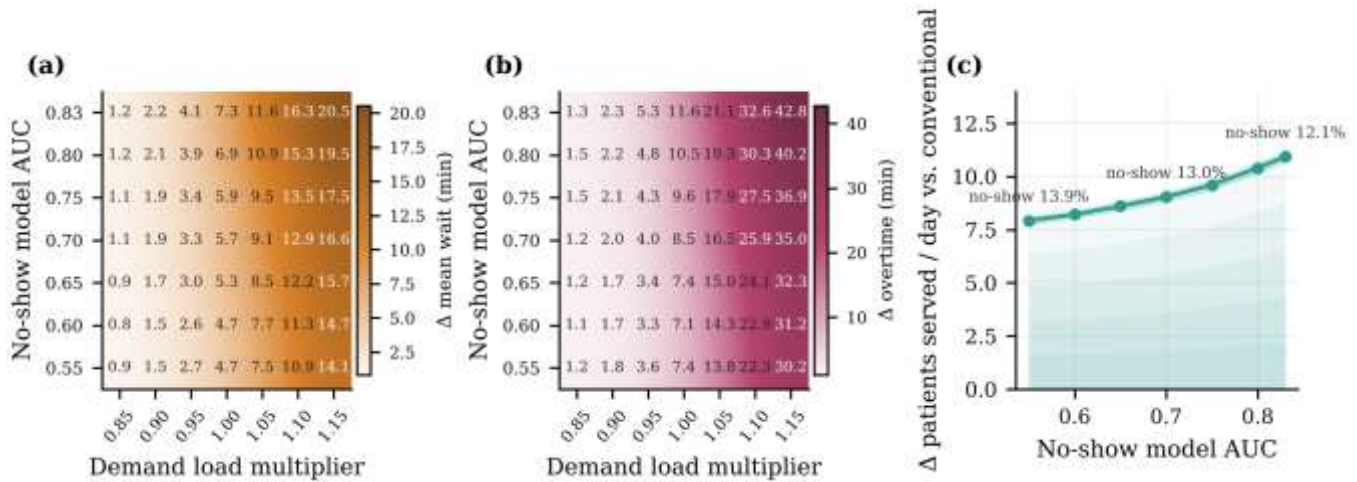


Fig. 5. NS policy vs. conventional across no-show model AUC and demand load: (a) change in mean wait; (b) change in overtime; (c) additional patients per session served (and 95% CI band) with residual no-show rate.

Distributions and "urgent access". Averages smooth out the tails (Fig. 6). NS extends the upper tail of the waiting distribution (with 90th percentile 28.1 min vs 16.8), and DC shortens it (11.2 min). The entire distribution is shifted to the left, as seen in Fig. 6b, for urgent patients with the full

AI configuration. The triage benefit is highly load dependent as can be seen in Fig. 7b. At nominal load AI triage saves urgent patients 0.58 min at a cost of 0.12 min to routine patients. At a load multiplier of 1.15 it saves 2.52 min (95% CI 1.99–3.05) at a cost of 0.42 min.

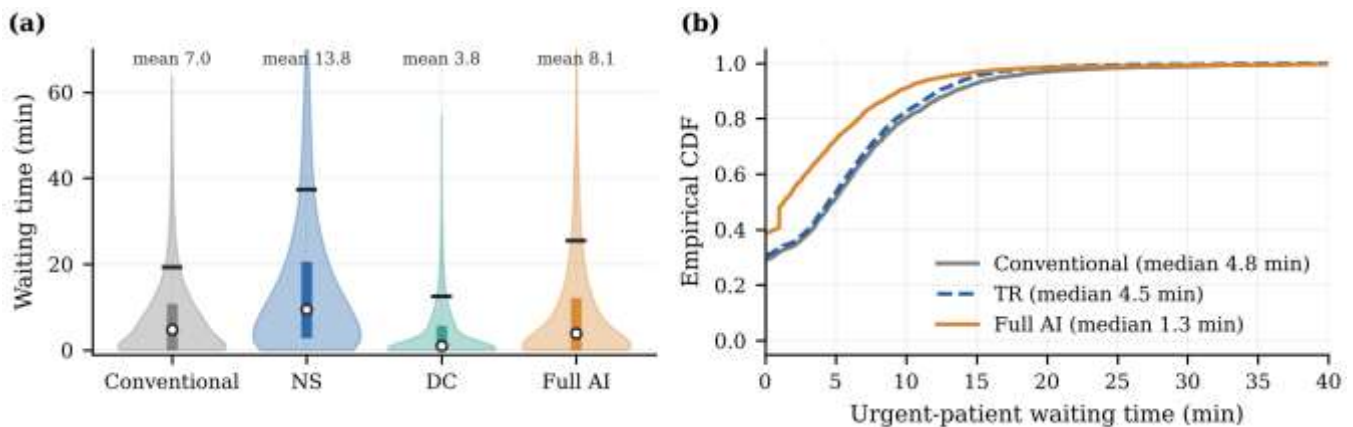


Fig. 6. Patient-level waiting-time distributions (violin; box = interquartile range; white dot = median; bar = 90th percentile). The empirical CDF of the waiting time for an urgent patient is shown in (b).

Overbooking frontier. The throughput as a function of mean wait, as κ changes, is shown in Fig. 7a. NS alone is not as intuitive as the conventional wait, even at $\kappa = 0$ (reminders only: 8.15 min, 92.7 patients). The mean wait is kept below the conventional 6.98 min, while with DC and TR, every κ up to 0.4 insures service of 97.5 patients (+7.6) with a mean

wait of 6.71 min and a mean overtime of 9.8 min. The design implication here is that no-show AI and digital check-in should be used in combination, and the amount of overbooking should not be predetermined but instead be decided on this frontier.

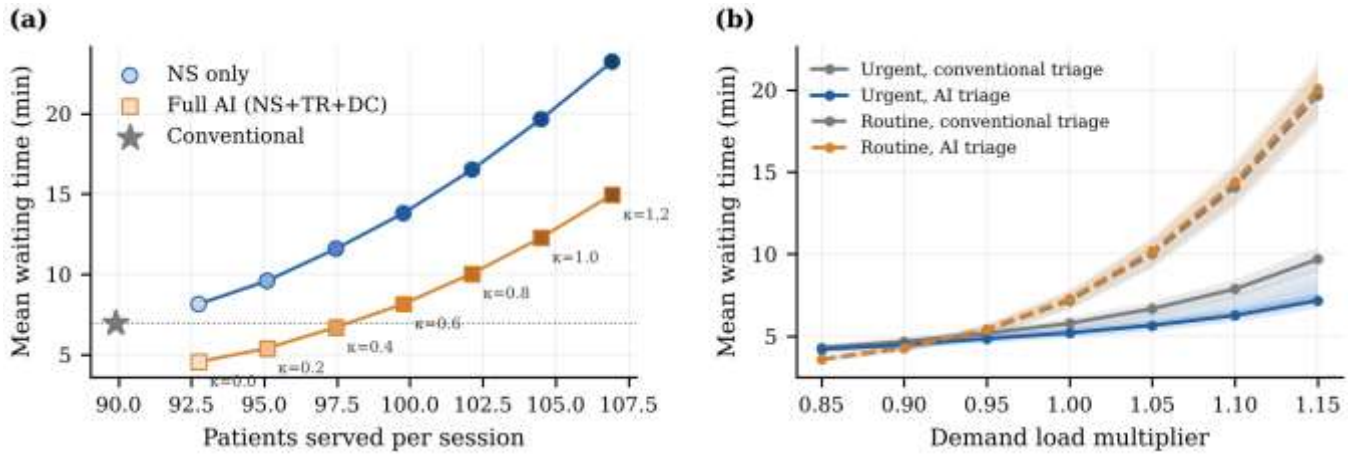


Fig. 7. Throughput–waiting frontier, varying the overbooking intensity κ , when only NS is used and when the full AI configuration (conventional = star) is used. (b) Mean wait times of urgent and routine patients for conventional and AI triage at increasing demand loads (shaded: 95% CI).

4.2 Satisfaction Consequences and Equity (RQ2)

These process changes are reflected in top-box satisfaction in the PSA (Fig. 8a, Table IV). NS alone impacted satisfaction in each parameter draw (mean decrease of -4.0 pp) as its throughput improvements cannot be seen by the surveyed patients, but its waits can be. DC raised satisfaction with 98% of draws (mean $+2.7$ pp). The nominal load of TR alone resulted in no measurable change

(-0.02 pp). The full AI configuration was centred near zero ($+0.31$ pp; $P(\Delta > 0) = 0.56$). That is, the satisfaction was about the same for the operational benefit ($+10.0$ patients per session). The net effect of digital contact, β_{dc} (PRCC = 0.99; tornado range -2.1 to $+2.6$ pp, Fig. 8b), was the most important measure to record during a real deployment, followed by wait sensitivity (PRCC = 0.83).

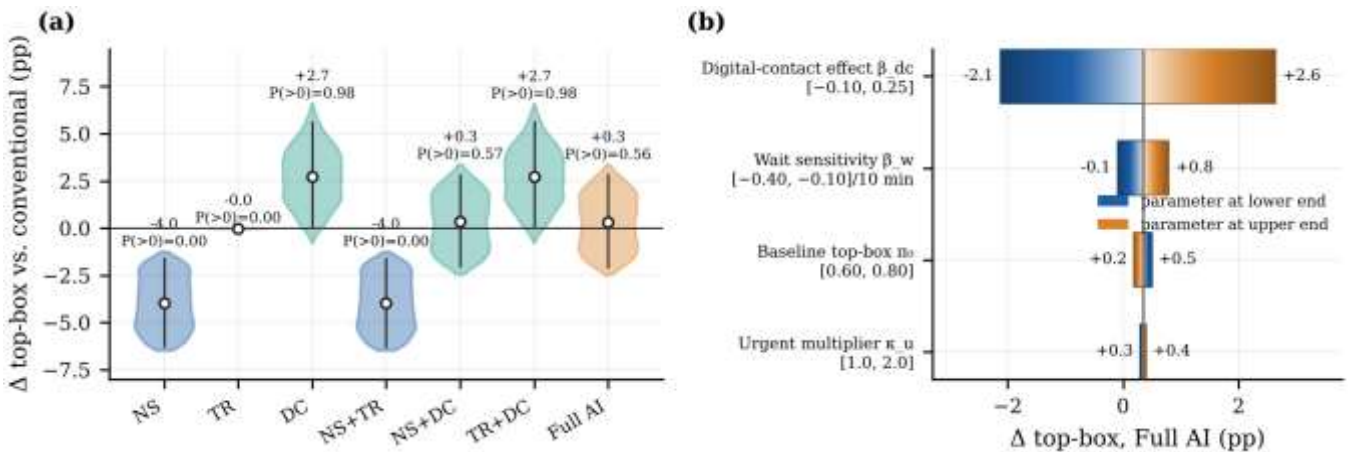


Fig. 8. 4,000 draws of probabilistic sensitivity analysis. (b) Distribution of difference in satisfaction at the top box compared to conventional operations (dot = mean, line = 95% interval). (b) One-way tornado for the complete AI configuration for the central parameter values.

Equity. The low-literacy patients in the full AI configuration waited 8.38 min, whereas the high-literacy patients waited 7.73 min (without fairness controls, 6.93 for low-literacy and 7.00 for high-literacy patients). Residual no showing rates were also higher for low-literacy patients (15.2% vs 11.5%), and much lower for low-literacy patients in regard to digital check-in uptake (35% vs 85%). Patients with high values of PSA in the centre experienced a dissatisfaction

that was lower than the usual (69.7%). and those with high literacy level experienced a satisfaction that was higher than usual (70.9%). The mitigated configuration decreased the waiting gap from 0.35 to 0.08 min and the satisfaction gap from 1.2 to 0.3 pp, both groups were above the conventional level and the throughput remained at 99.5 vs 99.6 patients. This residual no-show gap continued (15.8% vs 11.2%) because it is not due to the algorithm itself.

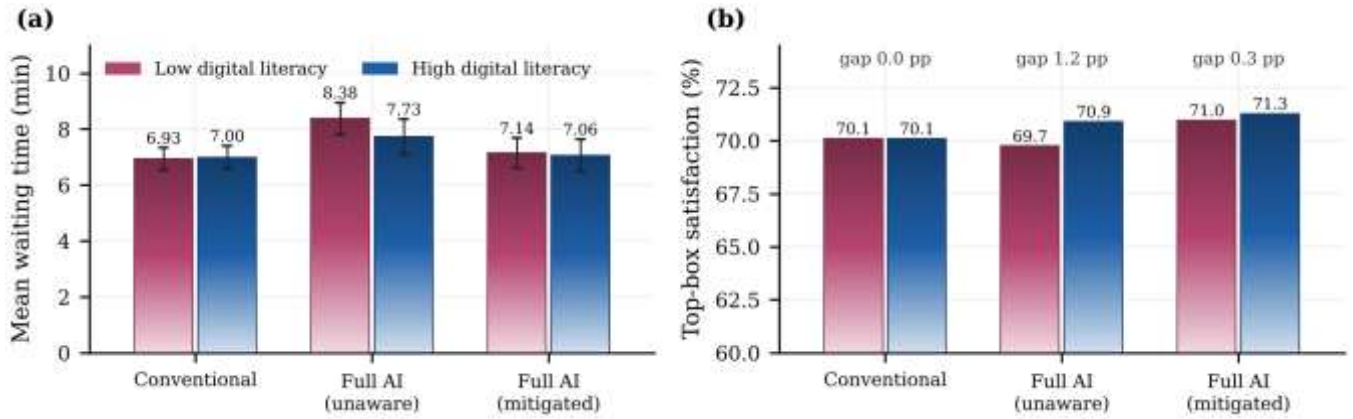


Fig. 9. Equity by digital literacy when the no-show model is less accurate for low-literacy patients: (a) mean waiting time (95% CI); (b) top-box satisfaction at central PSA parameters.

4.3 Measurement and Structural Assessment Protocol (Synthetic Demonstration)

To show the complete analysis protocol and to verify the software against known truth, one synthetic sample of $n = 400$ respondents was drawn from the population model and discretised to the five-point scale. These results describe simulated respondents, not patients. The measurement model met all common thresholds (Table 5): loadings ranged from 0.732 to 0.898, Cronbach's α from 0.776 to

0.865, ρ_C from 0.866 to 0.908, and every AVE was ≥ 0.632 . Discriminant validity held: the largest HTMT was 0.60, below the 0.85 threshold (Fig. 10b), and every $\sqrt{\text{AVE}}$ exceeded the construct's correlations with the other constructs (Fornell-Larcker). The SRMR of the composite-factor model was 0.034. Inner VIFs did not exceed 1.62, and the largest full-collinearity VIF was 1.80 (< 3.3), so the screen gave no indication of common-method bias.

Table 5: Measurement Model (Synthetic Sample, $N = 400$)

Construct	Loadings	α	ρ_A	ρ_C	AVE	Max HTMT
DTC	0.79–0.87	0.855	0.857	0.902	0.697	0.47
DL	0.76–0.90	0.776	0.860	0.866	0.685	0.33
REL	0.78–0.85	0.823	0.830	0.883	0.653	0.52
RESP	0.74–0.87	0.842	0.855	0.894	0.679	0.47
ASR	0.77–0.84	0.813	0.817	0.877	0.641	0.38
EMP	0.77–0.86	0.820	0.826	0.881	0.650	0.41
TAN	0.73–0.83	0.805	0.812	0.873	0.632	0.47
PS	0.80–0.87	0.865	0.867	0.908	0.712	0.60
LOY	0.81–0.89	0.808	0.821	0.886	0.722	0.60

Thresholds: loadings ≥ 0.708 , α and ρ_C 0.70–0.95, AVE ≥ 0.50 , HTMT < 0.85 [17], [39].

The structural model is given in Table 6 and Fig. 10a. There were also good comparisons between PLSc estimates and population values, with the CI for 95% of paths covering the actual value. PLS, on the other hand, underestimated the stronger paths, such as PS $\rightarrow$ LOY (0.509 vs true 0.60, PLSc 0.604). In this sample the following paths were not significant: H1d, H2c, H2e, H5. H5 is a real null and was retained as such, but the others are real small effects ($\beta =$

0.10 - 0.15) that a sample of 400 could not detect. The total indirect effect of DTC on PS was 0.263 (95% CI 0.171–0.367; true 0.270). The specific indirect effects through reliability (0.116) and responsiveness (0.076) were significant. The moderation effect H4 was estimated at 0.076 (95% CI -0.013 to 0.161; true 0.125), so the test missed a real effect. Simple slopes of DTC $\rightarrow$ RESP were between 0.27 at -1 SD of literacy and 0.41 at $+1$ SD of

literacy. R^2 was 0.377 for PS and 0.364 for LOY. All PS indicators (0.047–0.070) were positive and LOY was near zero in Q^2_{predict} . The PLS prediction errors were less than

the linear benchmark on 7 of 7 indicators. The following systematic power analysis is inspired by these misses at $n = 400$.

TABLE 6: Structural model (synthetic sample): population vs estimates

Hyp.	Path	True	PLS	PLSc	PLSc 95% CI	f^2	Sig.
H1a	DTC→REL	0.35	0.331	0.392	[0.29, 0.49]	0.182	yes
H1b	DTC→RESP	0.40	0.343	0.395	[0.28, 0.50]	0.195	yes
H1c	DTC→ASR	0.25	0.177	0.211	[0.10, 0.33]	0.047	yes
H1d	DTC→EMP	0.10	0.092	0.109	[- 0.01, 0.23]	0.012	no
H1e	DTC→TAN	0.45	0.394	0.473	[0.37, 0.57]	0.288	yes
ctrl	DL→RESP	0.15	0.206	0.227	[0.13, 0.33]	0.065	yes
H2a	REL→PS	0.25	0.252	0.295	[0.16, 0.43]	0.092	yes
H2b	RESP→PS	0.20	0.182	0.193	[0.06, 0.32]	0.037	yes
H2c	ASR→PS	0.15	0.097	0.095	[- 0.02, 0.21]	0.011	no
H2d	EMP→PS	0.20	0.137	0.141	[0.03, 0.26]	0.023	yes
H2e	TAN→PS	0.10	0.082	0.076	[- 0.07, 0.22]	0.006	no
H5	DTC→PS	0.00	0.060	0.045	[- 0.09, 0.17]	0.002	no
H3	PS→LOY	0.60	0.509	0.604	[0.52, 0.68]	0.573	yes
H4	DL×DTC→RESP	0.125	0.076	—	[- 0.01, 0.16]	—	no
H6	DTC→SQ→PS (total)	0.270	—	0.263	[0.17, 0.37]	—	yes

Percentile bootstrap with 2000 resamples; f^2 computed from PLSc. H4 uses the two-stage approach with PLS scores

(1,000 resamples). Rows marked "no" with a non-zero true value are Type-II errors at $n = 400$.

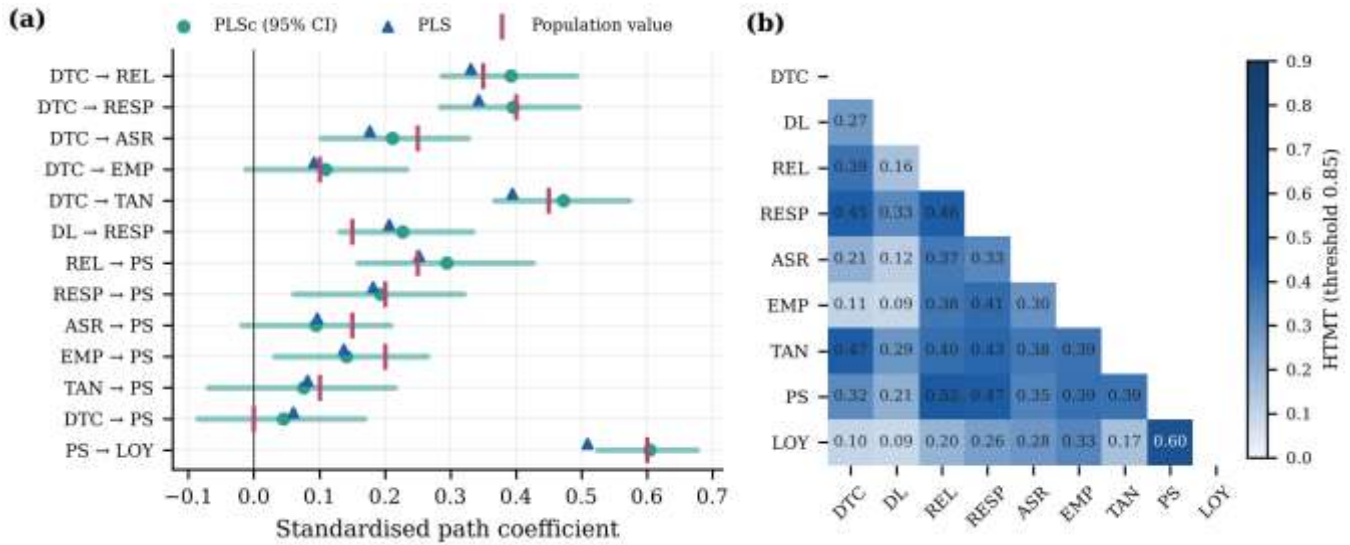


Fig. 10. Synthetic demonstration ($n = 400$, five point items). (a) PLS estimates with 95% bootstrap CIs, PLS estimates, and population values. (b) HTMT matrix with all values below the 0.85 discriminant-validity criterion.

4.4 Work Package Duration Estimation (Man-Days)

Fig. 11 shows the comparison of the three estimators over the 300 replications per sample size. PLS and sum scores both showed a nearly constant relative bias of 8.8% and 9.9% (median relative bias at $n = 400$), respectively, which will not decrease as n increases. The median relative bias of PLS was near unbiased (+3.5% at $n = 100$) and within $\pm 2.6\%$ from $n = 150$, and was less than 1% for larger numbers of data points, $n \geq 300$. The PLS bias does not change as the sampling error reduces, so that PLS confidence intervals are more likely to be incorrect as sampling error decreases. For $n = 1000$, percentile-bootstrap coverage for PLS was 71.8% while PLS was between 92.5% and 93.6%, slightly below the nominal 90%, due to

somewhat wide intervals for 100 resamples. This would result in a Type-I error of the true null direct path $DTC \rightarrow PS$ for PLS (8.0–11.3%) and sum scores (8.3–12.0%), and lead to a false rejection of a full-mediation hypothesis. McLeod et al. (2011) found that at $n = 100$, 1.3% of replications had at least one inadmissible path estimate ($|\beta| > 1$), which is similar to PLS; for $n \geq 200$, the consistency of PLS is 0%. The paths into responsiveness were primarily impacted by the extreme resamples in the pilot run using the standard-error-based tests, with the three-item literacy scale being a part of the predictor block for the path to responsiveness. This is rare and leads to overestimation of the RMSE of PLS for the smallest sizes (Fig. 11b).

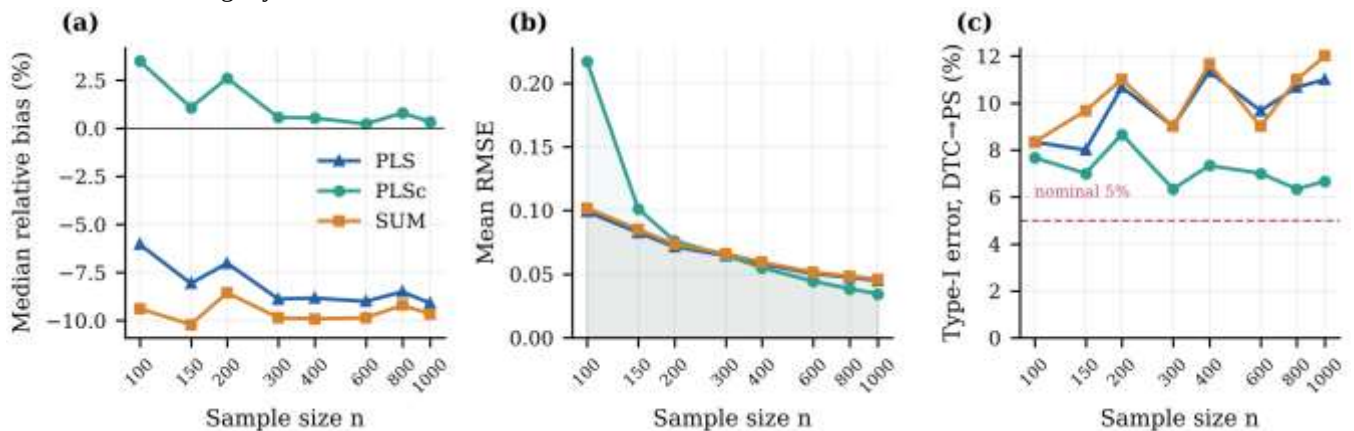


Fig. 11. Median relative bias averaged across the 12 non-null paths, RMSE, Type-I error for the true-null path $DTC \rightarrow PS$ (percentile-bootstrap test) of PLS, PLS and sum-score regression (300 replications per n).

4.5 Power and Required Sample Size (RQ3)

Fig. 12 and Table VII report power. The power to detect the larger paths ($\beta \geq 0.35$, H1a, H1e, H3) was approximately 80–150 respondents under any estimator. The small effects that had the most substantial interest were much harder! $DTC \rightarrow EMP$ ($\beta = 0.10$) required ≈ 950 (PLS) and ≈ 950 (PLS) respondents. $TAN \rightarrow PS$ ($\beta = 0.10$, one of five correlated predictors) required ≈ 680 and > 1000 . With the

two-stage approach, the literacy moderation H4 (standardised $\beta = 0.125$) needed ≈ 600 samples. The PLS (approximated at ≈ 200) and the PLS (approximated at ≈ 320) were required for the specific indirect effect via responsiveness. The reason that PLS appears to be stronger than PLS is that it has narrower intervals around a biased estimate, yet this is part of the reason for the decaying coverage.

Table 6: Sample Size For 80% Power By Hypothesis And Estimator (Monte Carlo) Vs. Inverse-Square-Root Rule

Hyp.	Path / effect	β_{pop}	PLS	PLSc	Sum score	ISR rule [48]	Power at n = 300 (PLS / PLSc)
H1a	DTC→REL	0.350	≤ 100	≤ 100	≤ 100	51	100% / 100%
H1b	DTC→RESP	0.400	≤ 100	≈ 130	≤ 100	39	100% / 99%
H1c	DTC→ASR	0.250	≈ 150	≈ 150	≈ 160	99	96% / 96%
H1d	DTC→EMP	0.100	≈ 950	≈ 950	> 1000	619	38% / 38%
H1e	DTC→TAN	0.450	≤ 100	≤ 100	≤ 100	31	100% / 100%
H2a	REL→PS	0.250	≈ 140	≈ 180	≈ 140	99	99% / 95%
H2b	RESP→PS	0.200	≈ 200	≈ 300	≈ 210	155	91% / 80%
H2c	ASR→PS	0.150	≈ 280	≈ 420	≈ 300	275	83% / 69%
H2d	EMP→PS	0.200	≈ 170	≈ 210	≈ 180	155	94% / 85%
H2e	TAN→PS	0.100	≈ 680	> 1000	≈ 700	619	51% / 31%
H3	PS→LOY	0.600	≤ 100	≤ 100	≤ 100	18	100% / 100%
H4	DTC×DL→RESP	0.125	≈ 600	—	≈ 600	394	46% / —
H6a	ind. DTC→RESP→PS	0.080	≈ 200	≈ 320	≈ 210	966	91% / 78%
H6	ind. total	0.270	≤ 100	≈ 140	≤ 100	85	100% / 99%

Interpolated on a log scale between simulated sizes (100, 150, 200, 300, 400, 600, 800, 1000); percentile-bootstrap test (100 resamples), $\alpha = 0.05$. "> 1000" = not reached

within the design. No consistent two-stage interaction estimator is available for PLSc, so H4 is reported for PLS and sum scores only.

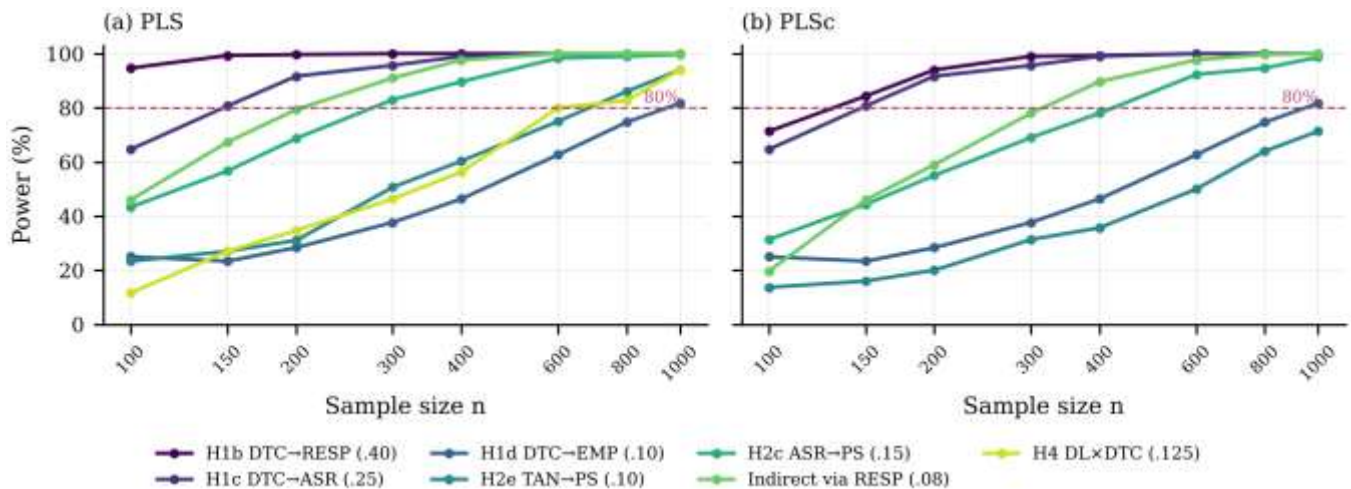


Fig. 12. Power curves of selected hypotheses for (a) PLS and (b) PLSc (300 replications per n; dashed line = 80%).

Fig. 13 summarizes the power of all the effects according to their size in terms of population. The ISR rule (triangles) was reasonable to use for PLS, with a $\beta = 0.15$ level, where it took about 280 respondents to reach 80% power with an ISR value of 275. But it was too idealistic for the shortest distances. The DTC → EMP ($\beta = 0.10$) needed ≈ 950 respondents with PLSc, while one of five correlated predictors (TAN → PS) needed > 1000. The rule was conservative about indirect effects (the actual indirect effect was $\beta = 0.08$, whereas the rule requires that ≈ 320

respondents be observed instead of 966). The effect size is provided by the DES in an empirical way. If there is a strong correlation between perceived waiting and actual waiting then a responsiveness path of around $\beta \approx 0.20$ is plausible. The patient-level point-biserial correlation between digital check-in and waiting time was $r = 0.20$. With PLSc, effects of this magnitude (RESP → PS, EMP → PS) were found to have 80% power at ≈ 300 and ≈ 210 respondents, respectively.

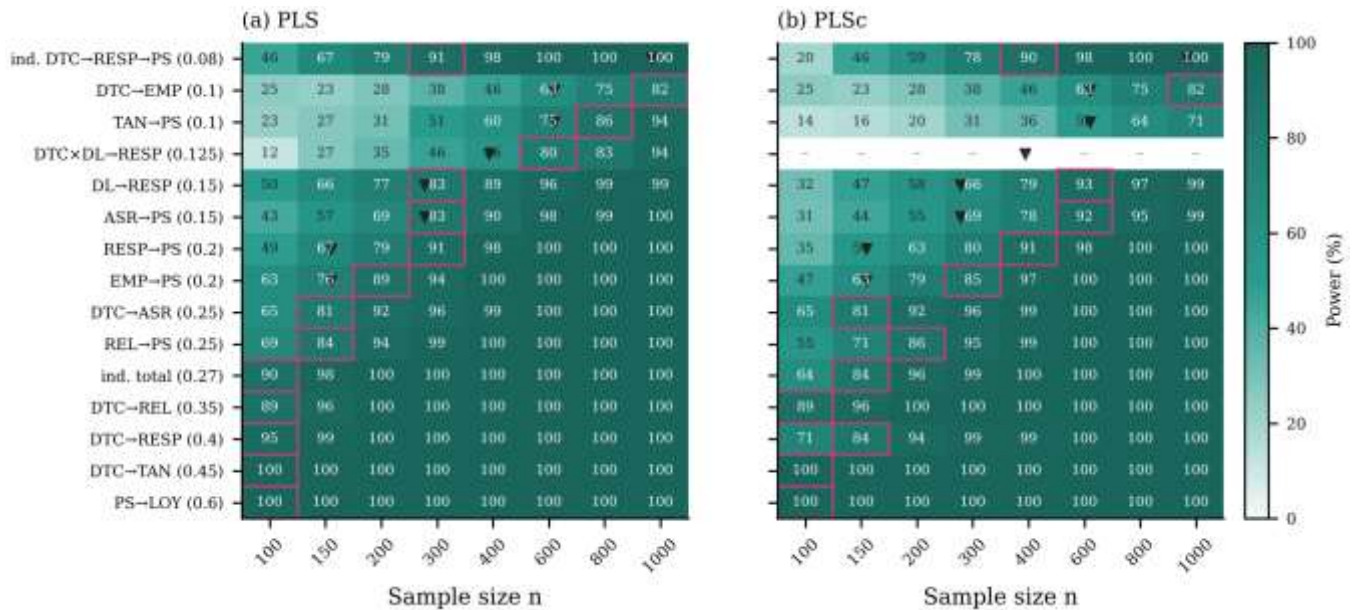


Fig. 13. Sample size dependent power (%) of each of the hypothesised effects (shown in parentheses in the rows, sorted by population size, (a) PLS, (b) PLSc). Magenta box: first n such that 80% of the values are in the box. triangle: sample size according to the inverse-square-root rule (displayed if the number of samples is in the simulated range). “—”: no PLSc estimator for the interaction.

4.6 Robustness to Measurement Conditions (RQ4)

Table 8 and Fig. 14 present the results of the robustness checks for PLSc. The median relative bias for 5-point discretisation with symmetric thresholds (+1.4%) remained virtually the same as that for continuous data (+0.1%) for $n = 300$. The distribution of satisfaction was skewed towards the ceiling, which reduced the paths by 4.9% ($n = 300$) and 5.6% ($n = 800$). A more insidious, though not more common, method variance was common, but one that happens to be not typically looked for. A method factor that loads on every item inflated the true structural effects by +11%, +21% and +35% for $\lambda_m = 0.2, 0.3$ and 0.4 ($n = 800$). This bias cannot be removed by PLSc, because there is no difference between the variance in the shared method and the true variance in the score. The Type-I error of the null direct path ranged between 3.5% and 10.0% in all

conditions, indicating that the majority of the method variance is shared by the mediators of the SERVQUAL model. The contamination was not detected either by the diagnostic or the other diagnostic. The mean maximum full-collinearity VIF was limited to 2.46 at $\lambda_m = 0.4$ which is lower than Kock's 3.3 threshold, meaning there were no samples flagged. The first factor accounted for 41% of the variance which is below 50% and indicated 0%. However, the net bias from a real survey is not predictable for two reasons: the first is a ceiling effect, the second is a method variance (which is partially offset by the former). The temporal disconnection of predictor and outcome measurement or the identification of objective time stamps that will be linked to responses, as suggested in this framework, are therefore crucial.

Table 8: Robustness of Plsc Inference (200 Replications)

Condition	Rel. bias $n=300$	Rel. bias $n=800$	Type-I $n=300$	Type-I $n=800$	Max VIF / Harman $n=800$
Continuous	+0.1%	+0.1%	3.5%	4.5%	1.84 / 26%
5-pt symmetric	+1.4%	+0.7%	7.0%	4.5%	1.77 / 24%
5-pt ceiling	-4.9%	-5.6%	10.0%	5.5%	1.67 / 22%
CMV $\lambda_m = 0.2$	+10.9%	+11.2%	9.5%	7.5%	1.99 / 30%
CMV $\lambda_m = 0.3$	+22.1%	+21.2%	9.0%	4.0%	2.18 / 34%
CMV $\lambda_m = 0.4$	+34.0%	+34.7%	6.0%	8.0%	2.46 / 41%
Ceiling + CMV 0.3	+15.3%	+13.7%	6.5%	3.5%	1.91 / 29%

Relative bias = median over replications, averaged over the 12 non-null paths. Type-I error is for the true-null path DTC → PS (nominal 5%). Max VIF = mean over samples of the

largest full-collinearity VIF (flag > 3.3); Harman = variance share of the first unrotated factor (flag > 50%). Neither screen flagged any sample in any condition.

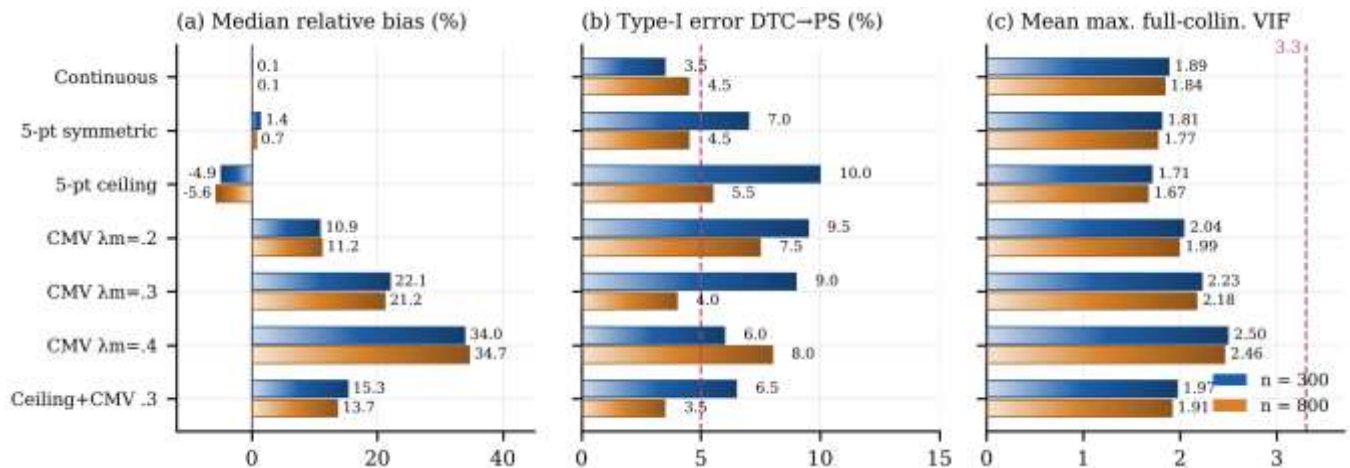


Fig. 14. 12 non-null paths (a) Median Relative Bias (MRB) at $n = 300$ and $n = 800$, (b) Type-I error of the null path at $n = 300$ and $n = 800$, (c) Mean of Maximum VIF (full collinearity) (dashed: Kock's 3.3 threshold) at $n = 300$ and $n = 800$.

5. Discussion

5.1 Theoretical Implications

The findings indicate that there was one major theoretical implication. AI-driven transformation and how it impacts patient satisfaction isn't something inherent to the technology but rather depends upon how it's set up. The same no-show model reduced the satisfaction, when applied to a maximisation of throughput. When used in conjunction with digital check-in (DCI), it was near neutral (Table 4) and at moderate levels of overbooking it maintained waits at the conventional level (Fig. 7a). The mean of such studies, which regress satisfaction on a global "digitalisation" score, therefore, tends to average out opposite-sign effects, reducing the estimated associations. The SERVQUAL lens helps to identify the places where the mechanisms are introduced. NS is about reliability (time of appointments) and responsiveness (waiting) and in opposing directions for the hospital and for the patient. DC responds and to the digital tangibles. The benefit of TR only seems to be experienced if queues are long enough and through TR for the most vulnerable patients comes assurance and responsiveness. The framework then connects the process quality, as defined by Donabedian, to the perceived quality in a testable manner.

The second point is from the equity analysis. The average satisfaction can be increased, while for the patients with low digital skills, it can be decreased, in the context of digital transformation. But this was despite the fact that there was no protected attribute in use in the algorithm; it was similar to mechanisms reported for clinical algorithms, which were found to be related to lower model accuracy and uptake of self-service. Digital literacy should thus be taught explicitly—like H4 is—and not be viewed as noise.

5.2 Managerial Implications

Implement AI Scheduling with front-end digitisation. Due to the lack of throughput, overbooking is used on its own to trade patient. Much of that trade-off is provided by digital check-in, and the frontier in Fig. 7a indicates the overbooking intensities that ensure that waiting is at or below its current level.

Just tune overbooking so it loads, not for the purpose of modeling accuracy. The more attendees that can be generated by a better no show model, the greater the resulting congestion will be—under a given fixed overbooking rule (Fig. 5). Update the value of κ whenever there are changes in demand or changes in model performance.

- Take advantage of AI triage in cases where queues are long. Between nominal and peak load its value for the urgent patients increased several times (Fig. 7b).
- To track results using digital literacy. Most of the satisfaction gap was eliminated at no throughput cost (Fig. 9) by both group-wise reminder quotas and staff-assisted check-in.
- The primary source of uncertainty in the PSA was the measure of the net effect of digital contact – this could be easily determined from a short pilot survey of comparing digital and desk check-in.
- Patient-satisfaction research has methodological implications, as described below. Patient-satisfaction research has implications for methodologies, as outlined below.

The following 4 recommendations follow on empirical studies on the digital transformation and patient satisfaction. The effect of PLS (or sum scores) is underestimated by ~ 10% and mediation hypotheses are over-rejected when the intent is to test theory. (ii) Make plans that will be affected by small effects. In this model, including rules of thumb (cited by others as being about 60 respondents for this

model), several hundred or in this case over 1,000 respondents would be required for the small but substantively important paths and the literacy moderation, respectively (Table 7). (iii) Be aware of 'ceiling' effects and think about using broader scales of response or top-box modelling. (iv) Control for common-method variance, by using the operational time stamps used by the DES to tie survey responses together. The moderate strength method variance raised the magnitude of each effect by about 10-35%, and did not fail any of the standard statistical screens.

5.4 Limitations and Validity

This is a simulation study, whose results are subject to the assumptions made. Construct validity: the validity of the DES parameters (Table II), the satisfaction response model and population path values are not the values from a particular hospital, but rather are reasonable design assumptions based on the literature. The PSA and the sensitivity grids deal with uncertainty in the parameters but not with uncertainty regarding the structure. Non-models include physician heterogeneity and day-to-day demand variation, multi-stage diagnostics, and patient balking. Internal validity: common random numbers, 500 replicated days per simulation scenario, and 300 Monte Carlo replications per condition ensure a small simulation error and provide CIs throughout. External validity: The synthetic respondents in Section IV-C exemplify the analysis protocol, and should not be read as a description of patients. The no-show model is an abstract score that is calibrated with AUC, instead of trained classifiers. Conclusion validity: Power estimates are based on the quality of the

measurement in Table V which would be higher with smaller samples. The code released will be re-run with local operational data and using the survey responses obtained using the instrument in Table 3.

6. Conclusion

This paper suggested and discussed an integrated framework for the study of impact of AI supported Digital transformation on Patient satisfaction. It utilizes a factorial discrete-event simulation of AI interventions, a probabilistic satisfaction-response model, and a structural model based on SERVQUAL for which the properties of the inferences were confirmed through Monte Carlo simulation. As long as the assumptions are true, AI no-show management improves throughput, but with only it in place, it will decrease satisfaction. Improved check-in time and mitigated overbooking delay issues, thanks to digital check-in. The main benefit of AI triage was to deal with urgent patients primarily during peak times. The overall system increased the throughput at a relatively constant level of satisfaction and a group of moderate overbooking intensities maintained waits similar to the traditional level. Satisfaction inequities were a result of accuracy gaps by digital literacy that were largely mitigated with simple operations. On the side of the measurement, unbiased testing of the theories was required and the small effects of central interest required a few hundred or over a thousand people. Future work will involve linking the data from the partner clinics to the DES and calibrating the instrument to time-stamp data and estimate the net digital-contact effect β_{dc} that the PSA determined to be the decisive effect.

References

1. Agarwal, R., Gao, G., DesRoches, C., & Jha, A. K. (2010). Research commentary—The digital transformation of healthcare: Current status and the road ahead. *Information Systems Research*, 21(4), 796–809. <https://doi.org/10.1287/isre.1100.0327>
2. Kraus, S., Schiavone, F., Pluzhnikova, A., & Invernizzi, A. C. (2021). Digital transformation in healthcare: Analyzing the current state-of-research. *Journal of Business Research*, 123, 557–567. <https://doi.org/10.1016/j.jbusres.2020.10.030>
3. Nelson, A., Herron, D., Rees, G., & Nachev, P. (2019). Predicting scheduled hospital attendance with artificial intelligence. *npj Digital Medicine*, 2, Article 26. <https://doi.org/10.1038/s41746-019-0103-3>
4. Dantas, L. F., Fleck, J. L., Cyrino Oliveira, F. L., & Hamacher, S. (2018). No-shows in appointment scheduling—A systematic literature review. *Health Policy*, 122(4), 412–421. <https://doi.org/10.1016/j.healthpol.2018.02.002>
5. Levin, S., Toerper, M., Hamrock, E., Hinson, J. S., Barnes, S., Gardner, H., Dugas, A., Linton, B., & Kirsch, T. (2018). Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the Emergency Severity Index. *Annals of Emergency Medicine*, 71(5), 565–574.e2. <https://doi.org/10.1016/j.annemergmed.2017.08.005>
6. Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Affairs*, 33(7), 1123–1131. <https://doi.org/10.1377/hlthaff.2014.0041>
7. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
8. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *The New England Journal of Medicine*, 380(14), 1347–1358. <https://doi.org/10.1056/NEJMr1814259>
9. Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12–40.
10. Cronin, J. J., Jr., & Taylor, S. A. (1992). Measuring service quality: A reexamination and extension. *Journal of Marketing*, 56(3), 55–68.

- <https://doi.org/10.1177/002224299205600304>
11. Babakus, E., & Mangold, W. G. (1992). Adapting the SERVQUAL scale to hospital services: An empirical investigation. *Health Services Research*, 26(6), 767–786.
12. Bleustein, C., Rothschild, D. B., Valen, A., Valatis, E., Schweitzer, L., & Jones, R. (2014). Wait times, patient satisfaction scores, and the perception of care. *The American Journal of Managed Care*, 20(5), 393–400.
13. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
14. Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research*, 69(10), 3998–4010. <https://doi.org/10.1016/j.jbusres.2016.06.007>
15. Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24. <https://doi.org/10.1108/EBR-11-2018-0203>
16. Gupta, D., & Denton, B. (2008). Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions*, 40(9), 800–819. <https://doi.org/10.1080/07408170802165880>
17. Guy, R., Hocking, J., Wand, H., Stott, S., Ali, H., & Kaldor, J. (2012). How effective are short message service reminders at increasing clinic attendance? A meta-analysis and systematic review. *Health Services Research*, 47(2), 614–632. <https://doi.org/10.1111/j.1475-6773.2011.01342.x>
18. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
19. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
20. Greenhalgh, T., Wherton, J., Papoutsis, C., Lynch, J., Hughes, G., A'Court, C., Hinder, S., Fahy, N., Procter, R., & Shaw, S. (2017). Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *Journal of Medical Internet Research*, 19(11), e367. <https://doi.org/10.2196/jmir.8775>
21. Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17, Article 195. <https://doi.org/10.1186/s12916-019-1426-2>
22. Donabedian, A. (1988). The quality of care: How can it be assessed? *JAMA*, 260(12), 1743–1748. <https://doi.org/10.1001/jama.1988.03410120089033>
23. Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing*, 60(2), 31–46. <https://doi.org/10.1177/002224299606000203>
24. Kruse, C. S., Krowski, N., Rodriguez, B., Tran, L., Vela, J., & Brooks, M. (2017). Telehealth and patient satisfaction: A systematic review and narrative analysis. *BMJ Open*, 7(8), e016242. <https://doi.org/10.1136/bmjopen-2017-016242>
25. Norman, C. D., & Skinner, H. A. (2006). eHEALS: The eHealth Literacy Scale. *Journal of Medical Internet Research*, 8(4), e27. <https://doi.org/10.2196/jmir.8.4.e27>
26. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
27. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), Article 115. <https://doi.org/10.1145/3457607>
28. Briggs, A., Claxton, K., & Sculpher, M. (2006). *Decision modelling for health economic evaluation*. Oxford University Press.
29. Marino, S., Hogue, I. B., Ray, C. J., & Kirschner, D. E. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology*, 254(1), 178–196. <https://doi.org/10.1016/j.jtbi.2008.04.011>
30. Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares*. Physica-Verlag. <https://doi.org/10.1007/978-3-642-52512-4>
31. Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295–336). Lawrence Erlbaum Associates.
32. Dijkstra, T. K., & Henseler, J. (2015). Consistent partial least squares path modeling. *MIS Quarterly*, 39(2), 297–316. <https://doi.org/10.25300/MISQ/2015/39.2.02>
33. Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50. <https://doi.org/10.1177/002224378101800104>
34. Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
35. Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall/CRC. <https://doi.org/10.1007/978-1-4899-4541-9>
36. Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J.-H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *European Journal of Marketing*, 53(11), 2322–2347. <https://doi.org/10.1108/EJM-02->

- 2019-0189
37. Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). A primer on partial least squares structural equation modeling (PLS-SEM) (3rd ed.). SAGE Publications.
38. Henseler, J., & Chin, W. W. (2010). A comparison of approaches for the analysis of interaction effects between latent variables using partial least squares path modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 17(1), 82–109. <https://doi.org/10.1080/10705510903439003>
39. Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration*, 11(4), 1–10. <https://doi.org/10.4018/ijec.2015100101>
40. Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
41. Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8
42. Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, 73(6), 913–934. <https://doi.org/10.1177/0013164413495237>
43. Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal*, 28(1), 227–261. <https://doi.org/10.1111/isj.12131>
44. Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354–373. <https://doi.org/10.1037/a0029315>
45. Reinartz, W., Haenlein, M., & Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM. *International Journal of Research in Marketing*, 26(4), 332–344. <https://doi.org/10.1016/j.ijresmar.2009.08.001>