From Descriptive to Predictive Analytics in Healthcare Service Operations

A Strategic Analysis of AI-Driven Digital Twins for Emergency Department Capacity Management

July 2026

Opher Baron, (SiMLQ; Rotman School of Management, the University of Toronto; Hebrew University of Jerusalem)

With

Shany Azaria (Tel Aviv University)

Dmitry Krass (SiMLQ, Rotman School of Management, The University of Toronto)

Abstract. Many modern business processes operate under congestion: stochastic demand, shared and limited resources, complex routing, and tight service-level constraints. From healthcare and public services to logistics, financial operations, and large-scale customer support, these systems exhibit nonlinear behavior where small disruptions propagate quickly and performance deteriorates sharply. This paper examines how Business Process Management (BPM) can evolve to address congestion-driven environments by integrating process mining, queueing theory, simulation, and machine learning and artificial intelligence (ML&AI) into *digital twins*. We discuss how descriptive analytics must move beyond static dashboards toward data-driven process models that reveal how processes actually unfold in practice—capturing routing variability, rework loops, synchronization delays, and resource contention. Process mining provides structural visibility; queueing-aware modeling explains performance; ML&AI supports prediction under uncertainty. Building on these foundations, predictive and comparative analytics enable counterfactual "what-if" evaluation of staffing, routing, prioritization, and scheduling policies before implementation. Finally, prescriptive analytics embedded within real-time digital twins allow organizations to intervene proactively—anticipating congestion, reallocating capacity, and mitigating cascading delays. Drawing on industrial deployments through SiMLQ, We demonstrate how combining process intelligence with congestion-aware operational modeling transforms BPM from retrospective analysis into continuous, real-time decision support. The implications extend across service industries where variability and resource contention define performance, resilience, and competitiveness. Emergency departments are not failing because clinicians lack commitment; they are failing because the operating system that coordinates demand, staffing, beds, diagnostics, specialist consultation, and inpatient transfers remains largely reactive. Digital twins that are based on extensive descriptive analytics can transform fragmented event data into a safe, continuously updated environment for prediction, counterfactual evaluation, and prescriptive action. The managerial implication is blunt: modern businesses, such as hospitals that rely only on after-the-fact statistics will remain trapped in decision blind spots. Businesses that build validated digital twins would move capacity management upstream, detect failure before it occurs, and redirect scarce resources to the highest-leverage bottlenecks in real time.

Keywords: emergency department crowding, digital twin, process mining, queueing theory, machine learning and artificial Intelligence, simulation, patient flow, capacity management.

1. The Global Crisis in Emergency Medicine Example

Modern systems and processes with different clockspeeds, from cloud computing tofinancial operations, large-scale customer support, manufacturing and logistics, healthcare, and public services, such as court systems, can all be thought of as *queueing networks*. In such systems stochastic demand, routing and service times create complexity in providing high service level. This paper examines how Business Process Management (BPM) can evolve to address congestion-driven environments by integrating process mining, queueing theory, simulation, machine learning and Artificial Intelligence (ML&AI) to create digital twins. We illustrate how descriptive analytics must move beyond static dashboards toward data-driven process models that reveal how processes actually evolve in practice—capturing routing variability, rework loops, synchronization delays, and resource contention. We illustrate these concepts using access to emergency care as an example, because of its direct impact on quality (and length) of life.

Emergency departments (EDs) are the open wound of modern health systems all around the world. They absorb unscheduled demand, function as a diagnostic front door, compensate for primary-care shortages, and serve as a holding area when inpatient beds are unavailable. From the perspective of a systems architect, the modern ED is far more than a clinical gateway; it can be conceptualized as a high-variability, high-acuity, multi-class queueing network. These systems operate under stochastic demand, shared and limited resources, and rigid precedence constraints. When downstream nodes, i.e., inpatient wards, are full or slow, the resulting "blocking" propagates upstream, manifesting as waiting-room crowding, ambulance offload delays, and the dangerous normalization of "hallway medicine" in the ED.

Recent Canadian data reinforce the scale of the problem. The Canadian Institute for Health Information (CIHI, 2025a; CIHI, 2025b) reported that in 2024-2025 half of ED patients spent four hours or more in care, and one in ten spent more than fourteen hours in the ED. One in ten admitted patients waited over two full days (48 hours) in the ED (regardless of their initial acuity level). The same data describes a system in which approximately 12 percent of ED visits ended in admission, a small share of volume that dominates congestion because admitted patients require beds, transfers, and inpatient coordination rather than rapid discharge.

The human consequences of operational failure are no longer abstract. Public reporting from Alberta has described waiting-room and ED deaths amid overcrowding, lack of capacity, and long waits (Canadian healthcare Technology, 2026). These episodes are not merely tragic clinical outliers; they are warning signals from a service system that cannot reliably distinguish between ordinary demand variability and imminent collapse. When patients with chest pain, sepsis risk, or possible aortic catastrophe wait for hours before assessment, the root cause is not only resource scarcity but also the absence of effective management.

In Quebec, the 2026 transition of dashboard responsibilities to Santé Québec led to a controversial reduction in publicly visible performance indicators, see (CanHealth, 2026; Gouvernement du Québec, 2025). For BPM, this highlights a critical governance flaw: descriptive dashboards that merely report yesterday’s congestion expose failure without providing a mechanism for improvement. We argue for a paradigm shift: moving from reactive reporting to proactive, ML&AI driven digital twins that convert fragmented event data into a management science instrument for real-time anticipation and intervention.

We highlight two main themes: (i) the objective of descriptive analytics starting with process monitoring, often presented in dashboards, should encompass an improved understanding of the system. Such an understanding requires modeling, and (2) with ML&AI digital twins for queueing network process are relatively easy to build.

2. The Analytics Maturity Framework

Management analytics framework from Baron (2021): We extend the management analytics framework from Baron (2021) the context of waits for healthcare organizations. The same ideas hold for other applications (with or without queueing). We emphasize the usage of ML&AI in supporting beneficial usage of management analytics in practice. For completeness, We copy the following definitions from Baron (2021):

- *Descriptive analytics* describes past performance of existing systems.
- *Predictive analytics* predicts future performance of systems.
- *Comparative analytics* compares performance of systems under different interventions.
- *Prescriptive analytics* prescribes interventions to improve future performance of systems.

In short- prescriptive analytics is the holy grail of analytics. Table 1 lists examples of managerial questions and decisions made based upon different management analytics components. The four-sphere framework in Figure 1 extends Figure 1 of Baron (2021); it separates analytics by time horizon and system state. This framework makes it explicit why descriptive dashboards, predictive models, causal comparisons, and optimization engines are complements rather than substitutes.

Sphere	Time / system state	Example	Typical ED decision
Descriptive	Past / existing	System monitoring: How long did patients wait?	Report Performance indicators.
Predictive	Future / existing	System operation: How long would future arriving in the future wait?	Forecast crowding and boarding risk for the next two to six hours.
Comparative	Past or future / alternative	System planning and what if analysis: What would change under alternative configurations?	Evaluate another nurse, faster consult request, or altered triage routing.
Prescriptive	Future / alternative	System improvement: What should we do now and in the future?	Select staffing, escalation, and patient-flow policies to optimize performance.

Table 1: Management Analytics- Spheres, Time, System, and Typical Decisions.

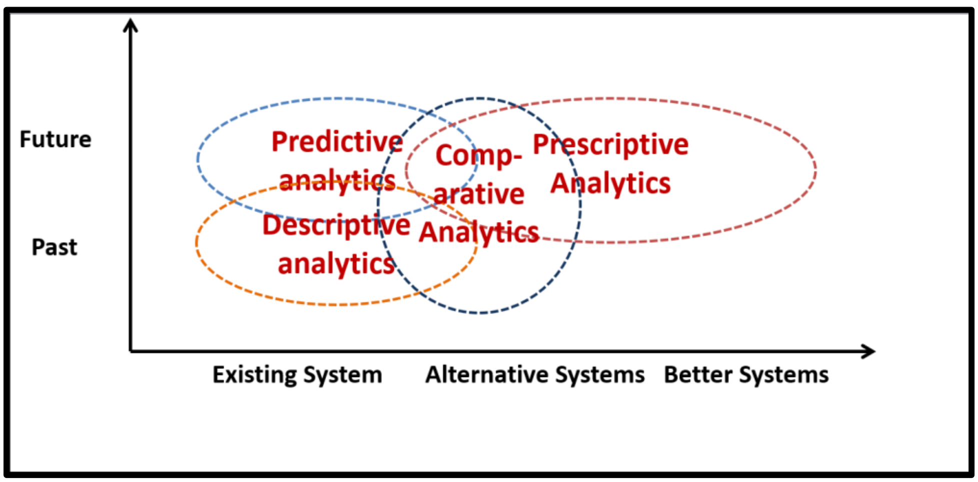

Figure 1. Management Analytics

The 5 faces of a Process: To better conceptualize how descriptive analytics can help understanding the system, consider the 5 faces of a process depicted in Figure 2: the process as (i) designed, by the process manager and its administrator, (ii) in reality, (iii) as seen by experts, such as employees and customers (thus there are many subfaces here) (iv) as captured in data, and (v) in its optimal, most effective, version. Obviously, these faces differ from each other. We think of management analytics as *using data to bring the real process towards the optimal one*.

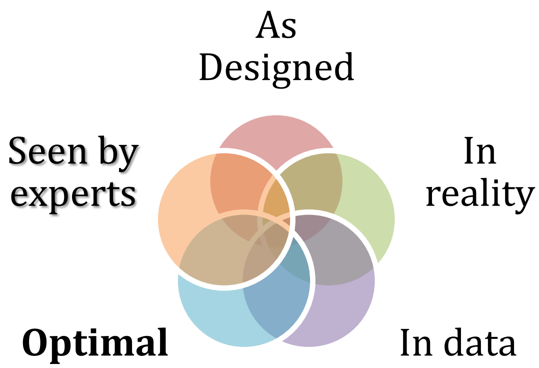

Figure 2. The five faces of a process.

System Understanding: We define a perfect understanding of a system as equivalent to being able to create *a digital simulator that responds as the system does*, i.e., can create both, the same data as is generated by the existing system, and the same data as would be generated by the real system after interventions. Having access to such a simulatior in real time implies perfect comparative analytics capabilities. Then, all is left for prescriptive analytics, is searching among feasible interventions for the optimal ones. Such optimization is still challenging and time consuming, and optimization in a data driven context often requires additional methods. An important part of this optimization is accurate and speedy performance evaluation, i.e., comparative analytics.

The focus of this paper is how to use ML&AI in descriptive analytics to support prescriptive one. While descriptive analytics is often understood as dashboards (no modeling), this view is too narrow. To identify interconnections within the "as is" system, modeling is often essential. The goal of descriptive analytics is to *understand* how the system operates, not just to describe its performance. This is where ML&AI can help- using descriptive analytics to understand, i.e., move towards comparative and prescriptive analytics.

3. How ML&AI Support Predictive, Comparative, and Prescriptive Analytics

We review the main usage of ML&AI in support for predictive, comparative, and prescriptive analytics today.

ML&AI in Predictive Analytics: Cheap, Accurate, and Timely. ML&AI tools have strategic importance to predictive analytics as they reduce the cost of prediction, improve its accuracy, and especially as more data is available, reduce the time to prediction. Agrawal, Gans, and Goldfarb (2018) describe AI as prediction technology: it makes prediction cheaper, which changes the economics of decisions that depend on uncertainty.

In ED operations, cheaper prediction enables a shift from periodic (monthly or weekly) review to intra-day control. The hospital does not merely ask whether last period's performance target was missed; it asks "which patients are at risk now?," "what bottleneck will bind in three hours?".

As ML&AI can quickly provide an accurate answer to this and similar questions, managers now need answers to questions on interventions and their impact, such as "which action is most likely to prevent unsafe crowding in three hours?" This extra time allows managers to plan and optimize their response to system disruptions before standard of care is impacted.

ML&AI in Comparative Analytics: Challenges and Opportunities: Comparative analytics asks what would have happened under a credible alternative. This distinction is the difference between a performance report and a management science instrument, such as BPM. ED leaders do not merely need to know that the time to Provider Initial Assessment (PIA) target was missed; they need to know whether another nurse, a changed triage policy, faster consultation, or earlier bed-management escalation would have improved the outcome.

In general, comparative analytics has two forms. Demonstrating these in the context of an ED, we think of *Retrospective* comparative analytics: How long would the patient who arrived at 11:00 a.m. this morning have waited if another nurse had been on shift or if a consult had been requested 45 minutes earlier? And *Prospective* comparative analytics: How long a patient that will arrive at 11:00 a.m., tomorrow, wait if the current admitted-patient queue remains unresolved, or if an additional physician is redeployed from fast track to acute assessment?

The causal logic draws on a broader empirical tradition. The 2021 Nobel Prize in Economic Sciences recognized David Card, Joshua Angrist, and Guido Imbens. The first two for their empirical and methodological contributions that advanced the analysis of causal relationships, and the last one for contributions to empirical macroeconomics particularly through natural experiments (Nobel Prize, 2021). The 2011 prize recognized Thomas Sargent and Christopher Sims for empirical research on cause and effect in the macroeconomy (Nobel Prize, 2011). These examples matter for healthcare operations because many operational changes cannot ethically or practically be randomized in real time and thus causal analysis has to be used.

Data-driven analysis faces the challenge of causality since it needs to provide performance evaluation of different systems than the observed ones. A/B testing is the gold standard for comparative analytics because it directly measures the effect of an intervention under controlled conditions. In its simplest form, patients, clinicians, shifts, units, or sites would be randomly split between a “control” condition that preserves the current process and a “treatment” condition that introduces a change, such as a new triage protocol, staffing rule, routing policy, or discharge process. The random assignment isolates the causal effect of the intervention from background variation in demand, case mix, staffing, and seasonal conditions. For this reason, A/B testing is widely used in digital services, clinical trials, and operational improvement programs as the clearest way to answer comparative questions.

In EDs, direct A/B testing is often impossible or is severely constrained by ethics, patient-safety concerns, operational disruptions the complexity of isolating a certain system within a highly interconnected network. Moreover, hospitals should not learn by experimenting blindly on live patients. This is where ML&AI can help.

In the next Sections we discuss how ML&AI can support descriptive analytics in building a digital twin of a complex queueing network system. Digital twins extend the logic of A/B testing. Rather than experimenting on real patients, the organization can simulate the changes inside a calibrated model of the ED, comparing the current pathway against alternative ones before implementation. Thus, A/B testing remains the conceptual benchmark for comparative analytics, while ML&AI-driven digital twins provide a safer, faster, and more scalable way to conduct “what-if” experimentation in complex queueing networks.

ML&AI in Prescriptive Analytics: Point Vs System Solution: Current applications of ML and AI in prescriptive analytics are still dominated by automation of existing processes. Quoting Gates, Myhrvold, and Rinearson (1996): *"The first rule of any technology used in a business is that automation applied to an efficient operation will magnify the efficiency. The second is that automation applied to an inefficient operation will magnify the inefficiency."* Thus, one needs to be careful with the processes that are automated. Moreover, such implementations are point solutions rather than redesigning the system itself. The trick is to use ML&AI to support better processes that may not have been feasible without this technology.

In many hospitals, ML&AI is deployed as a point solution: a model predicts patient deterioration, flags imaging results, forecasts arrivals, or even recommends a scheduling adjustment within a narrow workflow. Such point solutions can be valuable, but they often optimize a local decision without changing the system architecture.

The distinction between point solution and a system one is central to the argument in Agrawal, Gans, and Goldfarb (2022): the deeper economic value of AI emerges not merely when prediction is inserted into current decisions, but when organizations *redesign* decision systems around better prediction capabilities.

In ED operations, moving beyond isolated automation toward system-level prescriptive analytics can be done by integrating prediction, simulation, queueing, and operational constraints into a digital twin that can evaluate the consequences of decisions on the performance indicators of the ED.

4. Descriptive Analytics for Understanding the Systems -Using Process Mining and Queueing Theory

ML&AI provide several tools that can help elevate descriptive analytics from reporting to understanding. Process mining and queueing theory bridge the gap between recorded reality, including performance indicators, and the operations model.

Descriptive Analytics Deep Dive: PIs That Matter: An important way to test the digital twin is to compare its performance, in- and out-of-sample with performance observed in practice. A digital twin raises the standard for the descriptive metrics- as it intends to capture the entire system's behaviour- not just "its mean performance." The ED must define the right operational variables, maintain consistent timestamp logic, and segment performance by clinically meaningful groups. Table 2 summarizes the core metrics used in many EDs around the world.

Performance Indicator	Operational meaning	Significance?
Length of Stay (LOS)	Total time from ED arrival to departure from the ED.	Measures total system burden and validates patient-pathway simulations.
Left without Being Seen (LWBS)	Patients who depart before being seen by an appropriate clinician.	Signals front-door failure and potential safety risk, especially when high-acuity patients leave.
Time to Provider Initial Assessment (PIA)	Time from arrival or triage to physician initial assessment.	Captures the first major bottleneck and often predicts LWBS risk.
Time to Disposition (T2D)	Time to disposition decision: admit, discharge, transfer, or observation.	Separates diagnostic/clinical decision delay from bed-exit delay.
Time to Inpatient Bed (TIPB)	Time from admission decision to inpatient-bed transfer.	Measures back-door blocking and ED boarding pressure.
Time to Consult	Time from request for consult (often initial assessment) to arrival of the specialist.	Identifies coordination delays that can dominate admitted-patient LOS.

Table 2: Performance Indicators in Emergency Department

LOS is the most visible metric but not always the most actionable. A high LOS can arise from pre-triage waiting, diagnostic, consult, and disposition delay, or inpatient-bed boarding. Treating LOS as a single number is like treating a fever without diagnosing the infection. The digital twin can decompose LOS by pathway, daypart, and operational state. Most EDs measure LOS separately for admitted, non-admitted high acuity, and non-admitted low acuity patients. Similar distinctions exist for other performance measures recorded in Table 2.

LWBS rate is a key access quality measure. In standard queueing models, customers' balking or abandonment reflects low service valuation, and thus improve the efficiency of the system, see e.g., Hassin and Haviv (2003). In contrast, as diagnosis is a core ED service patients do not know their "valuation" ex ante since their true physical state can only be evaluated via assessment by the provider. Baker, Stevens and Brook (1991) shows that within a week after LWBS, up to 11% of patients require hospitalization and nearly half of LWBS patients needed the ED services. Thus, LWBS signals a failure of the ED.

PIA is the front-door bottleneck. It is also a managerial lever because it can often be improved without adding beds: rapid assessment zones, physician-in-triage models, nurse-initiated protocols, or dynamic provider assignment can reduce PIA. The digital twin can test whether those changes improve total flow or merely move congestion downstream.

T2D is an important metric as it captures the efficacy of patient release procedures. For example, a high difference between LOS and T2D typically indicates issues of bed blocking within the hospital (for admitted patients) and alternative level of care (for both admitted and non-admitted patients). Thus, this measure highlights that the ED system is just a subsystem of the overall healthcare delivery system.

TIPB is the back-door bottleneck. Boarding admitted patients consumes ED capacity and reduces the department's ability to absorb incoming unscheduled demand. In queueing terms, boarding creates blocking: downstream inpatient capacity constraints propagate upstream to ambulances, waiting rooms, and clinical staff workload. The fact that admitted patients represent a minority of visits but a large share of congestion is precisely why pathway-specific modeling is essential. Moreover, adequate capturing of the impact of bed blocking within the hospital can support change management of the procedures taken outside the ED.

Time to Consult further highlight that the ED is not an independent system. In most modern EDs, admitted patients require a consult from the admitted ward to sign off on admission to the ward. Similarly, kids seen in an ED often require a pediatrician to see them even before being released home. Long consult wait times add much congestion to many EDs.

Given the high variability in EDs and that long waits may deteriorate patients' health, it is common to look at the 90th percentiles of these and other (e.g., ambulance offload time) performance measures. Still, their averages are often also very informative. Moreover, looking at these and other PIs by hour of the day, acuity and patient pathways, paves the way for a better understanding the inter-dependencie in the system.

When completing descriptive analytics to understand the system- it is important to consider the relevant performance measures and learn how these are balanced by the system. For example, it is intuitive that LWBS and PIA are correlated- if patients do not wait much before initial assessment- they have little chance to be LWBS. With this in mind- shorter PIA will increase the throughput of served patients.

In sum the top 3 performance indicators are mostly internal to the ED and its procedures and the bottom 3 ones remind that the ED is a part of a much better system. A digital twin for the entire healthcare delivery ecosystem is the next step in improving access to health.

Process Mining: Process mining extracts knowledge from event logs generated by information systems and uses that knowledge to discover, monitor, and understand the actual processes. The minimum event-log structure required for analyzing queueing networks is simple: "who-" a case identifier, "what-" an activity label, and "when-" a timestamp. While time stamps could include arrival, start of service, and end of service for each activity, in many cases only arrivals or departures from stations are recorded. In healthcare, the case is usually the patient encounter; activities include registration, triage, physician assessment, lab order, imaging, consult request, and disposition decision; timestamps establish the observed sequence. (van der Aalst, 2016; Munoz-Gama et al., 2022; De Roock & Martin, 2022). Moreover, often there are many other relevant data, such as resources, i.e., beds, nurses, and doctors, and their availability; patients' characteristics, i.e., age, gender, and severity. The extra data supports better results of the ML&AI models, e.g., in understanding which patients follow specific pathways.

The directly-follows graph is the first operational translation of raw event data. It easily depicts if triage is followed by physician initial assessment in 65 percent of traces, and by waiting-room reassessment, lab-before-physician, or LWBS in the remainder. The difference between both pathways is not merely visualization-every discovered edge and node are a candidate delay, rework loop, bypass, or handoff failure.

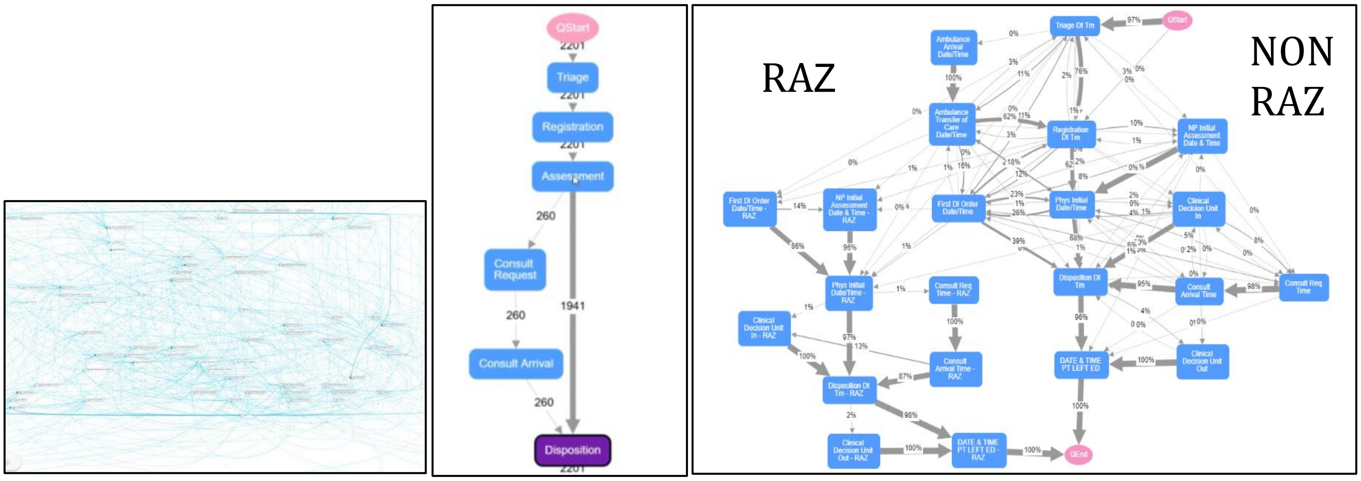

Figure 3: Left: a spagetti model, Center: a very simplicistic ED process view, and Right: an informative ED process view with a rapid assesment zone and a main (non RAZ) zone.

However, as depicted in Figure 3, left, unfiltered process maps often look as a "spaghetti models": dense diagrams that faithfully represent complexity but fail to support managerial action. Too simplistic models, see Figure 3, center, are not helpful either. The task is to filter without falsifying. High-frequency pathways, high-risk exceptions, and high-delay segments must remain visible, while rare noise and data-entry artifacts must not dominate the model. Figure 3, right, shows a useful simplified view of an ED that can support modeling.

Process mapping is part of the discovery process that reconstructs the real ED process from raw event logs. This process answers: What actually happens to patients, by acuity, mode of arrival, pathway, time of day, and disposition? This step often reveals that official protocols explain only part of the variation. Patients may move through diagnostics before physician assessment, wait for specialist consultation in undocumented holding states, or experience multiple cycles of order, wait, and reassessment.

The discovery process can expose differences between the designed and recorded workflows. Such conformance check compares actual flow with expected protocols. In the ED, conformance is not a rigid compliance exercise; many deviations are clinically justified. The value is to separate justified variation from preventable operational drift. For example, high-acuity patients should not normally experience long PIAs unless there is a clear clinical rationale or unavoidable capacity constraint.

Queueing Theory: Just depicting the process and supporting conformance checks delivers less than what descriptive analytics should. Queueing theory supplements process mining with the structural logic that captures the dependency of the system performance on its real time state. An ED is usefully conceptualized as a network of queues with stochastic arrivals, different patients' types, variable service times, finite servers, finite physical space, and downstream blocking. The same average arrival rate can produce radically different congestion depending on patient mix, variability, prioritization, pooling, and resource synchronization. Therefore, static staffing ratios routinely fail: they ignore the nonlinear relationship between utilization and delay. As utilization approaches capacity, waiting time increases exponentially, even before the nominal capacity limit is reached.

Queue mining that combines process mining and queueing theory allows us to gain much better understanding of the process by investigating the existing records. Moreover, with such an understanding of the process flow and internal dependencies, the analytics professional can start building a discrete-event simulation of the queueing network. Predictive analytics using ML&AI can calibrate arrival rates, service-time distributions, transition probabilities, and risk stratification. Queueing structure prevents the model from becoming a black-box curve fit, i.e., it maintains explainability. More importantly, changing the building blocks, such as the number of servers in a station, in the explainable simulation, i.e., provide comparative analytics, is straightforward. Together, these methods form a defensible digital twin: data-driven enough to reflect reality, structured enough to simulate interventions, and transparent enough to support leadership decisions.

Note that a digital twin depends on this discipline because bad process discovery produces bad simulation logic. Thus, an important part of the process to create a digital twin is taking advantage of its explainability and support discussion with business domain experts, so that the resulting model is not missing adequately representing the system.

5. The SiMLQ Solution: Automated Digital Twins

SiMLQ (www.simlq.com) is a tool that combines simulation, ML&AI, and queueing solution for automated digital-twin generation. Traditional simulation projects often require months of manual process mapping, distribution fitting, stakeholder interviews, model coding, and validation. Automation changes the economics of simulation. Once a simulation model can be generated and refreshed orders of magnitude faster than traditional simulation models, it becomes a digital twin, a live operating asset, rather than a one-off improvement project.

The live-digital twin loop depicted in Figure 4 has seven stages: data collection, creation of event logs, process map/dashboard (descriptive analytics), simulation and ML&AI (predictive analytics), what-if analysis (comparative analytics), and action selection, and action implementation (prescriptive analytics). The key is that the implemented actions impact the data collected, and thus feeds back to this 7-stage loop.

In contrast to traditional modeling, see Pidd (1999), where the model is an exogenous representation of reality, the management analytics approach that leads to a digital twin model requires it to continuously interacts with reality and data. Thus, reality, data, and model impact each other in real time.

The benefits from live digital twin simulation models are tremendous. In contrast to traditional simulation models that are expensive and thus used to support big projects, the marginal cost of a digital twin is negligible. This low cost implies that a digital twin can support operational decisions. Common examples in ED would be when a patient needs to undergo both diagnostic imaging and lab work- the digital twin can direct this patient in a fashion that reduces the overall waits.

The design principle is closed-loop control. A conventional dashboard ends when it displays a metric. A digital twin begins there. It asks what the metric implies, whether the state is likely to deteriorate, what alternatives exist, and which action should be taken next. Thus, the digital twin should be viewed as an operating model rather than an analytics report.

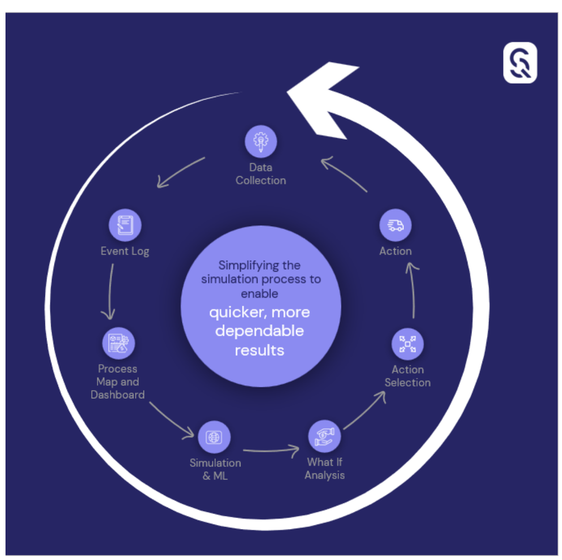

Figure 4. Closed-loop digital twin operating model.

Recent healthcare digital-twin literature supports this direction. Reviews emphasize that healthcare digital twins combine monitoring, simulation, prediction, and optimization, while emergency-care-specific work highlights potential applications in capacity management, triage, and operational decision support. Simulation-based digital twins have also been used to evaluate emergency medical communication-center scenarios and broader healthcare workflow designs. (Olawade et al., 2025; Li et al., 2025; Tang et al., 2024; Rojas et al., 2025).

6. Conclusion

Implementing digital twins using the four components of management analytics is not without difficulties. Moreover, our discussion here hints at further opportunities to use digital twins generated with modern ML&AI as a system solution for EDs and mother queuing network applications.

Risks, Limits, and Ethical Boundaries: The strongest argument for digital twins should not obscure their risks. The first risk is false precision. A digital twin can produce exact-looking numbers that depend on imperfect data, unstable demand, or undocumented clinical judgment. The solution is not to avoid modeling but to express uncertainty, validate forecasts, and communicate scenario ranges where appropriate.

The second risk is automation bias. Managers should not simply act in accordance with a recommendation just because it comes from ML&AI. A manger taking all decisions in accordance with “the algorithm,” manages no more. Governance should require that recommendations be explainable: which bottleneck is binding, what intervention is proposed, what trade-off is expected, and what safety constraints apply.

The third risk is misaligned incentives. If a hospital optimizes only average LOS, it may neglect tail risk, high-acuity waiting, staff burnout, or equity. E.g., if physicians are focused solely on throughput maximization during their shifts, patients’ waits may suffer. The objective must therefore include safety and fairness constraints. The goal is not to make the ED look good on paper; it is to improve real access and outcomes.

The fourth risk is privacy and cybersecurity. Digital twins depend on sensitive operational and clinical data. Implementation must follow applicable privacy law, role-based access, de-identification where feasible, audit logs, secure data pipelines, and clear data-retention rules. A hospital cannot build operational intelligence by weakening patient trust.

Toward the Optimal Healthcare System: The ED crisis is a service-operations failure with critical societal consequences. It is driven by stochastic demand, capacityand resource constraints, downstream boarding, diagnostic and consult dependencies. Currently decision systems often intervene only after the system deteriorates. The solution is not merely another dashboard. It is a progression from descriptive analytics to predictive, comparative, and prescriptive decision support.

Good descriptive analytics remains the foundation and should be pursued with modeling to support the next steps in the analytics journey. The ED needs a living model that can forecast crowding, test alternatives, and recommend action before patients experience avoidable harm. Many other queueing network systems can benefit from similar digital twin technology.

SiMLQ leads this strategic direction: a digital twin generated from operational data, calibrated by ML&AI, and activated through simulation. The premise of SiMLQ is not simply faster analysis (point solution) but a system change of the mode of ED management. In this mode leaders can see the system, anticipate failure, test interventions safely, and move toward the optimal hospital one operational decision at a time.

The transition from retrospective reporting to prescriptive digital twins marks a fundamental evolution in healthcare operations. By moving capacity management upstream, we transition from observing yesterday’s PIs to anticipating and preventing tomorrow's failures.

The ultimate vision is an ecosystem of interconnected digital twins—linking the ED, the hospital-wide bed management system, and even the patient's digital twin. Such a network would synchronize clinical and operational capabilities across the entire continuum of care, including long-term care and primary physicians. By integrating these systems, we seej to transform the healthcare system from a collection of disconnected silos into a synchronized, anticipatory network capable of delivering optimal care in real time.

References

1. Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction machines: the simple economics of artificial intelligence. Harvard Business Review Press.
2. Agrawal, A., Gans, J., & Goldfarb, A. (2022). Power and Prediction: The disruptive economics of artificial intelligence. Harvard Business Review Press.
3. Baker DW, Stevens CD, Brook RH (1991) Patients who leave a public hospital emergency department without being seen by a physician. causes and consequences. Journal of the American Medical Association 266(8):1085–1090.
4. Baron, O. (2021) Business Analytics in Service Operations—Lessons From Healthcare Operations. *Naval research Logistics*, Volume 68, Number 5, August, 1-17.
5. Canadian Institute for Health Information. (2025a). NACRS emergency department visits and lengths of stay, 2024-2025. https://www.cihi.ca/en/nacrs-emergency-department-visits-and-lengths-of-stay (Accessed, July 3, 2026)
6. Canadian Institute for Health Information. (2025b). Emergency department wait times in Canada: Longer ED stays reflect growing patient complexity and delays in admission. https://www.cihi.ca/en/emergency-department-wait-times-in-canada-insights-from-a-health-system-perspective/longer-ed-stays-reflect-growing-patient-complexity-and-delays-in-admission (Accessed, July 3, 2026)
7. Canadian healthcare Technology (2026) Doctors count 6 preventable deaths in Alberta EDs | Canadian Healthcare Technology, Jan 21, 2026.
8. CanHealth. (2026). Quebec drops many ER stats from public dashboard.
9. De Roock, E., & Martin, N. (2022). Process mining in healthcare: An updated perspective on the state of the art. Journal of Biomedical Informatics, 127, 103995.
10. Gates, B., Myhrvold, N., & Rinearson, P. (1996). *The road ahead* (Revised ed.). Penguin Books.
11. Gouvernement du Québec. (2025). Data on the Québec health system and its services.
12. Hassin R, Haviv M (2003) To queue or not to queue: Equilibrium behavior in queueing systems. International Series in Operations Research & Management Science (Springer Science).
13. Kuruppu Appuhamilage GDK, Hussain M, Zaman M, Ali Khan W. A health digital twin framework for discrete event simulation based optimised critical care workflows. NPJ Digit Med. 2025 Jun 19;8(1):376. doi: 10.1038/s41746-025-01738-4. PMID: 40537503; PMCID: PMC12179303.
14. Li H, Zhang J, Zhang N, Zhu B. (2025) Advancing Emergency Care With Digital Twins. JMIR Aging. Apr 2025 21;8:e71777. doi: 10.2196/71777. PMID: 40258270; PMCID: PMC12053090.
15. Munoz-Gama, J., Martin, N., Fernandez-Llatas, C., Johnson, O. A., Sepulveda, M., Helm, E., Galvez-Yanjari, V., Rojas, E., Martinez-Millana, A., Aloini, D., Amantea, I. A., Andrews, R., Arias, M., Beerepoot, I., Benevento, E., Burattin, A., Capurro, D., Carmona, J., Comuzzi, M., … Zerbato, F. (2022). Process mining for healthcare: Characteristics and challenges. Journal of Biomedical Informatics, 127, 103994.
16. Nobel Prize. (2011). The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2011.
17. Nobel Prize. (2021). The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2021.

18. Olawade, D. B., Ighodaro, O., Erhieyovwe, E. O., Hankamo, N. E., Hamza, I. T., & Analikwu, C. C. (2025). The role of digital twin technology in modern emergency care. International Journal of Medical Informatics, 106229.
19. Penverne Y, Martinez C, Cellier N, Pehlivan C, Jenvrin J, Savary D, Debierre V, Deciron F, Bichri A, Lebastard Q, Montassier E, Leclere B, Fontanili F. A simulation based digital twin approach to assessing the organization of response to emergency calls. NPJ Digit Med. 2024 Dec 31;7(1):385. doi: 10.1038/s41746-024-01392-2. PMID: 39741218; PMCID: PMC11688439.
20. Pidd, M. (1999). Just Modeling through: A Rough Guide to Modeling. *Interfaces*, *29*(2), 118–132. http://www.jstor.org/stable/25062474
21. Rabbi, F., Banik, D., Ibne Hossain, N. U., & Sokolov, A. (2024). Using process mining algorithms for process improvement in healthcare. Healthcare Analytics, 100305.
22.
23.
24. van der Aalst, W. M. P. (2016). Process mining: Data science in action (2nd ed.). Springer.