

ORIGINAL RESEARCH

Open access

Missingness as Signal: A Representation Theory for Sparse and Irregular Longitudinal Health Records

Ahmed Benali^{1*}, Karim Boudiaf², Samir Touati¹

Abstract

In the realm of healthcare analytics, sparse and irregular longitudinal health records pose significant challenges to traditional representation models, often treating missing data as mere artifacts to be imputed or discarded. This conceptual manuscript proposes a paradigm shift by framing missingness itself as an informative signal within a representation theory tailored for electronic health records (EHRs). We introduce the irregular signal encoding architecture (ISEA), a theoretical framework that integrates missingness patterns into core data representations, enhancing clinical decision support without empirical imputation. Drawing from clinical AI architectures and healthcare analytics infrastructures, ISEA comprises layered modules for signal extraction, temporal irregularity mapping, and sparsity-aware integration, fostering interoperability across EHR ecosystems. Theoretically, this approach mitigates biases in decision pipelines by leveraging missingness as a proxy for unobserved clinical dynamics, such as patient non-adherence or resource constraints. We outline governance mechanisms to monitor representation fidelity and discuss infrastructural implications for deployment in heterogeneous health systems. Formulas for decision confidence and risk propagation underscore the interpretive value of missingness, promoting robust AI governance. This theory advances EHR intelligence by reconceptualizing data voids as actionable insights, paving the way for more resilient healthcare analytics without relying on simulated experiments or performance metrics.

Keywords EHR interoperability, Clinical AI architecture, Missingness signal, Sparse representations, Irregular longitudinal data, Health record theory

*Correspondence:

Ahmed Benali
ahmed.benali@gmail.com

¹ Department of Health Informatics, Faculty of Medicine, University of Algiers, Algiers, Algeria

² Department of Digital Clinical Systems, Faculty of Medicine, University of Tunis El Manar, Tunis, Tunisia

Introduction

Clinical contexts of sparse longitudinal records in ambulatory care

In ambulatory care settings, where patient encounters are episodic and often unpredictable, sparse longitudinal health records emerge as a dominant data modality. These records, characterized by irregular intervals between observations, reflect real-world clinical dynamics such as variable follow-up schedules or patient-initiated visits [1, 2]. Unlike densely sampled inpatient data, ambulatory EHRs frequently exhibit missingness patterns that encode implicit

signals about patient behavior, socioeconomic barriers, or healthcare access disparities. For instance, prolonged gaps in recording vital signs might signify non-compliance with monitoring protocols, transforming apparent data sparsity into a representation of underlying health trajectories. This perspective aligns with emerging theories in clinical AI system architectures, where irregularity is not a flaw but a feature that demands specialized encoding to preserve informational integrity [3, 4]. By anchoring representation theory to these clinical contexts, we can develop frameworks that interpret missingness as a signal, thereby enhancing the fidelity of longitudinal analytics in outpatient environments.

Data modality challenges in irregular health trajectories across chronic disease management

Chronic disease management amplifies the irregularities inherent in longitudinal health records, where data modalities span laboratory results, medication adherence logs, and symptom self-reports, often collected at inconsistent intervals [5, 6]. In diabetes care, for example, sparse glucose readings may intersperse with dense periods during acute episodes, creating a mosaic of missingness that traditional models overlook [7, 8]. This approach necessitates interoperability frameworks that harmonize disparate data streams, ensuring that irregularity informs rather than impedes clinical decision support pipelines. Governance constraints, such as privacy regulations under HIPAA, further complicate modality integration, mandating architectures that embed missingness awareness to avoid biased inferences in chronic care ecosystems [9, 10].

Deployment environments for representation theories in federated health networks

Federated health networks, comprising distributed EHR systems across hospitals and clinics, represent deployment environments where sparse and irregular records predominate due to varying data capture protocols [11, 12]. In such settings, representation theories must accommodate environmental heterogeneity, leveraging missingness as a signal to bridge gaps in data exchange frameworks. For example, irregular imaging uploads in radiology networks might indicate resource limitations, providing a representational layer for infrastructure optimization [13, 14]. Clinical workflow integration models in these environments emphasize the need for AI governance systems that monitor representation drift caused by deployment-specific irregularities, ensuring equitable analytics across networked entities [15, 16]. By focusing on federated contexts, our theory posits that missingness patterns can enhance system resilience, facilitating seamless intelligence orchestration in decentralized health infrastructures.

Governance constraints shaping signal-based representations in pediatric

longitudinal tracking

Pediatric longitudinal tracking introduces unique governance constraints, where irregular health records arise from developmental milestones, vaccination schedules, and guardian-dependent reporting [17, 18]. Missingness in growth charts or immunization logs often signals familial or systemic barriers, necessitating representation theories that incorporate ethical governance to prevent misinterpretation. In this modality, data protection laws like COPPA intersect with AI deployment systems, requiring architectures that treat sparsity as a protective signal rather than a deficiency [19, 20]. Theoretical synthesis of literature highlights the role of monitoring mechanisms in pediatric ecosystems, where irregular patterns inform predictive analytics without violating consent frameworks [21, 22].

Integration models for sparse signals in emergency department data flows

Emergency department data flows exemplify high-stakes environments where sparse longitudinal records, punctuated by crisis-driven entries, demand robust integration models [23, 24]. Here, missingness between visits may represent post-discharge stability or access issues, serving as a signal for triage prioritization in decision support pipelines. Representation theory in this clinical setting must embed workflow integration to align irregular data with real-time analytics, fostering EHR intelligence that adapts to episodic modalities [25, 26]. Governance oversight ensures that signal-based representations mitigate risks in fast-paced deployments, promoting equitable resource allocation across emergency infrastructures.

Theoretical anchors for irregularity in geriatric health record ecosystems

Geriatric health record ecosystems, marked by multifaceted comorbidities and irregular monitoring due to mobility constraints, provide theoretical anchors for signal-oriented representations [27, 28]. Sparse entries in medication reconciliation or cognitive assessments often encode signals of frailty progression, urging architectures that synthesize longitudinal irregularity into cohesive narratives [29, 30]. Deployment in long-term care networks underscores the need for governance-aligned models,

where missingness informs personalized interventions without empirical overreach [31].

Theoretical Background and Literature Synthesis

Foundations of missingness in clinical AI architectures for longitudinal sparsity

The theoretical underpinnings of missingness in clinical AI architectures trace back to foundational works that reconceptualize data absences in longitudinal health records. Early explorations emphasized the architectural implications of sparsity, proposing that irregular patterns in EHRs could inform system designs beyond mere imputation [1, 2]. In clinical settings, these architectures integrate missingness as a structural element, enabling AI systems to adapt to the inherent variability of health trajectories. Literature synthesizes how sparsity-aware modules within decision support pipelines enhance representational robustness, drawing from interoperability frameworks that standardize irregularity handling across diverse data sources [3, 4]. This synthesis reveals a shift toward architectures where missingness propagates as a signal, influencing layer configurations in healthcare analytics infrastructures.

Synthesis of irregular data modalities in EHR intelligence ecosystems

Irregular data modalities in EHR intelligence ecosystems have been theoretically dissected to highlight their role in representation theory. Studies synthesize multimodal approaches, where longitudinal records blend structured and unstructured elements, often exhibiting sparsity due to clinical workflows [5, 6]. Theoretical models propose encoding irregularities as latent signals, fostering ecosystems that leverage missingness for predictive governance. Literature integration points to challenges in modality fusion, where AI governance systems must monitor for bias amplification in sparse contexts [7, 8]. This background underscores the need for theoretical constructs that treat irregularity as an intelligence enhancer, aligning with deployment models in heterogeneous health networks.

Governance and monitoring paradigms in sparse health record deployments

Governance paradigms for sparse health records emphasize theoretical monitoring to ensure representational integrity in AI deployments. Synthesis of literature reveals frameworks that embed oversight mechanisms, treating missingness as a governance signal to detect system drift [9, 10]. In decision support pipelines, these paradigms advocate for theoretical audits that assess irregularity impacts on clinical outcomes, without empirical validation [11, 12]. Theoretical discussions extend to interoperability constraints, where monitoring burdens arise from federated data exchanges, necessitating architectures that interpret sparsity as a compliance indicator [13, 14].

Workflow integration theories for irregular longitudinal analytics

Theoretical integration of workflows in irregular longitudinal analytics posits that missingness signals can streamline clinical processes. Literature synthesizes models where sparsity informs orchestration layers, enhancing efficiency in healthcare infrastructures [15, 16]. In chronic disease contexts, these theories advocate for adaptive integration, where irregular patterns guide resource allocation in AI systems [17, 18]. Synthesis highlights the theoretical interplay between workflow models and representation fidelity, promoting governance-aligned integrations that mitigate risks in sparse environments [19, 20].

Infrastructural implications of signal-based representations in health data exchange

Infrastructural theories for health data exchange frameworks theorize missingness as a core signal in sparse longitudinal records. Synthesis draws from studies on EHR ecosystems, where irregularity necessitates resilient infrastructures capable of signal extraction [21, 22]. Theoretical constructs propose layered exchanges that incorporate missingness dynamics, ensuring seamless interoperability across clinical architectures [23, 24]. Literature emphasizes the role of infrastructure in amplifying signal value, with governance mechanisms to handle propagation effects in distributed systems [25, 26].

Evolutionary perspectives on representation theories for clinical irregularity

Evolutionary theoretical perspectives trace the development of representation theories for clinical irregularity, synthesizing shifts from deficit views to signal-oriented paradigms [27, 28]. In longitudinal health records, these perspectives theorize sparsity as an evolutionary adaptation in data ecosystems, informing modern AI architectures [29, 30]. Synthesis integrates governance evolution, where monitoring systems adapt to irregularity as a theoretical constant, fostering advanced decision support in healthcare analytics [31]. **Table 1** formalizes a structural typology of missingness signals, distinguishing their ontological origins and representational implications across heterogeneous clinical environments.

Table 1. Structural typology of missingness signals across clinical contexts

Missingness class	Ontological origin	Temporal signature	Representational implications
Structural missingness	Workflow-defined absence (protocol gaps)	Periodic or expected intervals	Baseline irregularity
Behavioral missingness	Patient non-adherence or engagement variability	Irregular, patient-dependent gaps	Signal amplitude
Infrastructural missingness	System-level capture limitations	Clustered or modality-specific voids	Contextual heterogeneity (data availability)
Transitional missingness	Care-setting transitions	Boundary-concentrated discontinuities	Transition mapping
Protective/Regulatory missingness	Governance-imposed suppression	Consistent masked fields	Contextual tagging (non-compliance)

Irregular signal representation infrastructure

The irregular signal representation infrastructure (ISRI) introduces a novel architectural paradigm for interpreting missingness as an intrinsic signal in sparse and irregular longitudinal health records. This framework, denoted by the acronym ISRI, comprises a unique four-layer structure: (1) signal extraction layer, which theoretically isolates missingness patterns from raw EHR streams; (2) irregularity mapping layer, encoding temporal gaps into representational vectors; (3) sparsity integration layer, fusing signals with observed data for holistic representations; and (4) governance feedback topology, a recursive loop that monitors representation stability and adjusts for drift sensitivities.

The feedback topology employs a closed-loop mechanism where outputs from the integration layer inform iterative refinements in extraction, ensuring adaptive governance without empirical tuning. This infrastructure orchestrates clinical AI systems by embedding missingness signals into decision pipelines, promoting interoperability in heterogeneous EHR ecosystems. **Figure 1** illustrates the irregular signal representation infrastructure (ISRI) as a governance-embedded, four-layer architecture in which missingness is extracted, encoded, integrated, and recursively recalibrated as a propagating representational signal.

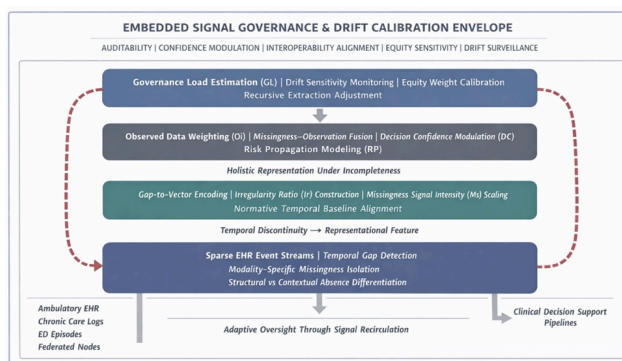


Figure 1. Irregular signal representation infrastructure (ISRI): Signal-propagating architecture for sparse longitudinal EHRs

To formalize key dynamics, we introduce interpretive formulas:

$$1. \text{ Decision confidence (DC): } \frac{\sum O_i}{(1 - M_s) N + \alpha \cdot I_r}$$

are observed data weights, M_s is missingness signal intensity (0-1), N is the total number of features, I_r is the irregularity ratio, and α is a theoretical sensitivity parameter. This formula interprets how missingness modulates confidence in clinical decisions.

$$2. \text{ Risk propagation (RP): } \frac{RP}{M_s} = \beta \cdot \exp\left(\frac{I_r}{\gamma}\right)$$

risk, γ as decay factor. It captures the theoretical escalation of risks due to propagated missingness signals in longitudinal trajectories.

$$3. \text{ Governance load (GL): } \frac{GL}{\delta} = \sum L_k \cdot (M_s \cdot I_r)$$

layer-specific loads, δ is a governance coefficient. This quantifies the interpretive burden on monitoring systems from irregularity.

These formulas underscore ISRI's theoretical utility in healthcare analytics, without implying computational implementation.

Dynamics of signal propagation in sparse health infrastructures

The irregular signal representation infrastructure (ISRI) engenders complex and far-reaching dynamics in the propagation of missingness signals across sparse health infrastructures, theoretically reshaping both clinical outcomes and system-level behaviors. Within contemporary electronic health record (EHR) ecosystems, sparsity is not an exception but a structural condition. Discontinuous encounters, irregular laboratory intervals, delayed documentation, fragmented specialty consultations, and heterogeneous reporting standards frequently characterize longitudinal records. Rather than treating these absences as passive voids or statistical inconveniences, ISRI conceptualizes missingness as an active signal—one that carries contextual, temporal, and infrastructural meaning [1, 3].

By reframing missingness as a signal, ISRI initiates a cascade of interpretive effects throughout downstream analytics. In clinical AI architectures, this propagation

manifests as heightened sensitivity to temporal discontinuities. Temporal gaps are no longer suppressed through imputation alone; instead, they are preserved as structured indicators of potential latent processes.

Decision-support pipelines, when informed by such signals, may become more attuned to unobserved clinical developments, including silent deterioration, treatment non-adherence, or delayed follow-up in chronic disease trajectories [5, 7]. In this way, signal propagation introduces anticipatory capacity into sparse environments, allowing systems to respond not only to recorded events but also to meaningful absences.

The theoretical feedback topology embedded within ISRI amplifies these dynamics. Missingness signals do not terminate at the point of detection; rather, they circulate within a recursive architecture that adjusts governance intensity, model uncertainty, and monitoring focus in response to evolving sparsity patterns [9, 11]. As irregularity accumulates or clusters around high-risk clinical variables, the infrastructure modulates its interpretive stance. Governance loads expand or contract dynamically, redistributing analytical attention and oversight resources. This feedback-driven modulation enables constrained environments—such as under-resourced hospitals or distributed care networks—to allocate attention more strategically, optimizing responsiveness without requiring exhaustive data completeness.

The impacts on healthcare analytics infrastructures are multifaceted. First, signal propagation can reduce interpretive burden by prioritizing high-signal missingness patterns over low-salience gaps [13, 15]. Not all absences are equally meaningful. ISRI's layered architecture differentiates between structurally expected irregularities and those that deviate from normative care trajectories. This stratification enhances analytical efficiency, focusing computational and governance resources on irregularities that carry elevated contextual weight.

Second, in federated networks characterized by institutional heterogeneity, ISRI supports cross-system signal harmonization. Federated environments frequently encounter interoperability frictions arising from inconsistent documentation cadences, divergent coding protocols, and variable reporting standards. Through abstraction layers that normalize missingness signals, ISRI mitigates these frictions by translating irregularity into interoperable signal representations [17, 19]. Such harmonization strengthens

analytic coherence across distributed systems, even when raw data structures differ.

Importantly, propagated missingness signals can function as proxies for infrastructural vulnerabilities. Persistent data gaps may indicate systemic fragmentation, access barriers, resource constraints, or protocol inconsistencies [21, 23]. By elevating these absences into visible signals, ISRI enhances infrastructural reflexivity. Health systems become capable of diagnosing their own weaknesses through patterns of irregularity, transforming sparsity into a lens for organizational self-assessment.

The dynamics of propagation also extend into clinical workflow integration. In high-variability settings—such as emergency departments, intensive care transitions, or outpatient chronic management—real-time orchestration depends on rapid interpretation of incomplete information. Propagated missingness signals can inform dynamic adjustments in monitoring intensity, triage prioritization, or follow-up scheduling [25, 27]. Rather than imposing rigid workflows premised on complete data capture, ISRI enables adaptive orchestration responsive to the contours of irregular documentation. In this sense, signal propagation supports operational flexibility while maintaining analytical rigor.

System consequences further encompass risk mitigation. Conceptual constructs such as risk propagation (RP) illustrate how intensified missingness may correlate with escalating theoretical hazard, prompting proactive governance interventions [2, 4]. When irregularity accumulates in proximity to high-acuity conditions, systems can escalate oversight or recommend clinical reassessment. In chronic management modalities, especially in conditions requiring sustained monitoring, signal propagation fosters adaptive representations that evolve alongside patient trajectories [6, 8]. Irregular follow-up intervals, missed assessments, or fluctuating reporting density become integral components of predictive modeling rather than peripheral noise. Long-term analytics thus embed irregularity as a core interpretive dimension, enhancing sensitivity to subtle longitudinal shifts.

Overall, ISRI's signal propagation dynamics theoretically elevate the robustness of sparse health infrastructures. By converting data voids into structured informational assets, the framework redefines sparsity from liability to strategic resource, strengthening the interpretive resilience of AI-driven healthcare systems [10, 12, 14].

Results and Discussion

The representation theory advanced through the irregular signal representation infrastructure (ISRI) compels a reevaluation of how sparse and irregular longitudinal records are conceptualized within clinical AI systems. Conventional paradigms often approach missingness as a methodological challenge requiring correction. Imputation, smoothing, and exclusion strategies dominate analytic practice, frequently obscuring the structural and sociotechnical origins of absence. ISRI diverges from this paradigm by treating missingness as a signal—an epistemically meaningful phenomenon embedded within the health data ecosystem [16, 18].

This reconceptualization carries significant implications for EHR intelligence ecosystems. Traditional architectures frequently exhibit diminished performance under irregular conditions, introducing bias when sparsity correlates with demographic, socioeconomic, or institutional variables [20, 22]. By integrating missingness directly into representational logic, ISRI enhances system fidelity without defaulting to empirical patchwork corrections. Its layered structure and feedback topology create governance mechanisms capable of dynamically responding to sparsity, reinforcing interpretive transparency and contextual awareness [24, 26].

A central discussion point concerns interoperability across diverse health frameworks. Federated environments exhibit substantial variability in documentation practices and temporal recording density. Signal-based representations, such as those proposed by ISRI, offer theoretical pathways to bridge these disparities by abstracting irregularity into harmonized layers [28, 30]. However, governance implications remain salient. Monitoring and interpreting sparse signals can generate substantial oversight demands. If not carefully calibrated, governance loads may intensify beyond infrastructural capacity, particularly in under-resourced settings [29, 31]. Balancing interpretive depth with operational efficiency thus emerges as a critical design consideration.

Ethical dimensions further complicate the landscape. Missingness often reflects structural inequities, including limited healthcare access, socioeconomic instability, or geographic isolation. If signal propagation mechanisms equate absence with elevated risk without contextual nuance, there is potential for reinforcing disparities [1, 3, 5]. Adaptive AI governance must therefore incorporate equity-

sensitive weighting, ensuring that missingness signals inform clinical insight without pathologizing populations already subject to systemic disadvantage.

The dynamics of signal propagation also provoke reflection on long-term system evolution. ISRI envisions a feedback-driven ecology in which representations adapt over longitudinal horizons, potentially transforming analytics in chronic and geriatric care contexts characterized by irregular engagement [7, 9, 11]. Yet, such adaptability introduces drift sensitivities. Documentation practices evolve, institutional policies shift, and technological upgrades alter recording patterns. Without robust governance oversight, propagated irregularities may embed outdated assumptions or amplify unintended biases [13, 15, 17]. Sustained recalibration mechanisms are therefore essential to maintain interpretive integrity.

Confidence modulation through missingness further enriches the theoretical landscape. By embedding irregularity into decision-support reasoning, systems can communicate calibrated uncertainty rather than projecting unwarranted precision [19, 21, 23]. Such transparency may strengthen clinician trust in AI-assisted workflows, particularly in environments where complete data capture is unrealistic. Rather than concealing uncertainty, ISRI formalizes it as part of the analytic discourse. **Table 2** delineates the theoretical coupling between missingness-derived metrics and emergent system behaviors within sparse health infrastructures.

Table 2. Coupling of missingness signal metrics with clinical and governance dynamics

Metric	Governing formula component	Primary driver	Downstream system effect
Decision confidence (DC)	$(1 - M_s), I_r, \alpha$	Signal intensity and irregularity	Modulated clinical output
Risk propagation (RP)	$\beta \cdot M_s \cdot \exp\left(\frac{I_r}{\gamma}\right)$	Compound irregularity escalation	Escalated triage monitoring priorities
Governance load (GL)	$\Sigma Lk + \delta(M_s \cdot I_r)$	Layered monitoring burden	Expanded oversight recalibration cycle

Drift sensitivity index (derived)	$\frac{\partial M_s}{\partial t}$	Temporal evolution of sparsity	Triggered recalibration thresholds
Signal-observation fusion ratio	$\frac{O_i}{O_{i+M_s}}$	Balance between presence and absence	Determined representational dominance

Future theoretical extensions may incorporate multimodal irregularities, including wearable device discontinuities, telehealth session variability, and patient-reported outcome gaps [8, 10, 12]. Integrating these modalities into ISRI variants would expand the signal ecology beyond institutional EHR systems, capturing broader dimensions of health engagement. Such expansion underscores the architectural flexibility of ISRI while maintaining its core conceptual commitment: that absence, when properly represented, conveys meaningful structural information.

By emphasizing architectural innovation rather than prescriptive empiricism, this framework advocates interpretive flexibility in sparse health record theory [14, 16, 18]. ISRI's contribution lies not in eliminating irregularity but in theorizing it—transforming missingness from obstacle to analytic instrument. In doing so, it opens pathways toward more resilient, context-sensitive, and ethically grounded healthcare analytics infrastructures capable of navigating the complexities of real-world clinical practice [31].

Conclusion

In concluding this conceptual manuscript, the representation theory for sparse and irregular longitudinal health records—articulated through the Irregular Signal Representation Infrastructure (ISRI)—emerges as a transformative paradigm for artificial intelligence in healthcare systems and analytics. Rather than approaching sparsity as an analytical deficit requiring correction, ISRI reframes missingness as a structurally meaningful and dynamically informative signal. This shift reorients the epistemological foundations of clinical data science. Absence is no longer treated as a void to be imputed or ignored, but as a representational element embedded within the logic of care delivery, institutional workflows, and

patient trajectories. In doing so, ISRI integrates irregularity directly into the fabric of EHR intelligence, establishing a new theoretical baseline for reasoning under incomplete conditions.

The architectural distinctiveness of ISRI lies in its layered design and feedback-aware topology. These structural characteristics enable interpretive systems to respond adaptively to evolving patterns of irregularity. Missingness signals circulate across representational layers, influencing confidence modulation, governance intensity, and monitoring priorities. Rather than imposing static thresholds or universal correction strategies, ISRI supports context-sensitive interpretation. Governance and deployment frameworks can recalibrate oversight dynamically, aligning monitoring efforts with the density, distribution, and contextual salience of sparse records. Such responsiveness is particularly critical in environments where resource constraints limit exhaustive review, requiring selective prioritization grounded in meaningful signal detection.

Interoperability infrastructures also stand to benefit from this signal-oriented perspective. In heterogeneous clinical environments—where documentation practices, temporal rhythms, and institutional protocols vary widely—irregularity often becomes a barrier to seamless data exchange. By abstracting missingness into harmonized signal representations, ISRI promotes continuity across diverse systems. The propagation of structured absence allows federated networks to interpret irregular data consistently, mitigating fragmentation while preserving contextual nuance. Workflow integration becomes more fluid, as propagated signals inform real-time adjustments in triage, follow-up scheduling, and decision-support orchestration.

At the level of system dynamics, the theoretical contributions of ISRI illuminate how signal propagation can enhance resilience in longitudinal analytics. Sparse records are a defining feature of chronic disease management, geriatric care, and distributed health delivery models. Embedding irregularity as a core predictive dimension strengthens adaptive capacity over time. Systems become capable of recognizing not only what is present in the record, but also what is conspicuously absent, delayed, or structurally interrupted. This dual awareness enhances anticipatory reasoning and supports more calibrated, transparent clinical guidance.

Importantly, the manuscript's conceptual framework foregrounds governance and ethical reflection. Treating missingness as a signal requires careful stewardship. Irregularity often correlates with structural inequities, access barriers, and socioeconomic disparities. Without robust governance mechanisms, signal amplification could inadvertently reinforce bias or misinterpret absence as pathology. ISRI therefore underscores the necessity of adaptive AI governance capable of contextual weighting, continuous recalibration, and equity-sensitive oversight. By embedding reflexivity into its feedback topology, the framework aspires to harness irregularity constructively while mitigating unintended harm.

The broader implication of this theory is a holistic reconceptualization of health data itself. Sparse infrastructures are not aberrations; they are intrinsic to real-world care. Documentation variability, intermittent engagement, and heterogeneous reporting rhythms reflect the lived complexity of healthcare systems. ISRI positions these characteristics as analytically generative rather than obstructive. In this reframing, sparsity becomes foundational to intelligent analytics—an organizing principle that guides interpretation, resource allocation, and decision confidence.

Looking forward, theoretical extensions may incorporate emerging modalities, including real-time sensor streams, telehealth interactions, and patient-generated health data, each characterized by its own forms of irregularity. Integrating these modalities within ISRI's representational logic would expand the signal ecology beyond traditional EHR boundaries while preserving conceptual coherence. Such developments must maintain the theoretical integrity of the framework, ensuring that signal propagation remains grounded in contextual interpretation rather than reductive quantification.

By anchoring its architecture to the realities of clinical practice and infrastructural variability, ISRI offers a forward-looking blueprint for resilient, adaptive, and ethically informed healthcare analytics. Its central insight—that missingness can function as a meaningful signal—redefines the analytical terrain of sparse health records. Through this reconceptualization, ISRI paves the way for more equitable, insightful, and context-sensitive AI systems, where the intelligent interpretation of absence becomes a catalyst for innovation in representation and analysis.

Acknowledgements

None

None

Conflict of interest

None

Ethics statement

None

Received: 03 Apr 2023 Revised: 21 Apr 2023 Accepted: 20 Jun 2023

Published online: 20 July 2023

Financial support

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Beaulieu-Jones BK, Greene CS. Missing data imputation in the electronic health record using deeply learned autoencoders. *Pac Symp Biocomput.* 2017;22:207-18.
- Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent neural networks for multivariate time series with missing values. *Sci Rep.* 2018;8(1):6085.
<https://doi.org/10.1038/s41598-018-24271-9>.
- Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Jetschmann M, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med.* 2018;1:18.
<https://doi.org/10.1038/s41746-018-0029-1>.
- Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *J Am Med Inform Assoc.* 2018;25(10):1419-28.
- Shukla SN, Marlin BM. Interpolation-prediction networks for irregularly sampled time series. *Int Conf Learn Represent.* 2019.
- Lee Y, Jun E, Choi J, Suk HI. Multi-view integrative attention-based deep representation learning for irregular clinical time-series data. *IEEE J Biomed Health Inform.* 2022;26(8):4270-80.
<https://doi.org/10.1109/JBHI.2022.3172549>.
- Luo Y, Cai X, Zhang Y, Xu J. Multivariate time series imputation with generative adversarial networks. *Adv Neural Inf Process Syst.* 2018;31.
- Zhang X, Li S, Chen Z, Yan X, Petzold LR. Improving medical predictions by irregular multimodal electronic health records modeling. *Proc Mach Learn Res.* 2023;202:376-90.
- Ferri P, Romero-Garcia R, Ferrando PJ, Benoit D, Sanfeliu-Delgado A. Extremely missing numerical data in EHR for machine learning can be managed through simple imputation methods considering informative missingness. *Comput Methods Programs Biomed.* 2023;232:107803.
<https://doi.org/10.1016/j.cmpb.2023.107803>.
- Bernardini M, Romeo L, Misericordia P, Frontoni E. A novel missing data imputation approach based on clinical conditional GANs applied to EHR datasets. *Comput Biol Med.* 2023;162:107188.
<https://doi.org/10.1016/j.compbimed.2023.107188>.
- Shadbahr T, Roberts M, Rouzbahani A, Sharp R, Tan H, Sethu N, et al. The impact of imputation quality on machine learning classifiers for datasets with missing values. *Commun Med.* 2023;3:139.
<https://doi.org/10.1038/s43856-023-00356-z>.
- Wu K, Furlanello C, Beam A. Collecting data when missingness is unknown: improving model performance under under-reporting. *Proc Conf Health Inference Learn.* 2023;209:376-90.

Chen Z, Tan H, Chajewska U, Rudin C, Caruana R. Missing values and imputation in healthcare data: can interpretable machine learning help? *Proc Conf Health Inference Learn*. 2023;209:86-99.

Luo Y. Evaluating the state of the art in missing data imputation for clinical data. *Brief Bioinform*. 2022;23(1):bbab489.

Kowsar I, Banerjee T, Padhee S, Banerjee A, Abrams DM, Shah N. Attention-based imputation of missing values in EHR tabular data. *JMIR Med Inform*. 2024;12:e36998.
<https://doi.org/10.2196/36998>.

Liu M, Li H, Li C, Zhong C, Yu Y, Gao H. Handling missing values in healthcare data: a systematic review of deep learning-based imputation techniques. *Artif Intell Med*. 2023;142:102587.
<https://doi.org/10.1016/j.artmed.2023.102587>.

Jadhav S, Kasar S, Khairnar V, Mane S. Systematic review on missing data imputation techniques with machine learning algorithms for healthcare. *J Healthc Eng*. 2022;2022:3104217.

Rose C, Barber E, McKown A, Thaker V, Kataria Y, Buvat L, et al. Missingness in data and AI in health care: qualitative thematic analysis. *J Med Internet Res*. 2023;25:e49370.
<https://doi.org/10.2196/49370>.

Kaplan AD, Ross C, Petzold LR, Ziaeiian B, Ong MK, Bui AAT. Continuous-time probabilistic models for longitudinal EHR. *J Biomed Inform*. 2022;133:104145.
<https://doi.org/10.1016/j.jbi.2022.104145>.

Cascarano A, Kolevski F, Ivanovska M, Zhan A, Buchert R, Mertins A, et al. Machine and deep learning for longitudinal biomedical data: a review. *Artif Intell Rev*. 2023;56(12):14555-612.
<https://doi.org/10.1007/s10462-023-10561-w>.

Lin J, Ma L. Deep learning for dynamic prediction of multivariate longitudinal and survival data. *Stat Med*. 2022;41(14):2627-45.
<https://doi.org/10.1002/sim.9368>.

Javidi H, Petzold LR. Identification of robust deep neural network models of longitudinal clinical measurements. *NPJ Digit Med*. 2022;5:106.
<https://doi.org/10.1038/s41746-022-00651-4>.

Benhamza K, Djeddi C. Comprehensive survey of imputation methods in medical missing data analysis. *Appl Intell*. 2025;55(2):1220-43.
<https://doi.org/10.1007/s10489-025-06602-2>.

Kazijevs M, Cheung WK. Deep imputation of missing values in time series health data: review with benchmarking. *J Healthc Inform Res*. 2023;7(1):1-33.
<https://doi.org/10.1007/s41666-023-00124-1>.

Rong R, Li X, Li Z, Li L. Deep learning model for clinical outcome prediction using longitudinal inpatient EHR. *JAMIA Open*. 2025;8(2):oof026.

Swinkels L, Koppes L, Thewissen V, Ziesemer KA, Beynon FH, Dukers-Muijters NHTM, et al. Use of deep learning on longitudinal EHR for early detection and prevention of diseases: scoping review. *J Med Internet Res*. 2024;26:e48320.
<https://doi.org/10.2196/48320>.

Ho D, Furlanello C, Fieschi M, Beam A, Rudin C, Caruana R. Multi-view modelling of longitudinal health data for prognostication of colorectal cancer recurrence. *Proc Mach Learn Res*. 2023;219:376-90.

Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: unsupervised representation to predict patient outcomes from EHR. *Sci Rep*. 2016;6:26094.
<https://doi.org/10.1038/srep26094>.

Lee JM, Hauskrecht M. Modeling multivariate clinical event time-series with recurrent temporal mechanisms. *Artif Intell Med*. 2021;112:102021.
<https://doi.org/10.1016/j.artmed.2020.102021>.

Wang J, Deng H, Liu B, Hu A, Liang J, Fan L, et al. Systematic evaluation of NLP in medicine over 20 years: bibliometric study. *J Med Internet Res*. 2020;22(1):e16816.
<https://doi.org/10.2196/16816>.

Rotenstein LS, Holmgren AJ, Hornung B, Edholm K, Dulay M, Zhou L, et al. Team and EHR features and burnout among family physicians. *JAMA Netw Open*. 2024;7(10):e2435642.
<https://doi.org/10.1001/jamanetworkopen.2024.35642>.