

Reducing Healthcare Waiting Times in Kuwait

Scheduling, Demand Forecasting,
and Administrative Automation

Jasem Al Salem

Version 1.0 / 4 October 2026

A detailed evidence review and proposed evaluation framework for improving timely access to non-emergency outpatient care.

Kuwait policy context · Intervention evidence · Data and system design · Controlled evaluation · Economics · Safety and equity

Research status

No patient records analysed. No clinical trial conducted. Numerical scenarios are hypothetical and reproducible. This report is not peer reviewed.

Contents

Abstract	3
Executive synthesis	4
1. Research purpose and scope	4
2. Evidence approach and limits	5
3. Kuwait as the study setting	6
4. Define waiting before attempting to reduce it	8
5. Why queues form	9
6. What the intervention evidence can establish	10
7. The proposed intervention package	11
8. Data and interoperability design	12
9. Forecasting and prediction design	13
10. Scheduling, clinical priority, and fairness	15
11. Reproducible numerical illustrations	16
12. Proposed prospective study	18
13. Economics, productivity, and growth	21
14. Safety, privacy, and accountability	22
15. Implementation and scale decisions	23
16. Limitations and unresolved questions	24
17. Conclusion	25
Research transparency statement	25
Appendix A. Selected intervention evidence	26
Appendix B. Minimum data dictionary	26
Appendix C. Outcome specification and analysis checks	28
Appendix D. Operational failure register	29
Appendix E. Interview and observation guide	29
Appendix F. Reproducibility and next research steps	30
References	31

Abstract

Background. Delays in obtaining and completing healthcare can arise from several interacting processes: referral handling, appointment allocation, missed appointments, uneven session capacity, diagnostics, and administrative work. Digital booking changes how a patient enters a queue; it does not necessarily increase the rate at which appropriate care is completed. For Kuwait, the research question is therefore which combinations of operational redesign and technology can improve timely access within actual staffing, clinical, and institutional constraints.

Objective. To develop a detailed, testable framework for reducing non-emergency outpatient waiting times in Kuwait using scheduling improvements, demand forecasting, and administrative automation, while preserving clinical priority, continuity, equitable access, and staff capacity.

Methods. A targeted narrative evidence review, completed on 4 October 2026, was combined with operations analysis and the design of a prospective evaluation. Official Kuwait and World Health Organization materials inform the setting; published intervention studies inform candidate mechanisms; reporting guidelines inform the proposed study. Reproducible arithmetic scenarios illustrate capacity and cost relationships. The review is not a systematic review, and the scenarios are not estimates of Kuwait's performance.

Synthesis and proposal. The recommended sequence is to establish reliable referral and appointment event records; improve cancellation, reminder, and rebooking processes; manage usable capacity and waiting-list priority; and then evaluate whether predictive tools improve on transparent rules. The proposed primary evaluation compares an operational improvement package with usual care, using a fixed-horizon measure of time without a completed first consultation. A subsequent comparison isolates the incremental contribution of predictive tools. Additional outcomes include completion within locally agreed clinical windows, in-clinic delay, safety, patient burden, equity, staff workload, and incremental cost.

Conclusion. Technology could reduce avoidable waiting where it changes a binding operational constraint. It cannot independently resolve sustained demand exceeding usable capacity. Kuwait-specific benefits require prospective measurement, appropriate local governance, and explicit accounting for patients who remain waiting. The appropriate output of an initial programme is a verified improvement in access and care delivery, rather than a high prediction score or a larger number of digital bookings.

Keywords: Kuwait; outpatient access; waiting time; appointment scheduling; health service productivity; artificial intelligence; demand forecasting; patient navigation; implementation research.

Executive synthesis

This report proposes a focused health service research programme for Kuwait. Its central hypothesis is that a coordinated package of better referral information, accessible reminders and cancellations, clinically governed waiting-list management, and realistic session planning can improve completed access to outpatient care. Predictive models should be evaluated as an additional component of that package. Their value depends on whether they cause a useful operational decision and whether that decision improves patient outcomes at acceptable cost.

Five decisions define the proposed programme. First, measure the patient's full pathway from the original eligible referral to a completed consultation, without erasing delays when a referral is returned or a booking changes. Second, distinguish the queue for an appointment from the queue inside the clinic. Third, protect people who cannot use a smartphone, respond immediately, travel at short notice, or navigate a service in its default language. Fourth, hold clinical priority and actual staffing constraints above predictions about attendance. Fifth, compare the package with both usual care and, at a later stage, the same package without predictive tools.

The proposed initial setting is a small number of non-emergency adult outpatient services with identifiable referral routes, stable clinical leadership, and auditable appointment records. This is a study-design choice, not a statement that a named hospital has agreed to participate. Emergency care, urgent cancer pathways, paediatric services, and time-critical specialist pathways require separate designs. A shared data definition can eventually support comparison across services, but a single scheduling rule should not be imposed on clinically different pathways.

An illustrative capacity calculation shows why the distinction matters. If 10,000 genuinely usable monthly slots produce 8,000 completed consultations, raising completion to 8,500 adds 500 consultations. If 8,500 equivalent new episodes arrive each month, that improvement only stabilises the queue. It does not clear the existing backlog. Every number in this example is an author-selected assumption. The calculation demonstrates a decision rule, not a forecast for Kuwait.

The immediate research deliverables should be a validated baseline, an evidence and data-quality register, a co-designed intervention manual, a registered analysis plan, and a controlled evaluation. A national rollout should depend on sustained clinical and operational benefit, acceptable equity and workload effects, and the ability to operate safely when software or communications fail.

1. Research purpose and scope

1.1 The question that the paper addresses

The primary question is: can an operational package supported by digital tools reduce the time that eligible patients in Kuwait spend without their first completed non-emergency outpatient consultation, compared with existing practice? The secondary question is whether demand forecasts, attendance-risk estimates, or automated administrative assistance add value beyond simpler workflow improvements.

These questions separate three forms of contribution. Digitisation makes information or transactions available electronically. Operational redesign changes responsibilities, handoffs, priority rules, or resource allocation. Artificial intelligence generates predictions, extracts information, or assists decisions. They may work together, but a beneficial package does not establish that its AI component was necessary. Attribution requires a comparison that holds the other components reasonably constant.

The unit of patient benefit is an appropriately completed care step. A booked slot is an intermediate event. A reminder sent is an activity. A model's discrimination is a technical property. None is sufficient evidence of better

access. A patient who receives several digital confirmations but still cannot obtain an appropriate appointment has not experienced the intended improvement.

1.2 Boundaries of the proposed initial study

The initial study would include new, clinically eligible, non-emergency referrals to participating adult outpatient services. Clinical staff would specify eligibility and urgency categories before recruitment. Follow-up appointments remain operationally relevant because they use the same capacity, but their appropriate timing differs from that of a new referral. They should be recorded and analysed separately, with explicit checks that improving new-patient access does not delay necessary follow-up.

The intervention would support administrative coordination and capacity decisions. It would not diagnose patients, autonomously determine urgency, deny care, or replace professional assessment. A forecast might recommend opening an additional session if authorised staffing exists; it cannot make unavailable specialists, rooms, equipment, interpreters, or laboratory capacity appear.

The paper is written for health service leaders, clinicians, researchers, technical teams, and policymakers considering a Kuwait pilot. It provides an implementation and evaluation design detailed enough to critique and adapt. It does not supply a clinically approved protocol, a procurement specification endorsed by the Ministry of Health, or a legal opinion.

2. Evidence approach and limits

2.1 Search and selection

The evidence search used web-accessible official publications, journal articles, and bibliographic records. Search concepts combined Kuwait, Ministry of Health, primary healthcare, outpatient appointment, referral, waiting time, reminder, missed appointment, advanced access, no-show prediction, artificial intelligence evaluation, and health economic reporting. Priority was given to Kuwait government and WHO sources for institutional facts, intervention research for effectiveness, and primary methodological guidance for evaluation design.

Selection was purposive: sources were retained when they addressed the setting, a plausible intervention mechanism, a consequential implementation risk, or a design decision. No claim is made that all eligible databases, languages, or unpublished evaluations were searched. There was no duplicate independent screening, formal meta-analysis, or comprehensive risk-of-bias scoring. Findings should therefore be read as a structured narrative synthesis and research proposal, not a pooled estimate of expected effect.

The search date was 4 October 2026. Publication dates and the periods studied are distinguished where relevant. An older clinical study may remain useful for a mechanism, while older Kuwait service descriptions cannot establish the exact current configuration of every facility. The companion evidence register records the narrow use and limitation of each source.

2.2 Four evidence categories

Verified source statements concern what a cited source reports, within that source's setting and dates.

Interpretations explain why such evidence may matter for a Kuwait study. **Proposed design choices** describe what a future research team could do. **Illustrative calculations** use stated assumptions to expose operational relationships. These categories must remain distinct when the report is adapted into policy summaries or presentations.

There is no national, patient-level waiting-time dataset analysed in this report. It does not establish Kuwait's median specialist wait, national missed-appointment rate, current administrative cost per visit, or the proportion

of appointments that could safely be recovered. These quantities belong in the baseline study. Absence of a verified estimate in this review is not evidence that the measure does not exist within government or provider systems.

2.3 Transferability and uncertainty

An intervention can work in one setting and fail elsewhere because the constraint differs. Reminders address communication and memory; they do not fix unaffordable transport or lack of suitable slots. An electronic referral can accelerate transfer while leaving specialist capacity unchanged. Faster triage may expose previously hidden unmet need and initially increase the recorded waiting list. Such changes require explanation rather than immediate classification as failure.

The report therefore uses external evidence to identify mechanisms worth testing, not to import an effect size into a Kuwait business case. Local language, entitlement arrangements, service organisation, communication preferences, and staffing patterns must be established through site work. Where a source measures attendance, the report does not automatically interpret it as evidence of shorter referral-to-consultation time.

3. Kuwait as the study setting

3.1 Build on existing services

Kuwait has existing public healthcare institutions and electronic service channels; the relevant design question is how an intervention fits those services. WHO's Kuwait primary healthcare case study examines the response during March to December 2020 and is useful historical context, rather than a current audit of outpatient operations. [1] Kuwait Government Online explicitly describes a hospital and specialist appointment service for citizens and residents, and its Salem application page describes appointment and reminder functions. These descriptions establish available service concepts, not universal coverage or measured effectiveness. [3, 4]

The pilot should begin with a site-specific map: where referrals originate, who verifies them, who assigns urgency, which systems store bookings, how patients change appointments, and how completion is recorded. The map should include telephone and in-person work. An electronic interface can conceal substantial manual reconciliation behind it; only observation and event-level data can establish the actual process.

The MOH standardised outpatient policy, code A-ADM-11, already specifies referral triage and appointment benchmarks. Section 8.3.4 lists 1-2 weeks for urgent stable referrals, 2-4 weeks for new non-urgent referrals, 4-8 weeks for existing-file follow-ups, and 12-24 weeks for regular follow-up or prescription reassessment. Specialty requirements and physician-directed earlier care qualify these ranges. Section 8.3.5 calls for three-monthly auditing and reallocation when standards are exceeded. The inspected cover prints effective date 1/10/2023 and revision date 1/10/2025. These are policy statements, not observed waiting times; the current applicable version must be confirmed with a participating institution. [2]

The study should map its event definitions to the institution's implementation of that policy, then report both policy-window performance and the independent patient-centred access measure. It should not invent a competing national target or assume that a range in the document defines identical measurement rules for every specialty.

3.2 Kuwait-specific design questions

The programme should test language and channel needs directly with patients and staff. Arabic and English are obvious starting points for interface review, but the language requirements of a participating population should not be inferred from nationality alone. Interpreting support, literacy, disability access, caregiver involvement, and shared-phone use can affect whether a message is understood and whether a patient can act on it.

Service eligibility and referral routing must be verified for each participating provider and patient group. The research should not presume that all residents follow an identical pathway or that the same administrative rules apply across public and private services. Where eligibility checks are required, their duration should be visible rather than disappearing before the recorded start of waiting.

Calendar effects are hypotheses to test locally. Ramadan, Eid periods, school calendars, annual leave, and changes in clinic hours may alter demand, attendance, or usable capacity. The model must distinguish a real demand change from a service closure. Heat, travel time, and transport arrangements may matter for individual patients, but the study should collect relevant information rather than treating them as established explanations for missed visits.

3.3 Baseline questions that need local answers

The baseline must determine how much delay is administrative, how much reflects unavailable clinical capacity, and how much arises from patient preference or difficulty accessing care. It should quantify referral arrivals, backlog age, completed visits, late cancellations, unused slots, staff absence, room availability, and session start-time reliability. It should also identify repeated testing, incomplete documentation, and appointments that occur before necessary investigations are ready.

A facility with unused suitable slots and communication failures is a plausible candidate for cancellation recovery and navigation. A facility with fully utilised clinicians and sustained excess demand needs a capacity or pathway response. A facility with long in-clinic queues but short booking delays needs a session-flow response. Selecting an intervention before identifying the bottleneck risks improving a visible metric while leaving the patient's problem intact.

Local automation research also warrants careful interpretation. A 2026 Kuwait pharmacy study surveyed 416 respondents from four governmental hospitals using data collected in June-July 2024; 356 reported perceived time savings. This supports investigating patient experience, but it does not quantify a causal reduction in waiting time. [19] Official health statistics and a WHO-partnered PHC costing report offer contextual resources; neither is a substitute for linked referral and appointment events at pilot sites. [22, 23]

4. Define waiting before attempting to reduce it

4.1 Separate clocks along the care pathway

Clock	Operational definition	Why it matters
Referral processing	Original referral receipt to completed clinical and administrative review	Exposes missing information, handoffs, and repeated returns
Appointment access	Original eligible referral receipt to first completed appropriate consultation	Captures the patient's main access delay, including rebooking
Booking lead time	Appointment creation to scheduled appointment start	Useful for scheduling analysis, but does not capture pre-booking delay
Clinic-day delay	$\max(0, \text{clinician start} - \max(\text{scheduled start}, \text{patient readiness}))$	Distinguishes service delay from early arrival; retain early starts separately
Total patient time	Arrival to departure, with component timestamps where available	Includes check-in, tests, consultation, pharmacy, and handoffs
Diagnostic delay	Valid test order to result available for clinical action	Identifies downstream constraints
Follow-up timeliness	Actual follow-up date relative to clinically requested window	Detects displaced delay when new referrals receive priority

Patient readiness should be defined before analysis. If readiness requires completion of a check-in step controlled by the service, report both arrival-based and readiness-based waits; otherwise slow registration can be excluded from the apparent delay. Do not use a later timestamp solely because it produces a more favourable result.

The main access clock should preserve the original referral date when information is returned for correction, a patient is rebooked, or a slot is moved. Record patient-requested deferral as a separate state and report both total elapsed time and any prespecified service-attributable measure. The patient-experience measure should remain visible even when a contractual reporting rule permits a clock pause.

4.2 Include patients who are still waiting

Completed-visit records alone create a selection problem. They exclude the patients with the longest unresolved waits, those who disengage, and those who seek care elsewhere. The study needs an incident referral cohort and a waiting-list census, not simply an average calculated from this month's completed appointments.

For each cohort, report completion by fixed horizons, the number still waiting, the number transferred, the number declining or deferring, and deaths or urgent escalations where ascertainable. A fall in the median among completed visits can coexist with an increase in unresolved long waits. Both distributions must therefore be shown.

4.3 Protect the denominator

Every eligible referral should have an auditable final or current state. Removing duplicates may be appropriate, but the rule must identify genuine duplicate episodes without deleting separate clinical needs. Administrative rejection and withdrawal need reason codes. A deletion log should preserve who changed the record and why. Performance incentives must not reward reclassification or silent removal of difficult cases.

Reports should contain counts alongside percentages. A clinic completing 95% of a much smaller accepted cohort may provide worse population access if referral acceptance became more restrictive. Referral acceptance, rejection, and transfer volumes should be reviewed alongside waiting times.

5. Why queues form

5.1 Demand, usable capacity, and variation

Nominal capacity is the number of slots shown in a schedule. Usable capacity is the number that can actually deliver appropriate care with the required people, space, equipment, and preparation. Completed capacity is what patients ultimately receive. The gaps between these quantities have different causes and require different responses.

A session may contain unbooked slots because the referral list is not visible, booked slots that become unused after late cancellations, or fully attended slots whose consultation durations exceed the template. More bookings address only some of these problems. A schedule filled to its nominal limit can still be operationally unsafe if the template ignores case complexity or necessary documentation time.

At a simplified monthly level, backlog evolves as $B(t+1) = \max[0, B(t) + A(t) - C(t)]$, where A is incoming eligible demand and C is completed equivalent episodes. Real services also have transfers, withdrawals, changing urgency, and repeated visits, which need additional state transitions. The simple equation exposes one essential condition: reducing an existing backlog requires completed capacity to exceed incoming demand for a sustained period.

5.2 The importance of variability

Even when average capacity exceeds average demand, uneven arrivals and variable consultation duration can create substantial delay. Reserving no flexibility for urgent cases, complex visits, or staff disruption makes the schedule fragile. Improving average utilisation without considering its distribution can therefore increase patient waiting and staff overtime.

Little's relationship connects the long-run average number in a stable system with throughput and average time in that system. It is useful for consistency checks when boundaries and units match. It should not be used to infer an exact waiting-time distribution or applied uncritically to a rapidly growing queue. [18] Section 11 provides a separate simplified queueing illustration and explains its restrictive assumptions.

5.3 Bottlenecks can move

A faster referral process may move the queue to diagnostics. A successful reminder programme may increase attendance enough to expose insufficient nursing support. A cancellation-recovery programme may help mobile patients while leaving those who need transport behind. A responsible evaluation follows the whole pathway and asks whether the original constraint has been resolved or merely shifted.

This is also why a service-wide productivity claim requires more than a local time saving. If clinicians spend less time searching documents but more time correcting automated summaries, the net benefit may be zero. If a new booking channel transfers work from reception to patients and caregivers, the provider's administrative metric can improve while societal burden rises.

6. What the intervention evidence can establish

6.1 Reminders and accessible communication

Appointment reminder research supports testing low-complexity communication interventions before assuming that advanced prediction is necessary. Message content, timing, and ease of cancellation can influence behaviour. External results should be interpreted at the outcome actually measured: reduced non-attendance is not automatically a reduction in specialist waiting time. [5, 6]

For a Kuwait pilot, the operational hypothesis is that reminders paired with an accessible response route will improve attendance or permit earlier release of unwanted slots. The test should measure successful delivery, comprehension in a sample, cancellations with enough notice, rebooking success, and completed care. Counting outgoing messages alone cannot distinguish a useful service from an unread notification stream.

6.2 Advanced access and schedule redesign

Advanced access seeks to reduce delay by managing demand and supply so appointments can be offered closer to need. The cited review concerns primary care; transfer to specialist referral queues is indirect because preparation, referral gating, and diagnostic resources differ. Its evidence does not justify assuming that every service will achieve the same improvement, or that shorter booking waits necessarily preserve continuity. [7] The relevant lesson for this proposal is to study the interaction between timely access, clinician continuity, case mix, and follow-up requirements.

A specialist service should not convert all capacity into short-notice appointments merely because that makes availability appear better. Some patients need time for transport, leave from work, interpreting support, investigations, or caregiver arrangements. A balanced template can preserve advance booking while reserving a clinically governed fraction for urgent or recently freed capacity.

6.3 Prediction must be attached to a beneficial action

Predicting non-attendance is a different task from preventing it. A model can accurately identify patients who face transport, language, or employment barriers while offering no effective intervention. A useful programme links predictions to support, such as a call to clarify a referral or help reorganise an appointment, and evaluates whether that support improves completed access. Prospective intervention evidence is more decision-relevant than a retrospective accuracy score. [8]

Neither risk scores nor predicted convenience should override clinical urgency. Patients with a history of missed visits must not automatically lose access or be allocated inferior slots. Such a rule could amplify the barriers that generated the prediction. The programme should document whether a prediction changes the offered service, who benefits, and who receives additional burden.

6.4 Bias and the choice of target

A widely cited healthcare algorithm study demonstrated that a seemingly useful proxy can encode unequal access and produce biased allocation. Its specific setting differs from Kuwaiti scheduling; its relevance is the warning against substituting an administrative proxy for the outcome that matters. [9] Here, past attendance, rapid message response, or low expected visit duration should not become proxies for deserving care.

The proposed programme should optimise timely completion subject to clinical and equity constraints. It should not optimise attendance percentage by discouraging patients with complex needs, or maximise visit count by replacing longer necessary appointments with shorter easy ones. The choice of target is a clinical and governance decision, not merely a modelling decision.

7. The proposed intervention package

7.1 A reliable referral work queue

Create a visible work queue for received referrals with ownership, status, age, and the next required action. Administrative completeness checks can flag missing fields, but a designated clinical team should control clinical eligibility and urgency. Returned referrals retain their original identity and timestamp, with a reason and a route to correction.

Use a small, controlled set of states: received, awaiting information, awaiting clinical review, ready to book, offered, booked, attended, completed, deferred, transferred, withdrawn, and unresolved. Sites may need extra states, but each must have an operational meaning. A dashboard should surface aged items and failed handoffs to the person responsible for resolving them.

7.2 Reminders, cancellation, and rebooking

Offer patients a choice of available contact channels and verify contact details at an appropriate encounter. Reminder timing should be piloted, not declared universally optimal. A practical starting experiment could compare a simple reminder with a reminder that includes a clear way to confirm, request help, cancel, or rebook. Content should state the location and preparation needed without unnecessarily exposing sensitive information on a shared device.

Cancellation should release capacity only after a valid request is confirmed. A request for help must not be interpreted as cancellation. If a patient needs a later date, the workflow should preserve continuity and avoid placing the patient at the end of an unrelated queue. Staff must be able to resolve ambiguous responses without asking the patient to restart the process.

7.3 A governed short-notice list

Invite patients to opt into earlier appointments when clinically appropriate. Record their practical availability, travel notice requirements, preparation needs, and whether an approved caregiver may respond. Declining an earlier slot must not remove the original booking or lower clinical priority.

Select eligible recipients using a documented rule based on urgency, waiting time, compatibility with the slot, and patient preferences. Do not make the entire service first-click-wins. One possible design is to offer slots in priority bands with a reasonable response window and a staff-assisted route; its performance and fairness need evaluation. Reserve the slot transactionally so two patients cannot receive a final confirmation for the same capacity.

7.4 Session templates that reflect actual work

Classify visit types with clinical input and measure their duration distribution. Templates should allow necessary variation for new assessments, follow-up, procedures, interpreting needs, and complex cases. Avoid inferring that a group should receive less time because historical consultations were shorter; historical duration can reflect constrained care rather than appropriate care.

Plan clinician, nursing, room, and equipment capacity together. A forecast-driven recommendation to add appointments should show the resource assumption and the expected effect on overtime and in-clinic delay. Managers should be able to decline the recommendation with a reason that becomes part of the evaluation record.

7.5 Referral and appointment navigation

A small navigation function can help patients resolve missing information, understand the location or preparation, and change an unsuitable time. The function should be available through staff-assisted channels, including for patients whom digital systems fail to reach. Navigation is a service intervention with staffing costs, not a free side effect of prediction.

Support should be framed as an offer of assistance. Staff should avoid language that blames patients for missed visits when the underlying reason is unknown. Reason codes should distinguish patient-reported barriers, inability to contact, service cancellation, duplicate booking, and other circumstances. A broad no-show category conceals actionable differences.

7.6 Administrative automation with verification

Begin with deterministic tasks: matching identifiers under approved rules, checking required fields, detecting inconsistent dates, generating approved reminder templates, and routing work to the correct team. These tasks are often easier to validate than open-ended text generation. The system should show the source of a field and preserve an audit trail when it changes.

If language models are introduced, restrict the initial role to draft extraction or summarisation for staff review. Any generated assertion about symptoms, urgency, medication, preparation, or eligibility must be traceable to an authorised source and checked before it affects care. Confidence language from the model is not a validation method. A missing or contradictory detail should produce a request for review rather than an invented completion.

7.7 Operating through failures

Maintain a documented downtime process: retain a current authorised schedule, record changes made during an outage, provide a contact route, and reconcile the authoritative system before sending confirmations. Duplicate messages after recovery should be suppressed. If a forecast service fails, a safe default schedule should remain available. If identity or slot state cannot be verified, the system should defer the transaction to staff rather than guess.

8. Data and interoperability design

8.1 Minimum information model

The study needs linked referral, appointment, encounter, resource, and communication events. These are related but distinct objects. A referral may generate several appointment offers; an appointment may be rescheduled several times; a patient may have multiple legitimate episodes. Collapsing them into one spreadsheet row loses the sequence necessary to measure waiting accurately.

The research dataset should use pseudonymous person and episode identifiers. A separately controlled linkage process should reconcile records where authorised. Phone numbers and direct identifiers belong in operational systems with restricted access, not in general model-development extracts. Text fields should be minimised because they often contain unnecessary sensitive information.

8.2 Event timestamps and provenance

Store event time, ingestion time, source system, actor or process, and the previous and new state. Keep the local service timezone explicit and use a consistent machine-readable representation. Scheduled start, actual arrival, clinician start, and encounter completion must not be treated as interchangeable. Corrected timestamps should retain their history.

Data-quality checks should identify impossible sequences, negative durations, duplicate confirmations, orphaned appointments, unlinked completions, repeated clock resets, and unexplained missing outcomes. Sampling should compare electronic records with the underlying operational record through an approved audit process. A high completeness percentage alone does not establish that the meaning of a field is correct.

8.3 Standards as a starting point

HL7 FHIR R4 provides appointment and slot resources that can inform the design of exchanges between systems. [17] Citing the standard does not establish that a Kuwait provider implements that version or exposes an integration interface. The pilot must verify local capabilities, profile requirements, terminology, authorisation, and event semantics with each system owner.

The architecture should preserve one authoritative booking state. Predictive services can read approved features and return recommendations, but a controlled transaction service should enforce slot availability, identity, consent or contact preference, and clinical constraints. Idempotent transaction identifiers should prevent duplicated bookings when network requests are retried.

8.4 Data access and retention

Use role-based permissions and separate operational access from research analysis. The project should specify lawful authority, ethics approval where required, data-sharing agreements, retention periods, deletion processes, and incident responsibilities before processing patient data. Local legal and institutional review must determine which Kuwait rules apply to each entity and data flow; this report makes no assertion that a general technology regulation alone authorises healthcare data processing.

Public research outputs should contain aggregate results, the analysis code, and clearly synthetic examples where useful. Pseudonymisation does not make detailed patient trajectories safe to publish. Small cells and unusual combinations of attributes can permit re-identification and need disclosure review.

9. Forecasting and prediction design

9.1 Forecast decisions before choosing a model

A demand forecast should answer a defined operational question. Next-day forecasts may support staffing adjustments where rosters permit them. Weekly forecasts may inform session allocation. Longer horizons may support recruitment or capacity planning. A prediction delivered after the roster is fixed may be accurate but unusable. The intervention should therefore specify decision lead time, responsible owner, available action, and cost of adjustment.

Forecast eligible referral arrivals and usable capacity separately. Completed visits are constrained by supply and therefore do not reveal all demand. Training a demand model only on completed appointments can reproduce historical rationing. Unbooked referrals, unsuccessful booking attempts where validly recorded, and pending investigations may provide additional information, but their meaning and completeness need checking.

9.2 Compare with simple baselines

Start with seasonal-naïve forecasts, such as the same weekday in recent comparable weeks, and transparent averages adjusted for known closures. Compare these with regression or time-series approaches before introducing more complex models. An international outpatient forecasting study illustrates model evaluation in a specific hospital setting; it does not establish that forecast accuracy itself shortens queues. [20]

Use rolling-origin evaluation: train on information available before each forecast date, generate predictions, and compare with later observations. Keep a final time period untouched until model choices are fixed. If several sites participate, test generalisation at a site not used for development where feasible. A random row split can leak shared patient, clinic, or future operational information.

9.3 Evaluate uncertainty and the decision

Report mean absolute error in clinically interpretable units, such as referrals per clinic-week, and the coverage of prediction intervals. Percentage error becomes unstable when counts approach zero. Examine systematic underprediction during high-demand periods, not just overall averages. For staffing decisions, underprediction and overprediction can have different consequences and should be costed separately.

The operational evaluation should compare schedules or staffing decisions made with and without the forecast. Record whether staff saw the recommendation, whether they acted, why they overrode it, and whether the expected resource change occurred. A forecast that is routinely ignored may indicate poor fit with operational constraints rather than a need for a more sophisticated algorithm.

9.4 Attendance prediction

The target must distinguish late cancellation, early cancellation, non-arrival, provider cancellation, and completed attendance. These events have different implications for capacity. Training examples should reflect what was knowable when the intervention would be triggered. For example, a response to a later reminder cannot appear in a score that was supposedly computed before that reminder.

Evaluate calibration as well as discrimination. If a model assigns a group a 30% risk, the observed event frequency should be approximately compatible with that estimate, allowing for sampling uncertainty. Report performance by site, appointment type, lead time, and locally justified patient groups. A high area under the receiver operating characteristic curve does not tell managers whether a small calling team will make useful calls.

For a constrained outreach team, compare model-based targeting with alternatives: oldest unresolved referrals, patients reporting difficulty, simple prior-history rules, and a random or broad offer of support. The relevant endpoint is the incremental completed care produced per unit of outreach resource, together with equity and patient experience. A rapid review of implemented prediction-based interventions reinforces the need to distinguish prediction from the action linked to it. [21]

9.5 Model versioning and change control

Freeze the evaluated model, features, thresholds, and user interface during a trial phase except for documented safety fixes. Store the version used for every recommendation. Retraining can change the intervention and should not happen invisibly. Evaluate calibration drift, missing-feature rates, recommendation acceptance, and subgroup outcomes over time.

The reporting plan should use TRIPOD+AI for prediction-model components, with enough detail to reproduce cohort construction, predictors, validation, and performance reporting. [11] Reporting compliance is not proof of clinical utility. A transparent weak model remains weak; a technically strong model can still produce an unhelpful intervention.

10. Scheduling, clinical priority, and fairness

10.1 A constrained allocation problem

An appointment allocation engine should first apply hard constraints: appropriate specialty, clinician capability, required equipment, clinically permitted timing, patient availability, and approved service eligibility. Within the remaining feasible options, it can seek to reduce waiting and unused capacity. A weighted objective may combine access delay, overtime, idle capacity, continuity, and patient travel burden, but weights must be agreed and documented rather than tuned to produce attractive utilisation figures.

Clinical urgency and safety are not merely small penalties in a numerical objective. They should be explicit constraints or escalation rules. The tool should explain why a slot is suitable, what priority rule was applied, and when staff intervention is required. Staff must be able to override an assignment while preserving an auditable reason.

10.2 Overbooking requires a separate decision

Do not include automatic overbooking in the initial package. Attendance improvement already increases workload within the existing schedule. Adding bookings on top of uncertain attendance can increase queues, shorten consultations, and extend staff shifts. A later overbooking study would need a separate risk assessment, simulation, and prospective safety evaluation.

If explored, the model must account for correlated attendance and consultation duration, not simply sum independent probabilities. Shared transport disruptions, weather, holidays, and changes in reminder policy can invalidate historical assumptions. The objective should include the upper tail of overtime and waiting, not only expected utilisation. Patients should not be treated as interchangeable units of short, predictable work.

10.3 Fair access to recovered capacity

Recovered appointments should not systematically favour patients who can answer instantly, own newer devices, or travel without assistance. Monitor offers as well as acceptances. A low acceptance rate may indicate that the offered times are unsuitable, not a lack of interest in care. Preserve the original appointment when an earlier offer is declined.

Choose equity strata through local consultation and lawful data access. Potential dimensions include age, sex, disability or assistance needs, preferred language, site, booking channel, and service eligibility group where appropriate. Missing or small categories should not be interpreted as evidence of no disparity. Publish uncertainty and use protected aggregate reporting.

10.4 Human factors

Staff should see a manageable work queue with reasons and next actions, not an unfiltered stream of alerts. Measure alerts per shift, time to resolve them, overrides, unresolved items, and interruptions to direct care. Co-design with reception, nursing, medical records, clinicians, and patients should occur before the system is treated as ready for evaluation.

DECIDE-AI provides reporting guidance for early live evaluation, including clinical utility, safety, and human factors. [10] Its relevance here is the need to evaluate the combined human and technical workflow before conducting a larger effectiveness study. The system should be tested in shadow mode and then in a bounded assisted phase, with a reliable route back to existing operations.

11. Reproducible numerical illustrations

11.1 Purpose and assumptions

All values in this section are hypothetical author-selected inputs. They are not estimated from Kuwait patient records, vendor quotations, or intervention effect sizes. The companion calculation script reproduces the tables and figures using deterministic arithmetic. There is no stochastic simulation and no claim of forecast accuracy.

Assume one homogeneous service-equivalent pool with 10,000 usable monthly slots. At baseline, 80% produce completed consultations. Consider hypothetical completion rates of 82%, 85%, and 88% after an intervention. Assume 8,500 new equivalent episodes arrive monthly, an initial backlog of 12,000, and no changes in demand, case mix, transfers, withdrawal, or staffing. These simplifications are deliberately strong and unsuitable for direct budgeting.

Illustrative condition	Completed visits per month	Extra visits versus baseline	Monthly backlog change
Baseline, 80% completion	8,000	0	+500
Scenario A, 82% completion	8,200	200	+300
Scenario B, 85% completion	8,500	500	0
Scenario C, 88% completion	8,800	800	-300

Scenario A improves output while the backlog still grows. Scenario B stabilises the queue without clearing it. Scenario C reduces the simplified backlog by 300 episodes a month and would require 40 months to clear an initial 12,000, if every assumption held. This is a fluid accounting illustration, not a prediction of individual waiting times. In practice, priorities, follow-up demand, and non-fungible clinical resources make the calculation more complex.

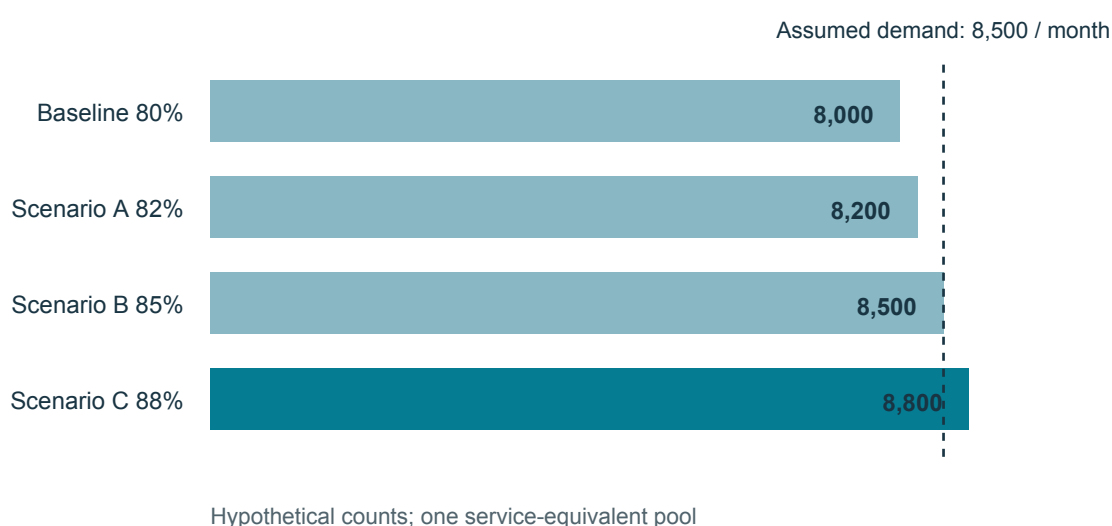


Figure 1. Hypothetical monthly completed capacity compared with 8,500 incoming equivalent episodes. The horizontal demand line separates backlog growth from possible reduction. Source: author calculations; not Kuwait observations.

11.2 Illustrative economic arithmetic

Assume setup expenditure of KWD 120,000 and annual recurring programme expenditure of KWD 60,000. For a simple three-year, zero-discount accounting illustration, annualised programme cost is KWD 100,000. This excludes additional clinical delivery costs and is not a cost-effectiveness result.

Illustrative condition	Additional visits per year	Annualised programme cost per additional visit	First-year cash programme cost per additional visit
Scenario A	2,400	KWD 41.67	KWD 75.00
Scenario B	6,000	KWD 16.67	KWD 30.00
Scenario C	9,600	KWD 10.42	KWD 18.75

The first-year cash numerator is KWD 180,000, including setup and one year of recurring expenditure. If no additional visits are produced, cost per additional visit is undefined, not zero. If output falls, report the higher cost and worse outcome directly rather than presenting a misleading favourable ratio. Additional investigations, consumables, clinical time, or capacity needed to deliver the extra visits must be included in a real evaluation.

11.3 Why near-full utilisation can be fragile

As a separate teaching example, assume a single-server M/M/1 queue with Poisson arrivals, independent exponential service times, first-come-first-served service, constant rates, no scheduled appointments, and a stable system. Let service capacity be four patients an hour. Under this model, mean queue waiting is $W_q = \lambda / [\mu \times (\mu - \lambda)]$.

Illustrative arrivals per hour	Utilisation	Mean queue wait under the model
3.0	75%	45 minutes
3.6	90%	135 minutes
3.8	95%	285 minutes

The nonlinear increase illustrates why the objective should not be 100% utilisation at all times. The numbers must not be applied to an actual Kuwait clinic: scheduled arrivals, multiple clinicians, prioritisation, late arrivals, and different service-time distributions violate the example's assumptions. A local discrete-event simulation could be developed later using validated distributions, but its predictions would still require prospective checking.

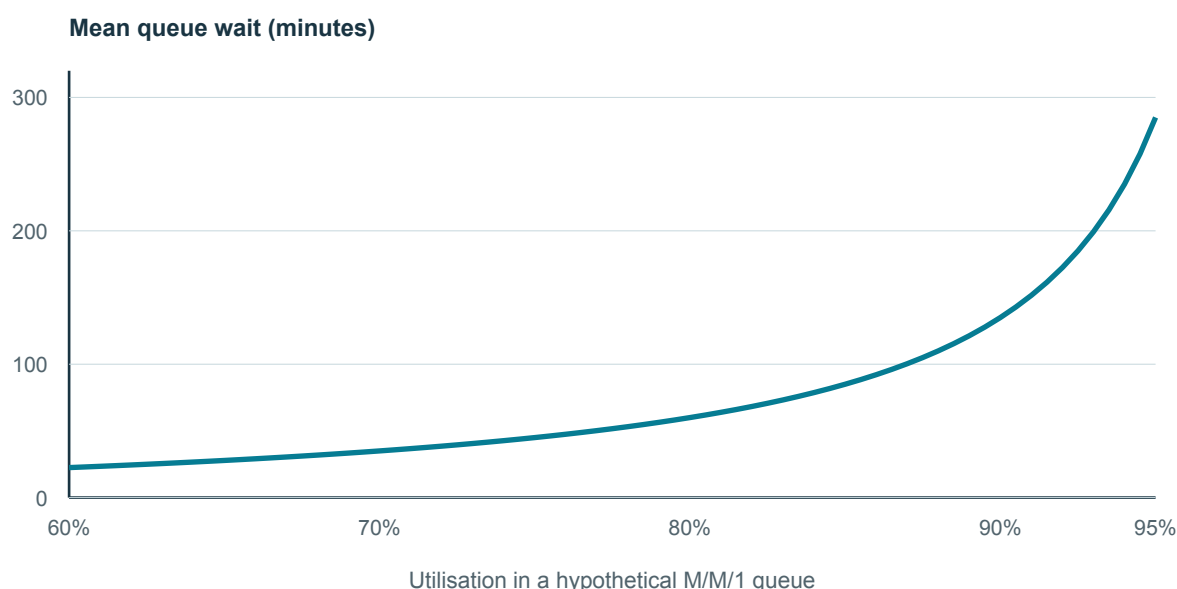


Figure 2. Mean queue waiting in a hypothetical M/M/1 system with service rate four per hour. This teaching figure illustrates sensitivity to utilisation; it does not model a participating healthcare facility.

12. Proposed prospective study

12.1 Study phases

Phase 0: governance and mapping. Confirm participating services, responsible clinicians, data access, ethics requirements, applicable policy, and patient involvement. Map the pathway and define the intervention. The report itself does not enrol sites or authorise data processing.

Phase 1: baseline and shadow evaluation. Aim to inspect up to 12 months of usable historical events to examine seasonality, supplemented by prospective observation and timestamp validation. If only a shorter reliable series exists, document the limitation and restrict seasonal conclusions. Run proposed forecasts or risk scores without changing care, and test the reliability of the data and proposed actions.

Phase 2: assisted feasibility. Introduce a bounded workflow with staff review at separate feasibility sites that will not become usual-care controls in Phase 3. Test identity matching, message comprehension, cancellation integrity, queue reconciliation, and downtime. The purpose is to identify failures and estimate implementation parameters, not to declare effectiveness from a short before-and-after comparison. If separate sites are unavailable, randomise before any workflow exposure and restrict implementation testing to assigned intervention sites under a prespecified design; an assumed washout does not undo organisational learning.

Phase 3: controlled effectiveness evaluation. Compare the defined operational package with usual care. If operationally feasible, randomise independent clinic or service clusters within strata defined by specialty and baseline access. Define clusters around shared staff and booking pools to limit contamination. The number of clusters must be determined from the baseline and power calculation rather than chosen for convenience.

Phase 4: incremental predictive evaluation. Compare the same operational package with and without predictive assistance. This can be a later cluster trial or, for selected non-interfering outreach components, a carefully designed individual randomisation. Do not claim a causal AI effect from Phase 3 if it compares a multi-component package with usual care.

12.2 Why a cluster design may be appropriate

Booking rules, staff training, and shared capacity affect multiple patients at once. Individual randomisation can contaminate the control group when recovered capacity benefits everyone in the same queue. A cluster design better matches such a shared intervention, but it requires enough independent clusters and an analysis that accounts for correlation within them.

If rollout must be staggered, a stepped-wedge design may be considered, but it is not automatically superior. Calendar trends, learning effects, uneven rollout, and changes in staffing can confound the contrast. A small number of clusters with sequential introduction may offer weak causal identification. A contemporaneous controlled interrupted time series may be more feasible, with limitations stated explicitly. Design selection should precede implementation and involve a trial statistician.

12.3 Primary outcome and estimand

For the incident cohort of eligible new referrals, the proposed primary outcome is restricted time without a completed first consultation over 90 days from original referral receipt. A patient completing the consultation on day 20 contributes 20 days; one whose noncompletion through day 90 is ascertained contributes 90. Unknown completion status is missing outcome information, handled under Section 12.7, rather than automatically assigned 90.

Eligibility should be assessed for all received referrals using common prespecified clinical rules independent of allocation, with adjudication blinded to study group where feasible. An incomplete but clinically eligible referral retains its original receipt date. Cases with unresolved eligibility should be reported separately, adjudicated where possible, and included in sensitivity bounds rather than silently omitted. Changes in eligibility or referral acceptance after assignment require independent review.

The primary estimand is a referral-average difference in mean restricted time under an intention-to-treat policy, regardless of uptake. For the primary cohort, use the first eligible referral per person within each participating service during recruitment; link repeat bookings to that episode. Separate legitimate referrals to different services may contribute, with sensitivity analysis restricted to one referral per person and accounting for repeated-person dependence where needed. Weight the analysis to target the defined referral population, not an inadvertently equal-weight average of clinics, and specify design-consistent weights before analysis.

The 90-day horizon is a proposed evaluation window, not a clinically acceptable waiting target. Report shorter, policy-relevant completion windows separately. For pathways requiring much earlier care, a different design and outcome horizon are necessary. Each referral must have 90 days of outcome ascertainment before the main analysis; recruitment should end early enough to permit this follow-up.

Death, withdrawal, transfer, or urgent escalation must not be silently treated as ordinary independent censoring. Death before consultation is a competing event and contributes 90 under this composite access-failure policy, with mortality reported separately. Transfer or escalation is not inherently absorbing: continue ascertainment and count a later consultation if it satisfies the prespecified qualifying-service definition. Where no qualifying consultation occurs through day 90 and that status is known, assign 90; where ascertainment is lost, apply the missing-data plan. A sensitivity analysis can include verified equivalent external consultations excluded by the primary definition. Report cumulative incidence with death as a competing event and separate safety outcomes. This definition measures completed access, not the clinical desirability of every route out of the queue.

12.4 Secondary and balancing outcomes

Secondary outcomes should include the proportion completing consultation within relevant clinical windows, time to clinical triage, number and age of unresolved referrals, rebooking count, late cancellation, non-arrival, recovered-slot completion, and follow-up timeliness. Patient-reported clarity, convenience, and burden should be measured using a short instrument tested for comprehension in the intended languages.

Balancing outcomes should include in-clinic delay, consultation duration, overtime, workload, duplicate bookings, inappropriate routing, delayed urgent escalation, and complaints. Record downstream diagnostic waits and avoidable repeat visits where attribution is feasible. A service should not be considered improved simply because the first encounter is faster if it creates additional incomplete encounters or defers necessary follow-up.

12.5 Primary analysis

Estimate the adjusted mean difference in the bounded 90-day outcome, with confidence intervals that account for the randomisation unit and weights consistent with the referral-average estimand. Prespecify adjustment for allocation strata and a small set of baseline prognostic variables. With few clusters, use a method appropriate to that limitation, such as appropriately weighted cluster-level analysis or randomisation-based inference; ordinary patient-level standard errors are invalid. Report an equal-clinic-weighted result as a distinct sensitivity estimand if useful, not as an interchangeable primary result.

Analyse patients according to the assigned intervention of their referral cluster and record cross-cluster transfers. Report the cluster and patient flow, missing outcome information, protocol changes, and actual delivery of each component. A per-protocol analysis may be secondary but cannot replace intention-to-treat when uptake is related to patient difficulty or staff workload.

For secondary time-to-event analyses, report the handling of competing events and censoring explicitly. For fixed-window completion, provide absolute risk differences as well as ratios where informative. Treat exploratory subgroup comparisons cautiously, using confidence intervals and interaction estimates rather than declaring benefit in one group and no benefit in another based solely on separate significance tests.

12.6 Power and sample size

A defensible sample-size calculation needs baseline outcome variance, intracluster correlation, cluster size distribution, baseline adjustment performance, expected missingness, and a clinically meaningful effect. None is established here for the intended sites. The report therefore does not assert a required number of patients or hospitals.

For illustration only, the common equal-size design-effect expression is $1 + (m - 1) \times \text{ICC}$. With $m = 200$ and $\text{ICC} = 0.02$, the factor is 4.98; with $\text{ICC} = 0.05$, it is 10.95. These assumed values show why thousands of patient records cannot compensate automatically for very few independent clinics. This expression is not a complete cluster-trial power calculation and does not address unequal sizes, covariate adjustment, or degrees of freedom.

The final calculation should use simulation or a validated method for the selected design, informed by the observed baseline and plausible effect sizes. Choose an effect meaningful to patients and operations, not simply the smallest detectable difference under an available dataset. Freeze the calculation and analysis plan before examining comparative trial outcomes.

12.7 Missingness, gaming, and sensitivity analysis

Investigate whether missing completion timestamps differ by intervention, site, or booking channel. Cross-check a sample against source records. Do not automatically impute an absent timestamp as either successful completion or non-attendance. If the main endpoint cannot be fully ascertained, report the missing fraction and conduct plausible best/worst-case bounds and model-based sensitivity analyses with assumptions stated.

Review referral rejection, transfer, and cancellation patterns for changes that could improve the metric without improving access. Retain a denominator audit and analyse the original eligible cohort. Predefine sensitivity analyses for verified external consultation, patient-requested deferral, major service disruptions, and different reasonable outcome windows.

12.8 Registration and reporting

Register a prospective interventional study in an appropriate public registry before recruitment, as applicable, and publish the protocol and statistical analysis plan. Use CONSORT 2025 for the randomised report and the relevant cluster extension, supplemented by CONSORT-AI where AI is part of the intervention. [12, 13] These are reporting frameworks, not substitutes for sound design or institutional approval. Early feasibility and prediction components should have their own clearly identified objectives and reports.

13. Economics, productivity, and growth

13.1 Three perspectives

The provider perspective includes setup, integration, licences, support, training, staff time, infrastructure, and additional clinical activity. The patient and caregiver perspective includes travel, time away from paid or unpaid work, communication costs, and repeated visits. A broader societal perspective combines relevant resource consequences while avoiding double-counting transfers or the same time saving twice.

Use a cost-consequence analysis initially: show costs next to access, safety, workload, and patient-burden outcomes. A cost per additional timely completed consultation may be useful if its definition is clear. A claim of cost-effectiveness in health terms would need a justified health outcome and appropriate modelling; a shorter wait alone does not supply quality-adjusted life years.

13.2 Measure resource use locally

Collect staff time through structured observation or validated logs rather than relying only on recall. Separate initial learning from steady-state work. Include the time spent reviewing recommendations, handling exceptions, correcting mistakes, maintaining contact details, and monitoring the system. Reduced clerical time may be redeployed to useful work without reducing the budget, which is a productivity gain but not a cash saving.

The WHO-partnered Kuwait costing report examines 84 selected public PHC outpatient services using 2019 activity. It provides a methodological and contextual reference, not a current price list for this intervention. [22] Local procurement and payroll data, under appropriate access controls, are required for the actual cost model. State the price year and distinguish accounting allocations from marginal resource costs.

13.3 Avoid counting the same benefit twice

If administrative time is redeployed to produce additional visits, reporting both its full monetised value and the full value of those same visits may double-count the benefit. Likewise, reduced appointment delay does not automatically equal recovered working hours. A patient might attend outside paid work, use annual leave, require a caregiver, or be unable to work for reasons unrelated to scheduling.

Report patient time in hours first, then value it in sensitivity analyses with transparent assumptions. Include unpaid care and avoid implicitly valuing access for higher-income patients as clinically more important. Eligibility for earlier care must never depend on estimated earnings or contribution to GDP.

13.4 Budget impact and uncertainty

Prepare separate first-year and steady-state budget statements. Include integration failure or delay, vendor price changes, higher-than-expected exception handling, and additional clinical demand generated by better access. Estimate the distribution of costs across participating institutions rather than assuming the organisation funding the software receives all benefits.

Use one-way and scenario sensitivity analyses for uptake, effect size, recurring cost, implementation delay, and clinical marginal cost. If probabilistic analysis is justified, use defensible parameter distributions and report uncertainty jointly with outcomes. CHEERS 2022 provides relevant reporting guidance. [16] A favourable illustrative ratio cannot replace measured resource use and a credible comparator.

13.5 The link to national productivity

Better healthcare access could support Kuwait's productivity through less time spent navigating services, fewer repeat visits, more reliable work attendance, and more effective use of clinical staff. Earlier appropriate treatment could also matter, but disease-specific clinical evidence would be required to quantify that pathway. These are plausible mechanisms for study, not measured national effects in this report.

The programme should establish a chain of evidence: intervention delivered, access improved, care appropriately completed, burden reduced, and resources used more effectively. Only after those links are measured should analysts consider broader economic modelling. Multiplying a hypothetical number of saved appointments by average wages and calling the result GDP growth would overstate what the data establish.

14. Safety, privacy, and accountability

14.1 Clinical responsibility

Assign a clinical owner for urgency rules, escalation, and inappropriate-routing review. Assign an operational owner for booking integrity and service continuity. Assign a data owner for access, retention, and quality. A vendor or model developer should not become the default clinical decision-maker because the interface presents a recommendation confidently.

WHO's AI ethics guidance emphasises autonomy, safety, transparency, accountability, inclusion, and sustainable responsiveness. [15] In this proposal, these principles are implemented through limited system authority, patient assistance routes, clear reasons for decisions, audited changes, and a practical ability to suspend or reverse automation.

14.2 Prespecified safety responses

Any suspected clinically significant delay caused by routing or prioritisation should trigger immediate clinical review. Confirmed wrong-patient booking or disclosure should trigger the institution's incident process and suspension of the implicated workflow until corrected. Repeated duplicate bookings or inability to reconcile schedules should revert the affected service to the authorised fallback process.

The final protocol should define severity, reporting routes, review times, and pause criteria with participating clinicians and governance teams. Numerical thresholds should be justified from baseline rates and risk tolerance;

this paper does not invent universally safe limits. Aggregate performance improvement must not overrule a serious preventable harm.

14.3 Patient choice and transparency

Explain the purpose of reminders and offers in understandable language, including how to request human assistance. Separate communication preferences from eligibility for care. Where research consent is required, refusal must not reduce ordinary access. Whether a waiver or service-evaluation route is appropriate requires an institutional determination, not an assumption by the technical team.

Patients should be able to correct contact information and report an unsuitable appointment or apparent mistake. The programme should distinguish a request to stop optional messages from a request to withdraw a referral. A model should not infer consent from engagement with an unrelated digital service.

14.4 Security and procurement

Require access logging, least-privilege accounts, encryption appropriate to the environment, tested backup and recovery, and a documented incident process. Evaluate whether any external service retains data, uses it for training, or routes it outside authorised infrastructure. Contract terms should cover data ownership, subprocessors, deletion, audit access, service continuity, and exportability.

Procurement acceptance should depend on a demonstrated workflow using approved test data: correct identity matching, reliable reservation, auditability, accessibility, and safe failure. A vendor demonstration or accuracy claim should not be treated as an independent effectiveness evaluation. Preserve the ability to export event data and change suppliers without losing the waiting-list history.

15. Implementation and scale decisions

15.1 A feasible sequence

An indicative sequence is four to six weeks for governance and mapping, six to eight weeks for baseline validation and integration design, four to six weeks for shadow and assisted testing, and at least several months of controlled operation plus outcome follow-up. These are planning assumptions, not promised delivery dates. Data access, ethics review, integration complexity, and the statistical design may lengthen the schedule.

Do not start the effectiveness clock while basic event recording is still changing. A service should enter the comparative phase only after staff can reconcile the referral queue, confirm booking state, identify exceptions, and operate the fallback process. Document implementation fidelity throughout the evaluation rather than classifying a site as fully implemented after training alone.

15.2 Decision gates

Gate	Evidence needed	Decision if unmet
Data readiness	Valid event meanings, acceptable linkage, explainable missingness, audited timestamps	Repair measurement before claiming effect
Workflow readiness	Staff ownership, accessible patient route, tested reservation and rollback	Continue bounded feasibility work
Clinical readiness	Approved urgency rules, escalation, review of failure cases	Do not automate affected decisions
Comparative benefit	Credible improvement in prespecified access outcomes with uncertainty reported	Refine, stop, or repeat a justified study
Acceptable consequences	No unacceptable safety, equity, workload, or downstream deterioration	Address harms before expansion
Sustainable operation	Real costs, support capacity, interoperable records, exit arrangements	Restrict scope or revise procurement

No single metric should decide scale. A small average wait reduction with higher staff exhaustion or a widening access gap may not justify expansion. Conversely, a modest effect in a difficult-to-reach group may be worthwhile even if overall utilisation changes little. Decisions should use the prespecified outcomes and values agreed with stakeholders.

15.3 From pilot to wider use

Replicate at services with different case mix and operating conditions before claiming national applicability. Preserve local clinical authority while standardising event definitions and reporting. Evaluate changes in demand induced by better access, since a successful service may attract referrals previously delayed or directed elsewhere.

Scale should include training, maintenance, data stewardship, and patient support, not only software licences. Publish null and negative results as well as positive findings. The research contribution becomes more valuable when it identifies where technology does not solve the binding constraint.

16. Limitations and unresolved questions

This report does not analyse patient-level Kuwaiti waiting data, validate a prediction model, estimate a causal intervention effect, or establish local implementation costs. Its evidence search is targeted rather than systematic. External studies differ in setting, intervention content, populations, outcomes, and age. Official service descriptions do not establish actual availability for every patient or institution.

The numerical examples use homogeneous capacity and constant demand, whereas real services contain interacting specialty queues, repeated visits, variable urgency, and constrained resources. The proposed study may be difficult to randomise if all candidate clinics share staff and booking infrastructure. Patient experience measures need linguistic and cultural testing. Some equity dimensions may be unavailable or inappropriate to collect.

The most important unanswered questions are empirical: how much waiting is avoidable within existing resources; which barriers explain missed care; whether released slots can actually be reused; whether administrative work is reduced or shifted; and whether gains persist beyond implementation. The study should be designed to answer those questions even if the answer is that additional clinical capacity or a different care pathway is required.

17. Conclusion

Kuwait offers a relevant setting for rigorous research on technology-supported access because the task can build on existing healthcare services, digital channels, and outpatient policy. The strongest initial intervention is a coherent operational system: reliable referral information, accessible communication, clinically governed allocation, realistic capacity planning, and audited completion.

AI may add value where a forecast or prediction changes an actionable decision. Its contribution should be measured against a simpler alternative and judged by timely completed care, safety, equity, workload, and cost. The proposed research programme is designed to make those distinctions visible. It supports a testable route from administrative improvement to health service productivity without presuming a national economic gain or an unobserved clinical result.

Research transparency statement

This document is a research report and proposed protocol prepared for RSB AI Research Lab. It is not peer reviewed and has not been submitted or registered as a clinical trial. No institution has been enrolled or represented as endorsing the proposal. No patient-level data were accessed. The only computed results are explicitly hypothetical deterministic illustrations.

AI assistance was used for literature discovery, drafting, calculation code, and document production. Sources were checked to the extent documented in the evidence register. Clinical, statistical, legal, and institutional review remains necessary before any implementation. Authorship contributions, funding, and potential competing interests must be confirmed by the named author before journal submission; this report does not assume declarations on the author's behalf.

Appendix A. Selected intervention evidence

This matrix summarises selected external findings. It does not pool effects or rank studies using a formal certainty framework. Differences in outcome definition and setting prevent direct application of these numbers to Kuwait.

Source and design	Reported finding	Limit on interpretation
Guroi-Urganci et al., 2013; review, SMS versus no reminder [5]	Seven studies with 5,841 participants; attendance risk ratio 1.14, 95% CI 1.03-1.26; moderate-quality evidence	Attendance outcome; not a demonstrated reduction in referral backlog
Hallsworth et al., 2015; two UK reminder-wording trials [6]	First trial: non-attendance 8.4% with specific-cost wording versus 11.1% with the existing message; OR 0.74, 95% CI 0.61-0.89	One setting; no AI; do not import the wording or monetary amount without local testing
Rose et al., 2011; advanced-access review [7]	All eight studies reporting third-next-available appointment time improved; only two reached less than 48 hours	Heterogeneous evidence; inconsistent findings on continuity and satisfaction
Shah et al., 2016; targeted-call RCT [8]	2,247 high-risk primary-care patients; non-attendance 22.8% versus 29.2%, difference -6.4 percentage points, 95% CI -9.8 to -3.0	Arrival, cancellation, and rescheduling proportions were not significantly different; fewer no-shows did not establish more arrivals
Percac-Lima et al., 2015; navigation RCT [14]	Among reported outcome denominators, non-attendance 167/1,631 (10.2%) versus 280/1,603 (17.5%)	4,425 appointments were identified as high risk; do not confuse that count with analysed patients; cancer setting differs from proposed initial scope

The matrix supports a staged test of communication, navigation, and scheduling. It also shows why the trial must retain several linked outcomes: classified non-attendance can change without an equivalent increase in completed visits, and completed visits can increase without reducing a backlog that continues to receive more demand than capacity.

Appendix B. Minimum data dictionary

The following is a proposed research schema, to be mapped to local systems. Fields should be collected only where necessary, authorised, and meaningful. Direct identifiers should remain in restricted operational systems.

Field or group	Definition	Validation or caution
person_key	Pseudonymous person identifier	Stable under authorised linkage; not a published patient identifier
episode_key	Identifier for one referral need	Do not merge separate legitimate clinical episodes
referral_received_at	First recorded receipt of the original referral	Retain through returns and rebooking; audit against source

Field or group	Definition	Validation or caution
referral_source	Referring service and route	Use a controlled vocabulary; distinguish missing from unknown
urgency and assigned_at	Clinician-assigned category and timestamp	Preserve reassessment history and responsible role
review_completed_at	Completion of required referral review	Do not substitute for original receipt
status_events	Time-stamped transitions and reasons	Append history rather than overwrite the previous state
appointment_key	Identifier of a specific booking	Link reschedules and replaced bookings to the episode
offered_at and offer_expiry	Time and response window for an offer	Record unsuccessful offers as well as acceptance
scheduled_start	Agreed appointment start	Preserve original and changed dates
arrival_at and ready_at	Physical arrival and defined readiness	Report registration delay separately
clinician_start and end	Start and end of the clinical encounter	Audit automatic defaults and batch entry
completed_at	Verified qualifying consultation completed	Define completion clinically, not merely a closed screen
cancellation_actor and reason	Patient, provider, system, or unknown; coded reason	Separate early/late cancellation and non-arrival
contact_channel and delivery	Permitted route, attempted contact, delivery state	Delivery does not establish reading or understanding
assistance_preferences	Language, accessible contact, caregiver support	Collect respectfully and only as needed
usable_session_capacity	Staffed, equipped, clinically suitable capacity	Distinguish from nominal template slots
staffing and disruption	Actual staff availability and material service changes	Necessary to interpret apparent intervention effects
model_version and action	Prediction version, recommendation, human decision	Retain override reason and timing
transfer and other exit	Destination or state, date, verification	Do not count transfer alone as completed consultation

For research extracts, replace precise dates or rare attributes where appropriate to disclosure control, while retaining enough information for valid analysis within the approved secure environment. Public datasets should be synthetic or sufficiently aggregated after formal review. The analysis code can be shared without sharing identifiable event histories.

Appendix C. Outcome specification and analysis checks

C.1 Core outcome definitions

Timely completion: numerator, eligible episodes with a verified appropriate consultation within the prespecified window; denominator, all eligible episodes in the incident cohort with adequate ascertainment. Report unknown outcomes separately and in sensitivity analysis. Use the clinically applicable window for each category, not a single lenient threshold for every referral.

Recovered-slot completion: count a slot only when it became available under a recorded cancellation or release event, was reallocated, and produced a completed consultation. Link the replacement to the original slot and avoid counting the same recovery after each rescheduling event. Report the notice interval, offer count, and staff effort.

Non-attendance: predefine the point at which a scheduled patient is classified as not arriving, account for cancellation rules, and distinguish provider cancellation. Report the denominator of scheduled appointments as well as unique patients and episodes. Do not mix appointment-level non-attendance with referral-level completion.

Backlog age: calculate elapsed time from original receipt for every unresolved eligible episode at a fixed census date. Report count, median, upper percentiles, and age bands. Backlog summaries are stock measures; incident-cohort completion is a flow measure. They answer different questions.

Workload: measure staff minutes by task and role, including exceptions and reconciliation. Report overtime and staff-reported burden alongside total task time. A lower average may conceal a small group handling substantially more difficult work.

C.2 Checks before releasing a result

Confirm that recruitment dates permit the required follow-up, that outcome definitions were fixed before comparative analysis, and that all clusters are accounted for. Reconcile referral, appointment, and encounter counts. Review excluded episodes and reasons with someone independent of the intervention team. Check whether missingness, rejection, or transfer changed after implementation.

Compare the main analysis with prespecified sensitivity analyses. Examine intervention fidelity, concurrent staffing changes, calendar effects, and contamination. Report confidence intervals and absolute effects, not only p values. Distinguish exploratory findings from primary results and avoid selecting the most favourable specialty or time period after viewing the data.

C.3 Worked patient-pathway examples

Example 1, administrative return: a referral arrives on day 0, is returned on day 3, corrected on day 8, and completed on day 24. Main access time is 24 days, not 16. The five-day correction interval is separately classified so the team can examine whether the service or referring process caused avoidable delay.

Example 2, earlier slot declined: a patient has a suitable appointment on day 30 and declines a day-12 offer because transport is unavailable. Preserve the day-30 booking. The declined offer is useful operational information but does not establish completed access or justify a penalty.

Example 3, no completion by horizon: an eligible referral remains unresolved at day 90. It contributes 90 to the bounded primary outcome and remains in the unresolved cohort. It is not removed because the clinic has no completed-visit record.

Example 4, transfer: a patient transfers on day 18 and has a verified equivalent consultation on day 27. The primary endpoint's treatment of that encounter must follow the prespecified qualifying-service rule; a sensitivity analysis can include equivalent verified external care. The transfer event alone is never counted as a completed consultation.

Appendix D. Operational failure register

Failure mode	Potential consequence	Proposed control and measurable signal
Wrong person matched	Incorrect appointment or privacy breach	Restricted identity checks, ambiguous-match review, incident count
Two confirmations for one slot	Patient travel wasted; staff overload	Atomic reservation, idempotent retries, duplicate audit
Help request read as cancellation	Loss of access	Explicit confirmation, reversible transaction, cancellation complaints
Risk score lowers priority	Disadvantaged patients wait longer	Prohibit risk-based deprioritisation; audit offers and waits by group
Unavailable resource ignored	Appointment cannot deliver care	Validate clinician, room, equipment, and preparation before offer
Automated summary invents detail	Misrouting or unsafe preparation	Source-linked draft, mandatory review, sampled factual error rate
Too many recovered bookings	Longer clinic-day queues and overtime	Workload limits, session monitoring, no initial automatic overbooking
Outcome timestamp is unreliable	False improvement	Source audit, missingness reporting, validation before trial
Silent model update	Evaluation no longer describes deployment	Version freeze and change log
System outage	Schedule inconsistency	Tested fallback, reconciliation owner, recovery audit
Faster first visit delays follow-up	Downstream harm or repeat demand	Track follow-up windows and diagnostic completion
Digital channel excludes patients	Unequal access	Assisted route, language testing, channel-specific outcomes

Controls should be tested through realistic, approved synthetic cases before live use. The register should remain open to newly observed failures rather than treating its initial contents as exhaustive.

Appendix E. Interview and observation guide

Patients and caregivers: ask how the appointment was obtained, which steps required help, whether the offered time was usable, how travel and work were arranged, whether messages were understood, what happened when a change was needed, and whether they felt pressured to accept a short-notice offer. Ask about time spent on calls and repeated visits as well as satisfaction.

Reception and medical records: observe how referrals are received, corrected, duplicated, and closed; how cancellations become visible; how conflicting booking states are resolved; and which information staff must repeatedly request. Identify unofficial spreadsheets or messaging practices without assuming that they are either unnecessary or safe.

Clinicians and nurses: establish appropriate urgency rules, preparation needs, variation in visit duration, reasons that visits cannot be completed, and the clinical consequences of delayed follow-up. Ask which recommended scheduling changes are infeasible and why.

Managers and technical teams: document capacity-setting authority, roster lead times, system interfaces, audit capabilities, downtime procedures, data retention, and support responsibilities. Determine who can change a booking rule and who can stop a malfunctioning process.

The qualitative analysis should preserve disagreement. A process experienced as efficient by management may create hidden work for reception or patients. Use a transparent coding approach and examine negative cases, including people who stopped engaging with the service. Recruitment through a smartphone app alone would miss some of the most relevant participants.

Appendix F. Reproducibility and next research steps

The companion package contains the editable manuscript, a calculation script, machine-readable illustrative inputs and outputs, an evidence register, and the PDF builder. The numerical script has no patient input and no external service dependency. All tables labelled illustrative should match its output. The document's publication date is not a trial-registration date.

Before an empirical study, replace illustrative parameters with a validated baseline; obtain institutional and ethics decisions; co-design the intervention; confirm outcome definitions; calculate the sample size; register the protocol; and independently review the analysis plan. Those steps should generate a new version that records what changed and why.

For journal submission, an author should verify every reference and declaration, select an appropriate article type, and adapt the length to the journal's requirements. The current document is deliberately a long-form research report. It must not be relabelled as an original effectiveness study without corresponding data and methods.

References

- [1] World Health Organization, Alliance for Health Policy and Systems Research. Kuwait: a primary health care case study in the context of the COVID-19 pandemic. 2024. Study period March-December 2020.
<https://ahpsr.who.int/publications/i/item/kuwait-a-primary-health-care-case-study-in-the-context-of-the-covid-19-pandemic>
- [2] Kuwait Ministry of Health. MOH standardized Outpatient Services policy. Policy A-ADM-11. Cover dates: effective 1/10/2023; revision 1/10/2025. Sections 8.3.1, 8.3.4-8.3.6, pp. 4-5.
<https://www.moh.gov.kw/TechnicalDepartment2025/MOHOPD.pdf>
- [3] Kuwait Government Online. Booking Hospitals and Specialized Clinics Appointments Service. Ministry of Health service description. Accessed 4 October 2026. <https://e.gov.kw/sites/kgoEnglish/Pages/eServices/MOH/HospitalClinicAppointments.aspx>
- [4] Kuwait Government Online. Salem App. Official application description. Accessed 4 October 2026.
<https://e.gov.kw/sites/kgoenglish/Pages/ApplicationPages/Application-Salem.aspx>
- [5] Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R, Car J. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database of Systematic Reviews. 2013;12:CD007458.
<https://doi.org/10.1002/14651858.CD007458.pub3>
- [6] Hallsworth M, Berry D, Sanders M, Sallis A, King D, Vlaev I, Darzi A. Stating Appointment Costs in SMS Reminders Reduces Missed Hospital Appointments: Findings from Two Randomised Controlled Trials. PLoS One. 2015;10(9):e0137306.
<https://doi.org/10.1371/journal.pone.0137306>
- [7] Rose KD, Ross JS, Horwitz LI. Advanced Access Scheduling Outcomes: A Systematic Review. Archives of Internal Medicine. 2011;171(13):1150-1159. <https://doi.org/10.1001/archinternmed.2011.168>
- [8] Shah SJ et al. Targeted Reminder Phone Calls to Patients at High Risk of No-Show for Primary Care Appointment: A Randomized Trial. Journal of General Internal Medicine. 2016;31:1460-1466. <https://doi.org/10.1007/s11606-016-3813-0>
- [9] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366:447-453. <https://doi.org/10.1126/science.aax2342>
- [10] Vasey B et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ. 2022;377:e070904. <https://doi.org/10.1136/bmj-2022-070904>
- [11] Collins GS et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. <https://doi.org/10.1136/bmj-2023-078378>
- [12] Hopewell S et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. BMJ. 2025;389:e081123. <https://doi.org/10.1136/bmj-2024-081123>
- [13] Liu X et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. BMJ. 2020;370:m3164. <https://doi.org/10.1136/bmj.m3164>
- [14] Percac-Lima S, Cronin PR, Ryan DP, Chabner BA, Daly EA, Kimball AB. Patient navigation based on predictive modeling decreases no-show rates in cancer care. Cancer. 2015;121(10):1662-1670. <https://doi.org/10.1002/cncr.29236>
- [15] World Health Organization. Ethics and governance of artificial intelligence for health. 2021. ISBN 9789240029200.
<https://www.who.int/publications/i/item/9789240029200>
- [16] Husereau D et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. BMJ. 2022;376:e067975. <https://doi.org/10.1136/bmj-2021-067975>
- [17] HL7 International. FHIR R4, version 4.0.1. Appointment and Slot resource definitions.
<https://hl7.org/fhir/R4/appointment.html> and <https://hl7.org/fhir/R4/slot.html>
- [18] Little JDC. A Proof for the Queuing Formula: $L = \lambda W$. Operations Research. 1961;9(3):383-387.
<https://doi.org/10.1287/opre.9.3.383>

- [19] Buabbas AJ, Alsaleh FM, Alsanafi M, Safar MA, Al-Sultan AT, Zaman TA. Patient Experiences With Outpatient Pharmacy Services in Hospitals Using Automated Pharmacy Systems: Cross-Sectional Study. *JMIR Medical Informatics*. 2026;14:e80963. <https://doi.org/10.2196/80963>
- [20] Luo L, Luo L, Zhang X, He X. Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. *BMC Health Services Research*. 2017;17:469. <https://doi.org/10.1186/s12913-017-2407-9>
- [21] Oikonomidi T, Norman G, McGarrigle L, Stokes J, van der Veer SN, Dowding D. Predictive model-based interventions to reduce outpatient no-shows: a rapid systematic review. *Journal of the American Medical Informatics Association*. 2023;30(3):559-569. <https://doi.org/10.1093/jamia/ocac242>
- [22] Elmusharaf K et al. The cost of health services delivered at primary care facilities in Kuwait. UNDP, WHO, UNIATF and Gulf Health Council. 2023. <https://uniatf.who.int/docs/librariesprovider22/default-document-library/kuwait-phc-costing-report.pdf>
- [23] Kuwait Central Statistical Bureau. Health statistics catalogue, including Annual Bulletin of Health Statistics 2024. Accessed 4 October 2026. Catalogue verified; numerical bulletin tables not analysed in this report. https://www.csb.gov.kw/Pages/Statistics_en?ID=59&ParentCatID=70

All online references were consulted during the research completed on 4 October 2026. Retrieval method and material limits are recorded in the companion evidence register. Citation of a source does not imply endorsement of this proposal by its authors or institution.