From Knowledge Graphs to Digital Twins: Perspectives on Modeling Patient Outcomes for Healthcare Quality Assessment

Anna K. Nitschke, Juan G. Diaz Ochoa, Simone Neumaier, Markus Knott

Submitted to: Journal of Medical Internet Research
on: August 06, 2025

Table of Contents

From Knowledge Graphs to Digital Twins: Perspectives on Modeling Patient Outcomes for Healthcare Quality Assessment

Anna K. Nitschke[1*]; Juan G. Diaz Ochoa[2] Dr rer nat; Simone Neumaier[3*]; Markus Knott[3]

[1] Department of Physics Heidelberg University Heidelberg DE
[2] PerMediQ GmbH Stuttgart DE
[3] Klinikum Stuttgart, Stuttgart Cancer Center Stuttgart DE
[*]these authors contributed equally

Corresponding Author:
Juan G. Diaz Ochoa Dr rer nat

PerMediQ GmbH
Pelargusstr. 2
Stuttgart
DE

Abstract

Medical applications of mathematical modeling, including Machine Learning (ML) models, Knowledge Graphs (KG), and Digital Twins (DT) - primarily involve the prediction of individual patient outcomes. However, to date, the systematic evaluation of these models' contributions to healthcare quality has been explored only to a limited extent. This perspective article systematically examines how mathematical modeling can be integrated into health care quality management. It begins with a comprehensive overview of quality assessment in health care, addressing the definitions of procedures, patient outcomes and Quality Metrics with a quantitative focus. The emphasis is subsequently placed on three categories of Patient-Centered Quality of Care (PCQC), namely, patient safety, procedure accuracy and procedure efficacy, which provide both conceptual and mathematical descriptions. Different levels of modeling tasks essential for managing PCQC are identified. This quantitative concept provides the foundation for the subsequent implementation of ML models. This article facilitates a deeper understanding of the topic by assigning relevant publications to these three quality categories. Focus is placed on the applicability of KGs and DTs to improve quality management in healthcare.

(JMIR Preprints 06/08/2025:81946)
DOI: https://doi.org/10.2196/preprints.81946

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
- ✓ **Please make my preprint PDF available to anyone at any time (recommended).**
- Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
- Only make the preprint title and abstract visible.
- No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
- ✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
- Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
- Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
- No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <a href="https:/

Original Manuscript

From Knowledge Graphs to Digital Twins: Perspectives on Modeling Patient Outcomes for Healthcare Quality Assessment

Anna K. Nitschke*
Heidelberg University, Department of Physics, Im Neuenheimer Feld 226, D-69120 Heidelberg, Germany
nitschke@physi.uni-heidelberg.de
https://orcid.org/0009-0009-5891-4298

Juan G. Diaz Ochoa*
PerMediQ, Pelargusstr. 2, D-70180 Stuttgart, Germany
Juan.diaz@permediq.de
https://orcid.org/0000-0002-9893-4068

Simone Neumaier
Klinikum Stuttgart, Stuttgart Cancer Center, Kriegsbergstr. 60, 70174 Stuttgart, Germany
https://orcid.org/0009-0005-0215-8105

Markus Knott
Klinikum Stuttgart, Stuttgart Cancer Center, Kriegsbergstr. 60, 70174 Stuttgart, Germany
https://orcid.org/0000-0002-9147-3203

* Corresponding authors

Abstract
Medical applications of mathematical modeling, including **Machine Learning (ML)** models, **Knowledge Graphs (KG)**, and **Digital Twins (DT)** - primarily involve the prediction of individual patient outcomes. However, to date, the systematic evaluation of these models' contributions to healthcare quality has been explored only to a limited extent. This perspective article systematically examines how mathematical modeling can be integrated into health care quality management. It begins with a comprehensive overview of quality assessment in health care, addressing the definitions of procedures, patient outcomes and Quality Metrics with a quantitative focus. The emphasis is subsequently placed on three categories of **Patient-Centered Quality of Care (PCQC)**, namely, patient safety, procedure accuracy and procedure efficacy, which provide both conceptual and mathematical descriptions. Different levels of modeling tasks essential for managing PCQC are identified. This quantitative concept provides the foundation for the subsequent implementation of ML models. This article facilitates a deeper understanding of the topic by assigning relevant publications to these three quality categories. Focus is placed on the applicability of KGs and DTs to improve quality management in healthcare.

Keywords: Quality Management; Medicine; Patient Outcomes; Machine Learning; Knowledge Graphs; Digital Twins.

1 Introduction

In recent years, considerable efforts have been made to improve patient outcomes by standardizing medical care for patients [1]. This has been achieved in particular through the creation and application of evidence-based medical guidelines [2]. Here, standardization, i.e., the design of medical procedures via guidelines aimed at the systematic integration of scientific evidence, plays a relevant role. This is achieved by selecting the most appropriate health decisions required during the patient's journey. Standardizing patient care is expected to improve cost efficiency, efficient workflows, and resource allocation [3].

Many forms of medical interventions (i.e., selection of a therapy) can be seen as part of well-defined processes, i.e., a series of actions taken to achieve an outcome [4]. This concept intersects with industry, where processes refer to a set of interconnected tasks that convert input into specific outputs. In this way, modern medicine can be seen as a set of planned processes recommended by guidelines. This leads the care of the patients from the initial state to the desired result. This overlap has led to industrialization in medicine, with both positive (evidence-based treatments) and negative aspects (dehumanization of medicine, focus on economic aspects) [3].

A misunderstanding of a process-oriented and industrial perspective of healthcare has led to a focus on process optimization and economic optimization instead of patient care (for a review of the industrialization of healthcare and its critics, see Da Silva [5]). On the other hand, a patient-centered and process-oriented perspective can lead to positive aspects such as silo breaking in clinical data, optimizing an outcome (i.e., patient recovery after therapy), minimizing medical errors and expenses for diagnosis and treatment, and maximizing clinical outcomes [6]. In such cases, the balance between standardization and customization is a paramount problem[3].

Records containing medical interventions when a patient receives medical treatment (for instance, electronic health records (EHRs)) are essentially defined as processes with different steps, consisting of diagnosis, medical interventions (such as medication, surgery, etc.), and tracking a patient's condition to prevent further health problems [7]. Thus, it is assumed that a patient's process can be modeled in a similar way to the dynamics of a productive industrial process. Through the logical integration of different steps, both the well-being of the patient and the economic result of the institution (e.g., hospital) can be optimized.

Mathematical modeling, particularly machine learning (ML), can significantly improve the way medical processes are implemented and meet quality standards. However, medical data, including image recognition data, are not used beyond direct patient care. The reasons for this are manifold and include a lack of digitalization, inadequate structured health data, stringent data protection laws, insufficient technical infrastructure, and limited personnel or financial capacities. In addition, healthcare providers often use medical jargon and classification schemes (for example, the International Disease Classification (ICD-10) codes for diagnosis classification) in records, which are country-specific and even center-specific and make it difficult to read machine data.

In addition, HCPs are still not sufficiently trained in ML technologies. On the other hand, IT specialists rarely understand clinical processes in healthcare facilities. In these dynamic times, it is crucial to bring different disciplines together if ML applications are to provide real benefits in everyday clinical practice [8]. Assuming that notions of causality can be identified in medicine and used to identify mechanisms and thus derive predictive models[9] [10], ML applications can provide information about patient outcomes, unexpected or undesired events, and support medical decision-making.

When reviewing PubMed, we found that there is currently no **perspective (and review) article** on **knowledge graphs (KGs**, i.e., the recording of medical information and knowledge via geometric representations) **or digital twins (DTs**, i.e., the digital or virtual representation of a patient) applied to analyze patient outcomes for quality assessment in medicine.

This perspective article presents a novel analysis of this topic and emphasizes the relevance of considering ML models within the derived quantitative concept of health care quality management. The analysis and prediction of patient outcomes is not only relevant from a clinical perspective but also an essential part of quality management in medicine. To date, this relevant topic has been explored only to a limited extent. This work aims to contribute to improving health and well-being (according to the UNO sustainability goals, SDG3)[11].

In Chapter 2, we present a definition of patient outcomes that can be represented quantitatively, enabling mathematical modeling. Chapter 3 provides a detailed introduction to the three categories of patient-centered quality of care, namely, patient safety, procedural accuracy and procedural effectiveness. These definitions provide a conceptual framework for mathematical modeling, to which exemplary publications such as GNNs or DTs are assigned in Chapter 4. Chapter 5 evaluates the perspectives presented in Chapter 4. Finally, in Chapter 6, we present future research directions in this area.

2 Definition of Patient Outcomes for Quantitative Quality Assessment

In medicine, it is generally difficult to define patient outcomes **quantitatively** (related to measurable information: exact metrics and observables captured by sensors) and **qualitatively** (related to nonmeasurable descriptions: categories) because patients are heterogeneous and complex organisms, and an overall assessment of their health status is a challenge.

For mathematical modeling, metrics must be introduced so that we can map the patient outcome to numerical values through observables (measurable parameters). Our goal in this study is to identify these variables of interest and place them in a broader understanding of quality assessment, thus paving the way for mathematical modeling in this field. To this end, we aim to introduce a description of patient outcomes as a fundamental concept for evaluating quality indicators (**QI**s). These indicators are relevant for decision-making because they provide an expected benefit; thus, in situations of uncertainty, one should prefer the option with the greatest expected desirability or value [12], which, in medicine, is the improvement of the quality of care (improvement of patient well-being, avoidance of patient death, the optimal use of resources, etc.). The mathematical description, however, covers a small portion of disease management and cannot be considered fully comprehensive.

2.1 Patient Outcome

From a medical perspective, a patient outcome is defined in terms of what is meaningful and valuable to the individual patient[13]. However, this definition is very general and difficult to relate to a quantifiable metric. As reported by Kersting et al., there is still a conceptual problem with the general definition of what results are in medicine [14]. Patient-relevant outcomes (morbidity, mortality and quality of life) and surrogate or biomarker outcomes are included. These outcomes can differ significantly depending on the underlying disease. After extensive research, Liu et al. [15] summarized the definition of patient outcomes related to clinical practice “as any change [within the patient's health status] that results from health care [for which each profession] has developed outcome measures that focus on the standards, activities, and impact of its discipline”.

To summarize, patient outcomes are manifold and range from **patient-centered quality of care** [14] to **institutional performance** [16] to **patient-reported outcomes** [17], which can be assessed from different data sources. A comprehensive overview of strategies for health care quality improvement and an overall conceptual framework was published by Busse et al. [18].

Table 1 **Subdivision of patient outcomes into three areas (columns)**: patient-centered quality of care, institutional performance and patient-reported outcomes. For each area, a definition and a further subdivision into categories are given to enable a quantitative description of the different areas.

	Patient Centered Quality of Care	Institutional Performance	Patient Reported Outcomes
Definition	Patient centered QIs in health care [19] measured or captured by HCPs [14].	Process execution and resource consumption on an institutional statistical evaluation level [16].	Patient’s subjective evaluation of health experience. Report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else [17].
Exemplary Categories	Medical evaluation of: ● Patient Safety ● Procedure Accuracy ● Procedure Efficacy	Financial evaluation of: ● Patient waiting time ● Economic efficiency of medical procedures ● Productivity	Questionnaires evaluating: ● Symptoms ● Quality of life ● Healthcare experience

In this perspective article, we focus only on the description of patient-centered quality of care (first column, Table 1), for which a comprehensive review is published by Kersting, Kneer, and Barzel [14]. We aim to provide a reference for the general concept, which is why we generally refer to medical procedures as follows:

*Within this article, we refer to a **medical procedure** as the overarching phrasing for any type of observation, measurement, diagnosis or treatment performed on the patient.*
Here is relevant to be cautious, since not all patient outcomes can be measured objectively and only a small portion of them can be translated into clear metrics. In other words, the mathematical model can reflect only a fraction of the clinical reality.

In the context of clinical trials, great efforts have been made to establish quantifiable patient outcomes. Clinical trials use predefined end points to measure patient outcomes. Typical endpoints are patient survival at a given time, incidence of clinical events (such as stroke), clinical performance measures, or patient-reported outcomes [20].

We introduce a definition of patient outcomes that can be considered acceptable for the quality management of medical procedures through a quantitative presentation called quality metrics. Considering the impact that electronic health records (EHRs) have on the quality of care, we provide a definition of patient outcome on the basis of changes registered in the health record and defined as events [21]**.** To this end, we assume that the data relevant for the definition of patient outcomes are machine-readable:

Within this article, we refer to ***patient outcomes*** *as events registered in the health records. Such events can include changes in a patient's health state or changes in medical procedures. They can be assigned to three categories Patient Safety, Procedure Efficacy and Procedure Accuracy.*

According to this definition, by considering such outcomes, providers can use a patient-centric approach while allowing organizations to understand the impact of their services on patients. It can also identify key areas for improvement to deliver better overall care.

2.2 Medical Guidelines

Medical guidelines are systematically developed statements that reflect the current state of medical knowledge. Medical experts prepare these in cooperation with professional associations. These evidence-based guideline recommendations are based on clinical studies published in the scientific literature and/or other existing evidence informed by expert experience and consensus.

The extracted information is then assessed and published by organizations, such as the World Health Organization (WHO), for international guidance or by country-specific professional societies (in Germany, the *Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V*. [22]). Guidelines are not a static corpus and are continuously adapted and further developed on the basis of current clinical studies and scientific findings. They are intending to support the decision-making of healthcare practitioners (HCPs) to enable appropriate care for certain health problems. Medical guidelines consist of complex workflows in mostly narrative form and can also contain multiple, equivalent treatment alternatives for the same disease unit.

2.3 Quality indicators

To assess how guidelines are applied and to improve the quality of care, several **QIs** have been developed by different institutions. These QIs serve as internal quality management tools for medical institutions, benchmark with other institutions, and aim to improve the quality of care for patients. In Germany and Switzerland, routine billing data from hospitals are used for this purpose if a health care facility decides to participate voluntarily in such initiatives [24]. In Germany, QIs for cancer treatment are defined annually by the German Cancer Society (DKG). Hospitals that fulfill the DKG criteria can be DKG certified. These DKG QIs are derived from medical guidelines and should make some specifications quantifiable. The outcomes of cancer patients treated in these DKG-certified centers are superior in terms of patient survival [25].

2.4 Quality Metric for Patient-Centered Quality of Care

Mathematical modeling is used to represent a real-world phenomenon by capturing and analyzing the most significant relationships between different observables through relationships between mathematical parameters. For the quantitative description of patient outcomes, they need to be described by numerically or categorically measurable parameters. These parameters capture various relevant aspects of the overall performance of patient-centered care quality. In this publication, these numerical and categorical parameters are referred to as quality metrics (QMs) because they enable an evaluation of patient care. Some QMs may be similar to the QIs currently used in the medical quality assessment process. Nevertheless, they are now placed into this new perspective and context of mathematical modeling.

Within this article, we refer to ***Quality Metric (QM)*** *as the quantitative QI, which is related to compliance with medical guidelines (Procedure Accuracy) as well as patient state centeredness of care (Patient Safety and Procedure Efficacy).*

The exact mathematical definitions of procedure accuracy, patient safety and procedure efficacy are provided in the next sections.

3 Patient-Centered Quality of Care and Use Cases

In this section, we present an overview of the broad concept of patient-centered quality of care. According to the classification provided in Table 1, we define in detail the three basic categories that we use for further discussion of healthcare quality assessment:

1. Patient Safety
2. Procedure accuracy
3. Procedure efficacy

Table 2 provides intuition about the different aspects of this field of quality assessment, as well as possible mathematical modeling applications. We will introduce the definitions of the prospective and retrospective analyses. In detail, we look at the three categories and discuss the quantities of interest in their quality assessment, which can be used as quality metrics (QMs).

Table 2 **Identified use cases for quality management**.

Use Case	Description	Example of Potential Application of Mathematical Modeling
Use Case 1: Patient Safety: ***real time patient/ biomarker tracking***	Real time information about the patient's condition provided by measurements (e.g. by wearables), allows monitoring of prespecified deviations indicating a critical patient's health state.	Software alerts health care professionals immediately so that information will not get missed and direct actions can be taken (avoids information delay, bad documentation, or miscommunication). This can be for example implemented for real-time measurement of blood glucose increases or decreases in outpatients (however, such technologies still require appropriate implementations and cautious use in order to avoid overdiagnosis [26])
Use Case 2: Procedure Accuracy: ***retrospective analysis of QIs***	Control of the overall quality of a health center by observing the compliance to guidelines (audit in the application of guidelines) or whether local QM specifications are fulfilled or tracking of potential	A quality manager uses a mathematical model on retrospective data to track if certain complications after an operation have significantly increased. If this is the case, investigations in the affected

	deviations in the healthcare process that might lower the quality of care (for instance adapting concepts from process mining in industry[27])	department need to be initiated.
Use Case 3: Procedure Accuracy: ***prospective reminder system (clinical decision support)***	Generated check lists/graphic visualization provides guidance to physicians by giving an overview of all possible procedures as well as a recommendation of the most suitable procedures regarding the medical guidelines. This is possible using a recommender system that filters all the possible procedures.	Depending on the patient's state, the mathematical model reminds HCPs for necessary procedures and checks if all previous procedures have been performed and otherwise reminds HCPs.
Use Case 4: Procedure Efficacy: prospective analysis for ***deep insight***	Many patients have complex and long-term diseases and have been treated in different hospitals. It is time consuming for HCP to overlook all aspects of a patient's journey.	A (graphic) Digital Twin could provide a longitudinal overview of the patient's medical journey and integrate all available data to the given procedures. It could then offer suggestions for further procedures within the frame of the medical guidelines by generating a Digital Twin[28].

For patient-centered quality of care, there is a strong reference to medical guidelines and appropriate clinical practice. Owing to the increasing complexity of medical knowledge, which is also reflected in the guidelines, it is becoming increasingly difficult for healthcare professionals to keep track of changes. The application of mathematical modeling is intended to help healthcare professionals comply with the guidelines, i.e., a sequence of recommended procedures. This is not only an analysis tool but also a form of cognitive or clinical decision support.

The cognitive aspects of practitioners are fundamental to medical decision-making, especially considering how biases and heuristics influence this decision-making [29]. As a result, how the use cases presented in Table 2 are implemented should provide a balance between data-/information-centric approaches (i.e., mathematics modeling) and heuristics.

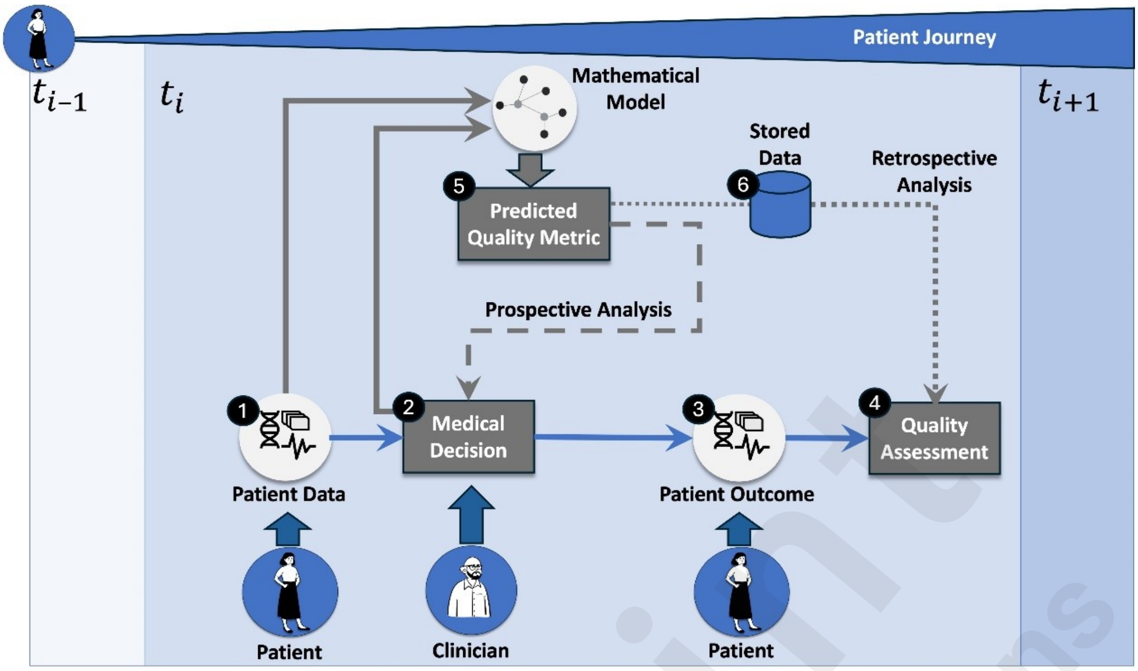

Figure 1 **Visualization of key parameters for health care quality management and mathematical model integration throughout the patient's journey over time.** The predicted quality metric of the mathematical model can be implemented in two ways: prospective (dashed line) or retrospective (dotted line). It includes the following elements: (1) data that are collected from the patient, (2) the medical decision is made by a clinician, (3) the patient outcome after the procedure is performed, (4) the quality assessment and (5) the predicted quality metric from the mathematical model can be stored (6). This can analogously be performed at the population level (patient–cohort predictions) instead of at the patient level (patient–centered predictions).

Mathematical modeling can be included for predicting QMs to enhance healthcare quality. In this context, the patient's journey can be understood as a complex chain of procedures that can be logically interconnected. Therefore, it is imperative to look at the whole patient journey as a process for which each procedure is just one point in time $[t_{i-1}, t_{i+1}]$. Like in Figure 1, the general way in which health care quality is assessed can be described as follows: collecting data from a patient (1), which a clinician uses to make a medical decision regarding the patient's further procedure (2) and whose impact is measured after the procedure is carried out, by looking at the patient outcome (3), which allows assessment of the quality of the healthcare provided (4). On the basis of the examples included in Table 2, one can distinguish between two methods of implementation: use the prediction in a prospective way (dashed arrow, meaning that before the medical decision is made, the clinician receives feedback from the model and can adjust the decision, leading to potentially different outcomes and improved health care quality); and in a retrospective inclusion (dotted arrow), allowing an evaluation of the observed patient outcome and an assessment of the quality of care.

Another distinguishing feature of mathematical modeling is whether the model works for individual patients or for patient cohorts. Hence, for modeling QM, we can consider the following representation levels:

I. Patient-cohort predictions for quality management of patient populations: statistical modeling of the patient population for controlling domestic and global medical practices. At this level, mathematical models describe an “average and black box virtual patient" that tests or corroborates published and domestic guidelines.

II. Patient-centered predictions: Modeling of specific targets and inclusion of personal information in gray (partial knowledge of model details) or white-box (full knowledge of the model details) models to model “virtual individual patients” and evaluate in this way published or domestic guidelines at an individual level. This implies that models should reflect and predict individual deviations from the expected results of the main patient cohort (see, for example, Peng et al. [30]).

The patient journey can be represented as a chain of observed patient outcomes $\{O_t\}$ (for instance, the main disease and corresponding codiseases) and possible medical decisions $\{MD_t^1,\ldots,MD_t^k\}$ that lead again to distinct patient outcomes $\{O_{t+1}^1,\ldots,O_{t+1}^l\}$ over time that are patient–journey dependent. Medical guidelines specify possible medical decisions following patient outcomes. The quality of the health care received during the patient journey can be assessed by comparing different elements within this tree structure. Three categories of patient-centered quality of care can be distinguished: patient safety, procedure accuracy and procedure efficacy. In the following, the 3 different categories of modeled patient outcomes are defined by describing their context to the patient journey in a qualitative and quantitative way, as well as identifying suitable quality metrics.

3.1 Patient Safety

Providing an exact definition of patient safety is challenging, as this term remains under debate in the literature [31]. Patient safety refers mostly to the avoidance of accidental and unwanted events in health care. Patient safety is usually measured **within** health care facilities via predefined inpatient quality indicators (IQIs). These QIs are mostly determined from the billing data of the hospitals and the medical documentation data contained therein (ICD-10 codes for diagnoses, OPS codes for medical procedures) [32] [https://qualityindicators.ahrq.gov/]. Some examples of IQIs that reflect patient safety are as follows (see Table 3 [Extracted from the AHRQ [33]).

Table 3 IQIs and corresponding descriptions

IQI	Description
Complication	**Complication-**related indicators (CREs), procedure-related errors (PREs), medical-related errors (MREs) and hospital system indicators (MSIs) are often used in several studies and applications. This indicator is, however, problematic when the cause of mortality is difficult to establish, especially in multimorbid patients.
Hospital readmission	Equivalent to patient movement in the hospital, for instance, the probability of (re)-admission to an emergency department or intensive care unit.
Patient safety indicators	Unexpected clinical interventions related to poor management of a disease (for example, infections after operations).

Assuming that physicians are rational decision makers, the medical decision is an evaluation of the risk of previous and future outcomes (according to the classic *Theory of Rational Decision Making* of Debreu, see, e.g., B. Walliser[34]); this decision is selected on the basis of a probability (assuming that physicians are either using heuristics or making rational decisions on the basis of available information[29]) to initiate a change in the patient's health state.

There are a variety of possible outcomes - $\tilde{O}_t$ -, some of which are desirable - $\tilde{O}_t^D$ - and some of which are adverse - $\tilde{O}_t^A$ - (implying that we analyze medical decisions producing physical and clinical effects, such as surgical procedures and radiation therapy; observe that an adverse outcome is evaluated depending on the specific clinical context). A mathematical model for patient safety MI_s can check (retrospective) or predict (prospective) how likely the outcome of a medical decision will be an element of the set of adverse outcomes (e.g., alert system from Use Case 1 - Table 2).

Owing to their rutinary nature, IQIs are suitable for quantitative representation. By integrating relevant data, it should be possible to establish correlations between IQIs either at the patient population level or at the single-patient level. Thus, from the IQI's appropriate **quality metrics, the following** can be defined:

- **The mortality rate** and **failure to rescue** have emerged as important quality metrics [35].
- Emergency **readmission to the hospital** after a procedure in outpatients. This can happen, for instance, if prophylaxis has been forgotten or if safety guidelines have not been properly followed.
- **Number of reoperations or complications** shortly after the operation. This might indicate a need for investigation within the facility.
- **Admission to intensive care units (ICUs)** and length of stay at the ICU among inpatients (different from the intermediate medical care unit (IMC)).

3.2 Procedure accuracy

The purpose of medical guidelines is to reduce errors and disparities in care while supporting best practices and responsibilities in medicine. Guidelines are more likely to lead to improved clinical outcomes if they are followed [36], [37], [38].

Another appropriate measure of patient-centered quality of care is adherence to medical guidelines, as well as the correct and accurate application of a medical procedure, which we refer to as procedure accuracy. To date, there is no structured or automated assessment of whether the recommended medical guidelines have been followed.

In general, comparing performed procedures with guidelines is challenging, especially considering that guideline recommendations do not have to be understood as rigid limits. In medical guidelines, a distinction is made between several grades of recommendation on the basis of available evidence [see, e.g., AWMF register for prostate carcinoma (in German) [39]].

For the inclusion of mathematical modeling for health care quality assessment, it is necessary to identify metrics that can be defined precisely (if possible, with binary or clear multilevel indicators). In most cases, these metrics cannot be defined universally and require disease-specific development. Thus, with respect to procedure accuracy, mathematical modeling can provide useful insights into the following:

- **Anomaly detection** and mathematical models can be helpful in discovering common patterns in the relationship between diagnoses and medical procedures [40] or in retrospective analysis of clinical narratives via natural language processing (NLP) [41]. Such pattern recognition can be used to discover for which patient the pattern deviates from the guidelines and if such anomalies are an indication of a patient-centered procedure [42].
- The creation of **local guidelines** (which are not merely deviations from national standards but adaptive tools shaped by real-world experience) is drawn up by hospitals or healthcare institutions. These include local adjustments to the general guidelines when differences are identified, and the efficacy of alternative methods has been demonstrated [43].
- The creation of **patient-specific guidelines**, for which the integration of additional data, such as physiological or genetic data, into the models can provide recommendations for action that consider specific patient variations. In this way, models enable the implementation of

patient-centered quality management.

Clear results include checking whether the recommendations of a guideline for performing a standardized procedure or a set of procedures have been followed in an individual patient. This information can be retrieved from EHRs, laboratory information management systems (**LIMS**) or even reimbursement data. In this way, not only is compliance with a guideline checked, but medical errors can also be reduced (quality metrics described in Table 4).

Table 4 Quality metrics.

Quality Metric	Description
Detection of patient outliers	Average amount of hospitalization required by a medical procedure (length of stay in hospital compared with average) and the appearance of adverse events (number of infections, etc., compared with average).
Expected medical procedure	On the basis of the ICD diagnosis, specific OPS procedures can be predicted, and OPS code deviations from the prediction can be detected. (OPS classification, which is equivalent to the International Classification of Health Interventions (ICHI) [44]).
Revision operation	Number of necessary additional surgical interventions after surgery, e.g., due to complications of the initial surgical intervention.

While these metrics are easy to extract at the site level, they risk missing relevant contextual factors. It would be desirable to integrate additional qualitative or clinical data points for a more holistic assessment. However, in clinical practice, easy numbers are used.

3.3 Procedure efficacy

A general definition of efficacy is provided by Lynch [45], for which "efficacy is the capacity [of a medical intervention] to produce an effect. For quality assessment, procedure efficacy now refers to the optimal selection of procedures for the best possible patient outcome.

Retrospective models could compare the medical decisions chosen by HCPs with other possible choices, leading to better outcomes. Hence, procedure efficacy can also be considered for monitoring a decision (made under imperfect information) by observing the patient's response or for making a prognosis and a subsequent decision under imperfect information.

For example, quantitative measures that can be inferred to quality metrics include the following (see Table 5).

Table 5 Metrics for Procedure Efficacy

Efficacy metric	Description
Biomarker levels	The prediction and projection in time of **biomarker levels** and their relation to accepted levels, to assess and predict the patient's condition (acceptable, i.e., considering parameters below a threshold; or deteriorating, i.e., parameters above the threshold), can lead to specific medical

	decisions. For example, a stable blood glucose range over time indicates the efficacy of diabetes treatment.
Tumor-free survival	Time (in months or years) from complete tumor resection until tumor recurrence.
Overall survival	Indicates the time (in months or years) from diagnosis until death from any cause.
Pathologic complete remission	Describes the histological assessment of the former tumor region after neoadjuvant therapy. For pathological response described in the report, further treatment can be adjusted.
Risk scores	Quantify, for example, the risk of an improvement or deterioration of a specific medical condition.

4 Mathematical Modeling of Patient-Centered Quality of Care

The mathematical representation of the patient's condition to predict the outcome is essential for a precise evaluation of quality management, both for the evaluation of compliance with guidelines and for the overall quality of medical procedures.

Although defining a suitable metric for quality management is still a research topic, we have identified some exemplary metrics from Section 3, which are summarized in Table 6. The "Combination Event and Time Period" column indicates that the observed metric is provided in conjunction with a time parameter (such as the period after hospital discharge).

Table 6 **Classification of quality metrics according to the taxonomy introduced in Section 3**. In this table, OPS refers to medical procedures. In this table, the symbol Δ refers to a difference with respect to a reference value. If the datatype is binary, the data include only two possible values {yes, no} or {0, 1}.

Patient Centered Quality of Care Categories	Quality Metric	Datatype	Correlation: Patient status	Combination Event & Time Period
Patient Safety	Failure to Rescue	Binary	OPS ↔ Mortality	Yes
	Hospital Readmission	Binary	-	Yes
	Admission to Intensive Care Unit (ICU)	Binary	OPS ↔ Critical care	Yes
Procedure Accuracy	Detection of Patient Outliers	Binary	OPS(Patient) ↔ OPS (Population)	No
	Δ OPS Classification	Multilabel	OPS(Patient) ↔ OPS (Population)	No
	Revision Operations	Binary		Yes
Procedure Efficacy	Δ Biomarker (minimal residual disease, CEA...)	Time Series	Biomarker ↔ OPS	-
	Tumor-free Survival	Binary	Biomarker/Time	Yes
	Recurrence-free	Binary	Biomarker	Yes

	Survival			
	Overall Survival	Binary	Biomarker	Yes
	Pathologic Complete Remission	Binary	Biomarker	Yes

In this table, we refer to tumor-free survival, not survivability, which is inherent capacity or likelihood not used as an outcome but for legal purposes when talking to patients.

On the basis of the data stored in the database/graph structure, single mathematical models (3), for which we focus on **ML** models, are trained to accomplish tasks that can either be **(A) classification tasks** of the patient state or **(B) predictions of parameter** evolution owing to the patient response. Specifically, Task A is used when predicting patient safety and procedure accuracy quality metrics. The focus of Task B is on the efficacy and safety of procedures. On the basis of the proposed metrics, the model predictions can be either prospective or retrospective.

In the present article, we focus on KGs to integrate information from single events within the patient journey. By analyzing patterns from these KGs, we will be able to better understand the changes in patient outcomes and evaluate the efficacy of medical procedures. These patterns can be extracted via methods such as graph neural networks.

Additionally, the use of Case 4 within Table 2 demonstrates that for some quality metrics, it is necessary to consider the medical process in its specific context within the patient's journey, which can be understood as a chain of events. This is an important aspect of interest when accessing Procedure Efficacy, as it strongly depends on the previous and following procedures. We want to present the concept of DTs as a model integration framework, enabling patient path representation.

Table 7 **Number of publications in the last 6 years for queries considering "patient outcomes" AND: "Machine Learning" (ML), "Graph Neural Networks" (GNNs) or "Digital Twins" (DTs).**

Year	2019	2020	2021	2022	2023	2024
ML	897	1513	2282	2651	3159	2381
GNN	13	8	12	18	24	23
DT	2	5	7	7	19	27

This growing interest and relevance of graph modeling is indicated in Table 7, which includes the number of published articles from the last 6 years (obtained after querying PubMed23 and entering "patient outcomes and *row* names"). There is a growing interest in the application of KGs (in particular, graph neural networks (GNNs)) to predict patient outcomes.

We also analyzed and observed similar trends when we queried terms such as patient outcomes and DTs. However, the number of publications in these fields is only a fraction of the total number of publications about the application of ML to predict patient outcomes.

Although GNNs and DTs are methods suitable for multimodal modeling leading to precision medicine (in the framework of quality management), this topic will not be the main focus of the present work; in such cases, the reader should refer to articles that focus on this specific topic, such as the scope review article written by Kline et al. [46]. In the following chapter, we present literature that can be assigned to the different quality management categories derived from the previous chapters.

4.1 Graph-based methods for modeling patient safety and procedure accuracy

With electronic health records, extended research integrating different patient data for health care quality assessment is now possible. Regarding this development, Si et al. published a comprehensive survey about the use of current advances in patient representation on the basis of data stored in EHRs [47].

In recent years, the application of linked knowledge in graphs has become relevant, in part because this allows holistic individualized representation, either considering individual characteristics in an entire population [40] or for a patient-centered representation of the disease and the corresponding outcomes. On the one hand, data represented as a graph is a method to leverage transparency in model construction. Graph models are also a way to map unstructured data, such as clinical narratives, into machine-readable data via large language models (LLMs). Additionally, graph models, including graph neural networks (GNNs), have been considered a way to close the gap between ML and symbolic reasoning [48], introducing the desired interpretability for ML methods.

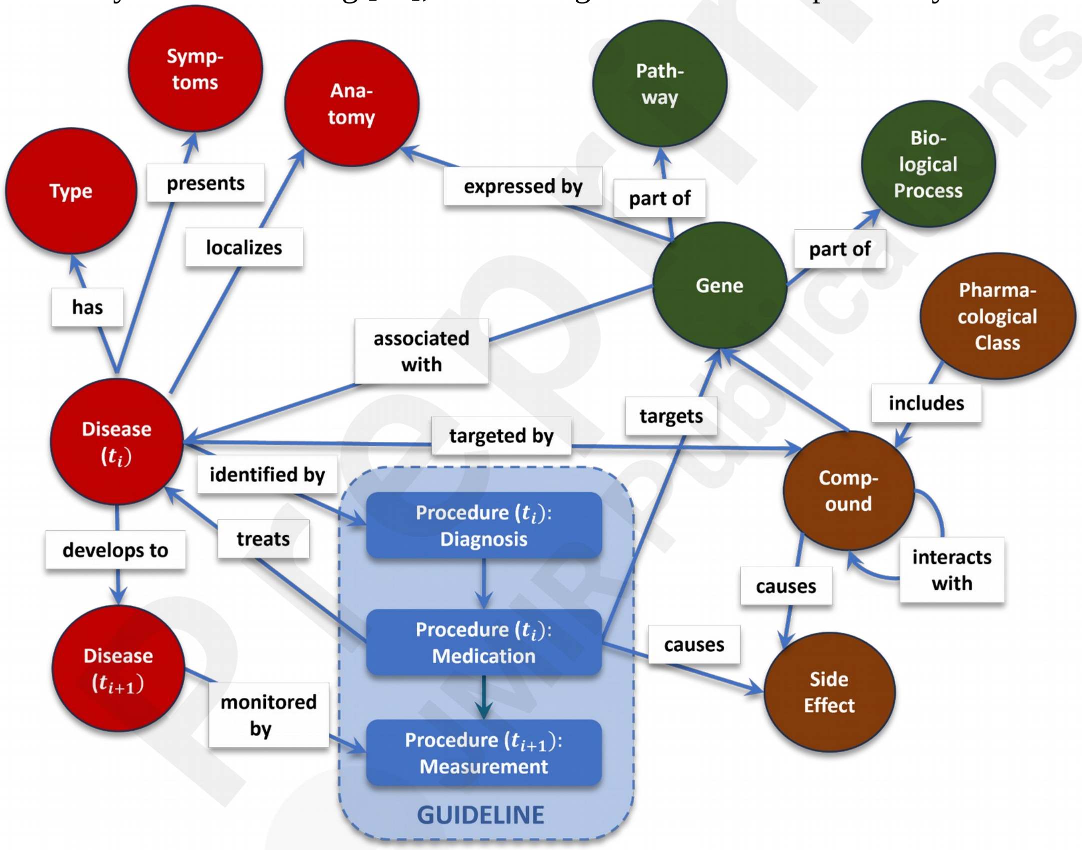

Figure 2 **A simple healthcare knowledge graph inspired by the article of Abu-Salih et al. (Abu-Salih et al. 2023).** The abstract graph shows how a disease is connected to its attributes (such as type, symptoms and anatomy), as well as to compounds (and attributes such as side effects and pharmacological classes), and is associated with genes (and attributes such as pathways and biological processes). Interactions with the disease, which we refer to as medical procedures (such as diagnosis, medication and measurement), can be added to the graph and lead to the development of the disease over time ¿¿).

The sample healthcare KG in Figure 2 can serve to represent the concept of graph models in health care: the health status of a patient can be modeled as the interlinking between various factors, such as the disease's attributes (including type, symptoms and anatomy), the patient's genetics (including pathways and biological processes), the kind of drug compounds, and the characterization of side effects. In this way, graph models can function as holistic representations of a patient's health status through the integration of different data sources into a single model.

In graph models, *representation learning can be applied to convert multivariate time series (MTSs)* *and static features into nodes of the graph* [49], for example, when multiple steps, such as various procedures in different time steps or disease stages, are converted into nodes in the whole graph.

Hence, starting from the disease stage at time point t_i, procedures such as a diagnosis can be performed, which introduces a new node into the KG. As clinical work is ideally performed within the frame of the guidelines, suitable medication will be proposed and applied. Any action through the performance of a medical procedure leads to an evolution of the patient's disease state, which at time point t_{i+1} can be monitored by the next procedure, again inducing new nodes in the KG. Relating medical procedures to specific patient characteristics (such as the genetic profile of the patient, side effects, etc.) enables a holistic view allowing guidelines to be tailored to the individual patient and to differ from institutional guidelines (such as in Use Case 2 of Table 2). Graphs are not limited to representing a single event but can also represent a whole process (i.e., the patient's journey). This is useful for digital twin implementation, which will be discussed in the following section.

In addition to patient-centered representations, graph models for disease-centered representations can also be constructed, as shown in Figure 3. This can be created from electronic patient records because hospital records contain rich information about the interrelationship of diagnoses (diagnoses and corresponding codiagnoses), the interrelationship of medical procedures (one medical intervention leads to another) and the interrelationship of diagnoses and procedures (due to the causal relationship between the diagnoses and the associated procedures). Thus, if a malignant neoplasm of the lungs (C34) is diagnosed, it is already known from earlier observations that it is often accompanied by chronic obstructive pulmonary disease (J44) and secondary malignant neoplasms of the brain (C79.3). Against the background of these relationships, four possible methods can be identified: computed tomography, bronchoscopy, magnetic resonance imaging and positron emission tomography.

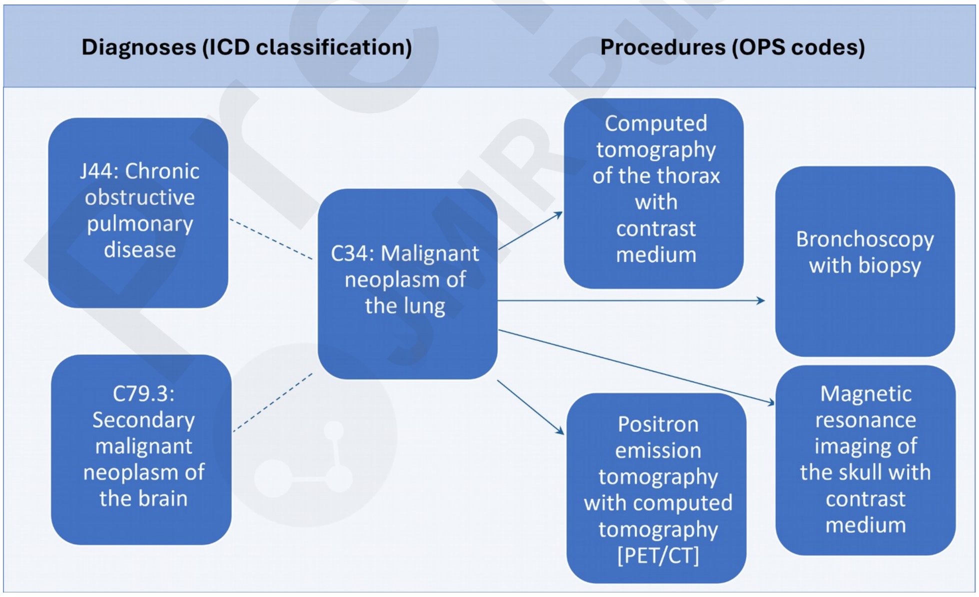

Figure 3 **Exemplary knowledge graph related to disease diagnoses and medical procedures.** In this example, we present the relationship between a malignant neoplasm of the lung (C34). We also present typical codiagnoses of this disease on the left (J44 and C79.3). On the right of the figure are corresponding medical procedures that integrate the main and codiagnoses (such as computed tomography, bronchoscopy, magnetic resonance imaging and positron emission tomography). This example has been extracted from guidelines for lung cancer in oncology *[50]*.

In Table 8, we present an overview of the application of graph models to model patient safety, procedure accuracy and procedure efficacy.

Table 8 List of relevant publications ordered by patient safety, procedure accuracy, and procedure efficacy.

Quality categories	Publication Details		
Patient Safety	Description - Publication	Model Target	Reference
	Prediction of Admission to Intensive Care Unit, where each node represents a patient with clinical data embeddings. The goal is to label a patient with either admission (1) or no admission (0) to the Intensive Care Unit (binary output).	Prediction Hospital Readmission	Recheteau et al. [51]
	Hybrid model combining Long Short-Term Memory networks (LSTMs) for extracting temporal features and Graph Neural Networks (GNNs) for extracting patient neighborhood information patient to perform outcome predictions, like In-Hospital Mortality (IHM, i.e., Failure to Rescue) and Length of ICU Stay (LOS).	Outcome Prediction: IHM, LOS	Tong. et al. [52]
	Model integrating image data and EHR in subnetworks. They have also developed a method to fuse outcomes from the different subnetworks, to analyze the probability of patient readmissions in cardiology.	Probability Patient readmission	Tang et al. [53]
	Graph Neural Network for patient centered modeling with the integration of relevant clinical information to identify potential 30-day Hospital Readmissions. In this work, they provided diagnoses of acute myocardial infarction of anterolateral wall (ICD-9 code: 410.0) and acute myocardial infarction of inferolateral wall (ICD-9 code: 410.2), which are grouped under the general family diagnosis acute myocardial infarction (ICD-9 code: 410).	Potential 30-Day Hospital readmission	Theodoropoulos et al. [54]
	method combining GNNs and deep learning to predict Hospital Readmission from nonstructured data with a method that extracts deep representations of clinical notes using a feature aggregation unit on top of a state-of-the-art Natural Language Processing (NLP) technique - Bidirectional Encoder Representations from Transformers (BERT). By exploiting these deep representations, a patient network is	Prediction Hospital Readmission	Golmaei et al. [55]

	built, and a Graph Neural Network (GNN) is trained to predict Hospital Readmissions.		
Procedure accuracy	Description - Publication	Model Target	Reference
	Graph model representing patients grouped by similar diagnoses [40]. In this model, patients with similar gender, age and diagnoses (coded with ICDs) are linked. This methodology allows the representation of a cohort by means of a graph. Based on this structure the authors assumed that procedure predictions are a form of patient classification task. Naturally, this model is limited in regard to performing predictions for single patients as it is excluding metabolomic and genomic data.	Procedure Prediction	Diaz Ochoa & Mustafa[40]
	G-BERT (Graph – BERT method) for medical code representation and medication recommendation (i.e., procedure prediction) [56]. In this work, the authors implemented GNNs to represent the internal hierarchical structures of medical codes. In this investigation the authors integrate the GNN representation into a transformer-based visit encoder and pretrain it on EHR data from only a single visit. Pretrained visit representations are then fine-tuned for downstream predictive tasks on longitudinal EHRs from patients with multiple visits.	Procedure Prediction	Shang et al. [56]
Procedure efficacy	**Description - Publication**	**Model Target**	**Reference**
	Geometric Graph Neural Networks (GGNN) are a well-suited method to incorporate geometric features, associated with genomic data, into deep learning for enhanced predictive power and interpretability for multiomics data. Zhu et al. implemented this method for the evaluation of patients with myeloma as well as other cancers from the Cancer Atlas to perform a prognosis prediction and in this way assist physicians in their decision-making process [57]. This implies the possibility of using such methods not only for patient monitoring but also as a basis for the evaluation of potential responses of patients to current	Prediction of patients with probability to develop Myeloma	Zhu et al. [57]

	or potential procedures.		
	This study introduces a Graph-based Multimodal Late Fusion (GMLF) deep learning framework that integrates histopathological images and transcriptomic data to predict therapeutic response in muscle-invasive bladder cancer. By leveraging a knowledge graph built from known biological interactions, the model guides the representation of molecular data in a biologically meaningful way. The approach enables cross-modal learning, improves predictive efficacy, and provides interpretable insights into relevant biomarkers. The integration of domain knowledge into the graph structure enhances the model's ability to support personalized treatment decisions in oncology.	Prediction of therapeutic response in muscle-invasive bladder cancer	Bai et al.[58]
	Graph neural network (GNN) model to predict immunotherapy response in skin melanoma. By identifying key pathways and constructing a gene signature called response Score, the model achieved high predictive accuracy and revealed strong associations with prognosis, genetic alterations, and the tumor microenvironment. Among the 13 contributing genes, one was identified as a potential therapeutic target due to its negative correlation with immune activity and poor prognosis. The findings support the clinical utility of GNNs in guiding immunotherapy and identifying biomarkers in melanoma.	Predict immunotherapy response in skin melanoma	Ye et al. [59]

4.2 LIMITATIONS OF KGs

As the graph grows in size, querying and updating can become increasingly resource intensive, potentially impacting the model’s performance. In addition, algorithms are often based on the message-passing paradigm, which consists of node-by-node iteration (where each node aggregates information from neighboring nodes), which has significant limitations for graphs consisting of highly connected nodes [60]. For this reason, graphs with explicit heterogeneous edge distributions pose a problem and are difficult to evaluate because of the excessive compression of exponentially growing information in fixed-size vectors [61]. Recently, methods based on a different approach than the node-centered approach, which is based on differential geometry, have shown promising results for overcoming graph bottlenecks [62].

KG often suffers from incomplete data. If the graph does not have all the relevant entities and relationships, the model may be missing important information. Incorrect or outdated information can lead to erroneous insights and conclusions. This problem can be solved by defining aggregation functions to derive partial graph folding and compensate for incomplete node categories [63]. Finally, keeping the graph updated with new knowledge and ensuring consistency across different sources can be challenging [64].

4.3 Health Digital Twins (HDTs) and Health Digital Shadows (HDSs) for Modeling Procedure Efficacy

To model procedure efficacy, it is necessary to analyze the system's (i.e., patient's) response to an external perturbation (e.g., administration of a drug). This can be accomplished by **health digital twins**[65] **(HDTs)** and **health digital shadows (HDSs)**, or "digital traces" in medicine; we keep the definition of HDSs in this article. With respect to this definition, however, we will be careful: while industrial objects can be fully digitalized, patients are too complex and cannot be objectivized. As a result, we prefer to use the term "shadow" instead of "Twin". For this reason, in this section, we refer to **health digital shadows (HDSs)**.

According to Case 4 of Table 2 and the quality metrics summarized in Table 3, it is crucial to gain deep insights into a patient's conditions to assess the efficacy of the procedures.

The targets are not only binary values indicating patient outcomes or medical procedures but also biomarkers recorded as time series. For example, it is currently not possible to predict in advance whether a patient will face an allergic reaction or intolerance to medication. However, this is relevant, as these reactions can be life-threatening and lead to treatment delays or morbidity.

Another problem is that physicians lack the time to review all available medical records of a patient. In particular, in the case of complex diseases such as cancer with various genetic mutations, which can also change during the procedure, prior or regular modeling is needed. In this way, procedure adaptations or medication adjustments can be made in cases of side effects or low procedure efficacy.

For example, genetic polymorphisms in drug metabolism are already known for some substances, which must be determined before the substance is administered (e.g., dihydropyrimidine dehydrogenase testing before 5-fluorouracil administration). Possible drug interactions must also be considered if several diseases are present. Many patients take several medications, which sometimes lead to unforeseeable interactions. For patients, these interactions are increasingly difficult to oversee and evaluate, and individual factors also play a role in tolerability. The combination of different target values, along with different time spans, makes the problem definition much more complex than for patient safety and procedure accuracy.

Hence, an algorithmic architecture capable of considering a wide range of patient information and potential interactions would help model patient responses to medical procedures and predict likely patient outcomes. In this way, we model the integration of different procedures within a process. Technically speaking, this task represents the need for the integration of various models along a process (i.e., patient's journey). In this context, the concept of a DT, which originated from engineering sciences [66], can be adapted into a clinical context.

The aim of a DT is to generate a digital representation of an existing object within a process or time evolution in a digital space. Its concept is designed to combine time-evolving information from real objects, such as data generated from the Internet of Things, and mathematical modeling in the digital space (Richter et al. 2023). Therefore, this digital representation allows the adaptation or optimization of processes in a digital space before transformations or modifications of real processes are made. For example, by integrating different and disparate information sources, it is possible to create different models not only for industrial production but also for the solution of critical and strategic problems such as pollution by generating universal avatars [67].

Thus, just as KGs aim at integrating different information sources in a medical framework, as shown in Figure 4 (4), the concept of a DT aims at the integration of various models. The main characteristics of this algorithmic architecture are as follows [68]:

1. **Enabling bidirectional communication between the model and real assets.**
2. **Different information sources**, such as genetic information, comorbidities, medication history, medical history, and family history, should be integrated.
3. **Integration of different models,** such as mechanistic models (e.g., PBPK models [69]) or inductive models (based on ML), at one time point/for one procedure.
4. **Couple individual procedure models within a process,** for instance, in a patient journey.
5. **Simulation of procedures over time and process**.
6. **Model improvement** through learning from the digital cohort (sum of all DTs).

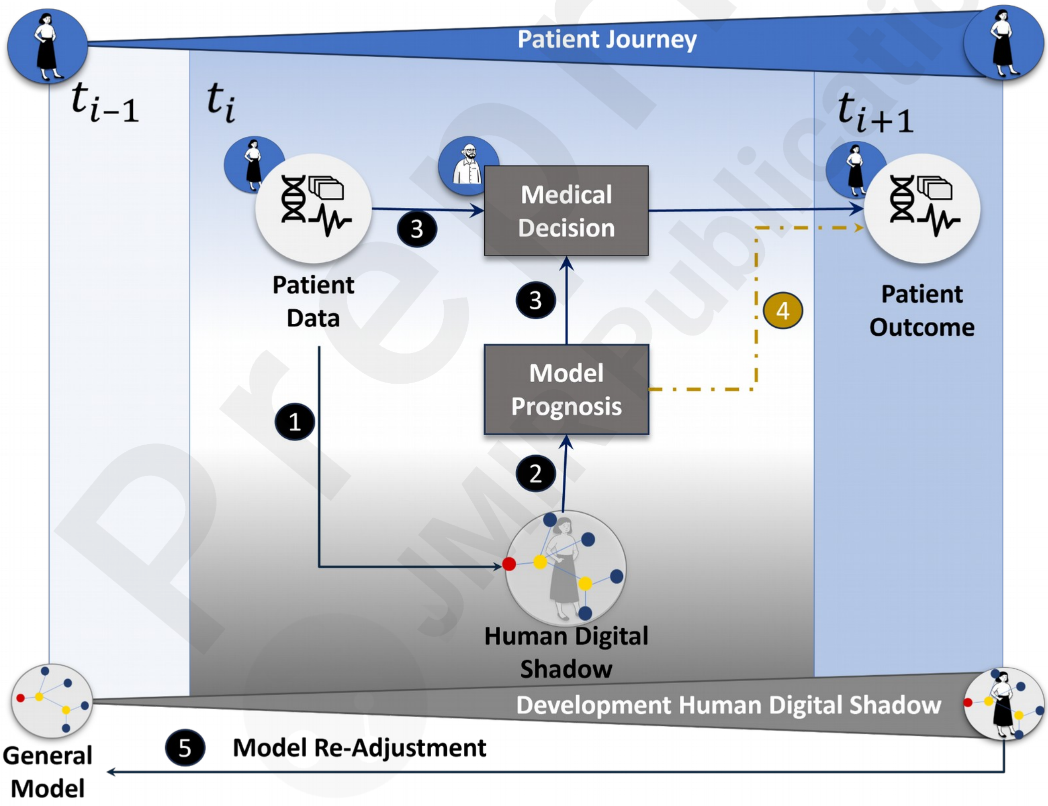

Figure 4 **Representation of a human HDS as an information system.** The HDS integrates patient data into a generally derived mathematical model (1) and allows model prognosis (2). Thus, the model supports the work of medical personnel as additional information for medical decision making (3), which leads to a patient outcome at the next time point. In this way, bidirectional communication between the HDS and the real world is enabled. The model is also able to simulate the patient's journey, although the model prognosis can be used to directly estimate the patient's outcome (4). The information gained by the individual human HDS can be used to readjust the general model constructed by the digital cohort (5).

Furthermore, the connection between KGs and HDSs represents a logical step in the way different and disparate information sources are connected to create descriptive and predictive models. As shown in Figure 5, relevant patient information as well as medical guidelines (in the form of KGs) can be integrated into a full graph containing patient data (for example, genetic or socioeconomic data) as well as additional relevant information (for example, drug structure and drug interactions).

In its role as an information and model integrating method, HDSs are well suited for [70]:

- **Generic representations:** The predicted value of a quantity of interest is within the expected ranges of the global population. The representations are on an abstract and qualitative scale.
- **Population-specific representations:** The predicted value of a quantity of interest is evaluated with respect to a patient cohort. Thus, such representations are more precise and are quantitative predictions.
- **Specific representations:** The predicted values of the quantity of interest have high accuracy in relation to the expected values for single patients. These models are used for personalized predictions.

The HDS concept is interesting for prospective analysis, i.e., in the simulation and prediction of a patient's health status and response to procedures. For prognostic purposes, they include the ability to integrate all patient information available. In this way, doctors treating patients can verify in advance the planned treatment regimen, such as medication or radiotherapy, on the patient's digital representation.

The HDS model runs through all the interactions between the procedure and the patient's organism. In the case of pharmacological treatment, medication interactions are simulated (for example, by integrating ML models for drug–drug interactions [71]). Then, an individual recommendation for the substance and individual dosage and medication could be obtained by applying appropriate mathematical models (i.e., ML). Additionally, if the treatment response is not sufficient, HDS could propose additional procedures or further treatment strategies.

Current databases for potential drug targets could also be considered. All these steps are included in a close information loop. Bidirectional communication between the real world and the model is enabled. Therefore, the general model obtains patient-specific information at time point t_i (1).

This enables the human HDS to make a prognosis (2), which can prospectively be used by the clinician to adapt his medical decision (3). Alternatively, model prediction can be used retrospectively to evaluate medical decisions by examining patient outcomes at time points t_{i+1} (4). As time evolves, the patient's journey produces increasing patient-specific information, leading to better possible adaptation of the general model to a personalized human HDS. As the cohort of HDSs grows, model readjustments for the general model are performed after certain time periods (5).

Thus, HDS could increase the safety of drug therapy and its efficacy by providing personalized recommendations on the basis of all available information [72]. For example, a similar concept, which is based on whole-body representation via PBPK models (physiologically based pharmacokinetic modeling) and the integration of multiagent systems representing hepatocyte metabolism, can be used for pharmacology, particularly for the representation of the avatar's response to a medication [73], [74]. This example demonstrates that through the analysis of the avatar's response, it is possible to plan procedures and even consider individual patient characteristics (for example, personalized metabolite levels that depend on the patient's genetics) that lead to personalized patient care. This example implies that procedures can be tested on the DHS for forecasting procedure success; therefore, the recommendation level of a procedure can be assessed before being applied to a real patient. Such an application is equivalent to the so-called dynamic HDT, where forecasts based on time series are performed [75].

Thus, HDS can be useful for modeling procedure efficacy, as in the following examples [76]:

- Human HDS for inpatients:
  - To analyze which combination of medical procedures is most appropriate for patients depending on their socioeconomic context, environment, genetics, etc.
  - Analyze whether medical decisions will have an effective and desired outcome.
- Human HDS for outpatients:
  - Physicians can help physicians recognize dangerous changes in a patient's health state through the analysis of patient data that can be delivered in real time (wearables or machines, for instance, in nephrology).

HDS are increasingly being used for the estimation of the quality of patient care [66]. To this end, the prediction of patient outcomes is relevant, for example, in a personalized patient assessment to decide whether a procedure with high risk (death, side effects, patient burdens, etc.) is needed. This prevents the deterioration of a patient's quality of life. Figure 5 shows an exemplary representation of a human HDS that integrates different data sources (such as omics, EHR and documentation), which are measured in different time periods along the patient journey, into different models to predict patient outcomes [75]. In Figure 5, a medical decision for the performance of a procedure according to medical guidelines is made (1). A whole-body (PBPK) model (2) can be used to integrate patient-specific data within the HDT to predict a patient's pharmacological response. Using pharmacological response prediction and patient data, another model can estimate patient outcomes, e.g., through cluster analysis, for survival time prediction (3). In this way, simulation along the patient journey is enabled.

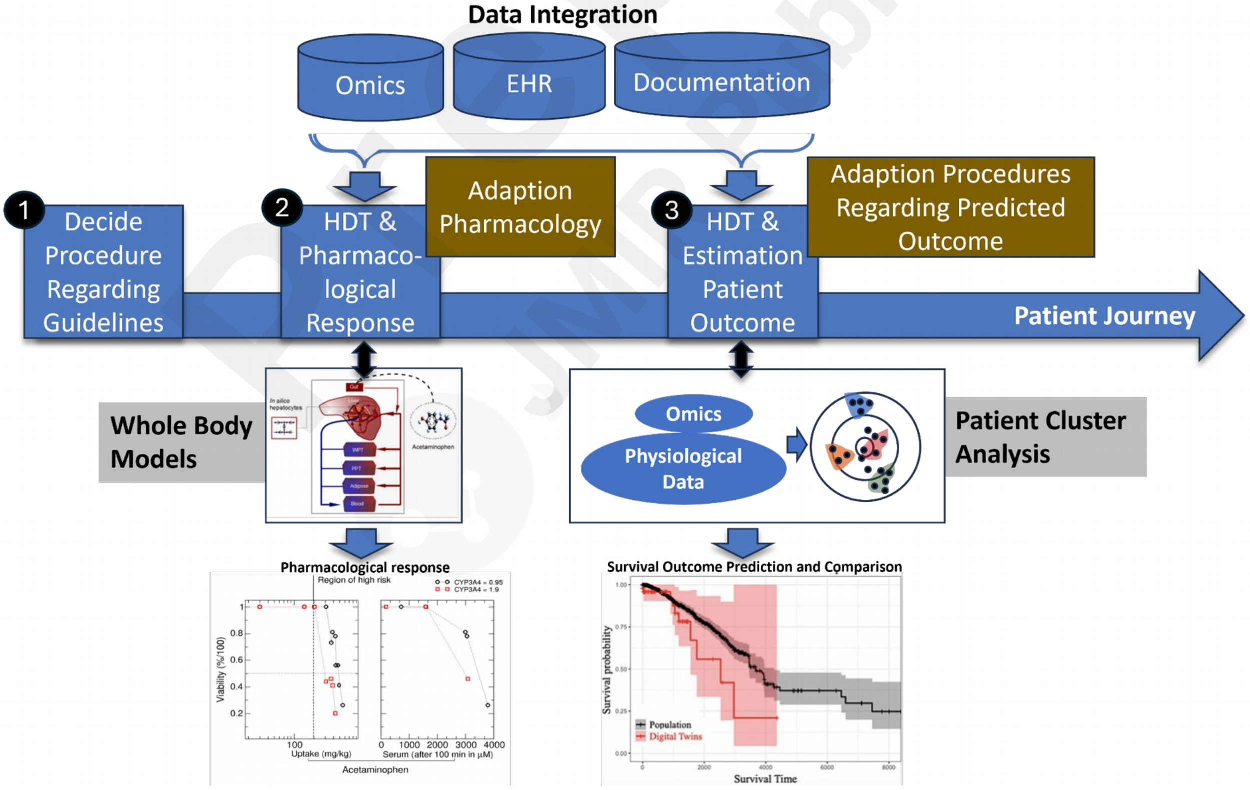

Figure 5 **Schematic representation of a human HDS**. A human HDS is not a single model but a collection of different models to perform different predictions depending on the system's parameters of interest. Such predictions can include the analysis and prediction of time series at different time points of the patient's journey. In this figure, we present a schema of a patient journey where after a medical decision, (1) different information sources are integrated to predict quality metrics in different time periods depending on the patient's evolution. For example, the pharmacological response can be estimated via a whole-body model, e.g., the PBPK model (2) *[77]*, and the patient outcome can be estimated via cluster analysis (3) *[78]*.

Chang et al. presented a remarkable example of possible DHS realization through the creation of a pipeline for predicting patient outcomes, particularly survival heterogeneity, for patients with breast cancer. In this implementation, three different data sources were integrated, where low-dimensional embeddings of clinical and molecular features enriched by an external annotation database were the main data sources. In this algorithm, a dimensional reduction method, UMAP, is implemented.

This is a general-purpose manifold learning and dimension reduction algorithm to reduce the data space dimension. A dynamic HDS requires a pipeline that is accordingly adapted, i.e., new data and additional models must be integrated and appropriately adapted according to other outcomes [78].

Digital twins often oversimplify real systems. These approximations can lead to discrepancies between the model and the actual system. The accuracy of the digital twin depends on the quality and granularity of the data used. Since more than one model is coupled to generate a full digital twin, different models trained with different data and different granularities can lead to unbalanced representations. Finally, poor or sparse data can lead to unreliable models [79]. Notably, all these problems are exacerbated by issues related to the aggregation and integration of different types of data that occur in healthcare, along with the strict privacy policies that must be followed [80], a problem that may be alleviated with the introduction of electronic health records [81].

Creating and maintaining a DT can be computationally intensive, requiring significant resources for simulation and real-time data processing. Integrating a DT with other systems, sensors, and data sources can be complex and may require sophisticated interfaces and protocols. Ensuring that the DT reflects real-time changes in the physical system can be challenging, especially in systems with high dynamics [82].

An interesting option is the relation and validation of DTs with in vitro methods such as organs on chips. This option opens the possibility to compensate for potential defects in model predictions and identify their potential problems by validating and comparing results with advanced in vitro patient representations [83]. This, however, is still a challenge in the field of oncology and requires more research to obtain a correct representation of patient outcomes.

5 Issues and future research directions

Our analysis revealed that patient safety, the accuracy of the procedure and the effectiveness of the procedure are paramount. Using the present taxonomy, we can define the types of challenges and the type of ML implementation to be performed. However, this taxonomy is far from perfect. For example, distinguishing between a diagnostic procedure and a therapeutic procedure may require a more specialized and precise taxonomy. In addition, the metrics required to evaluate patient outcomes and, in this way, to estimate the QI may not always be perfectly calibrated [84], which can limit the applicability of such methods. In general, patient outcomes are still a much-debated subject [85]. Despite this, we believe that it is important to keep any classification and taxonomy as simple as possible to help facilitate the implementation of mathematical models.

In our analysis, we concluded that holistic quality management analysis (especially regarding procedure efficacy) requires the integration of several data sources. This is because there is no single decision (in decision support) or event (assignment of a diagnosis), but whole processes impact the patient's health status. Thus, quality management and assessment can be very challenging for model implementation since it requires the integration of a constant information flow, which can continuously change within the process. As a possible solution to address these issues, we have introduced the concept of the human HDS.

It is also relevant to consider guidelines that serve as guidance to define and implement processes necessary for the management of patients' health status. They are the result of clinical studies and are intended to help doctors make decisions. However, they contain complex workflows that should be evaluated individually. Thus, while a deviation from a guideline may be a potential problem in disease management for one patient because a step in the workflow has not been taken into account (which is a reduction in the quality of the hospital, i.e., a decrease in procedural accuracy), for other patients, this deviation can be seen as a necessary adjustment needed for more personalized healthcare (i.e., increasing the effectiveness of the procedure). In this context, especially in the evaluation of personalized practices, ML methods can influence the way guidelines are applied.

We are, however, cautioned about full-model-assisted personalization as well. Some critics have reported that data integration and target modeling have the potential to drive away from real personalized patient centers and more human medicine [86]. In this case, individual modeling is just a way to quantitatively evaluate guidelines and their potential effect on the expected quality management of the disease at the patient level.

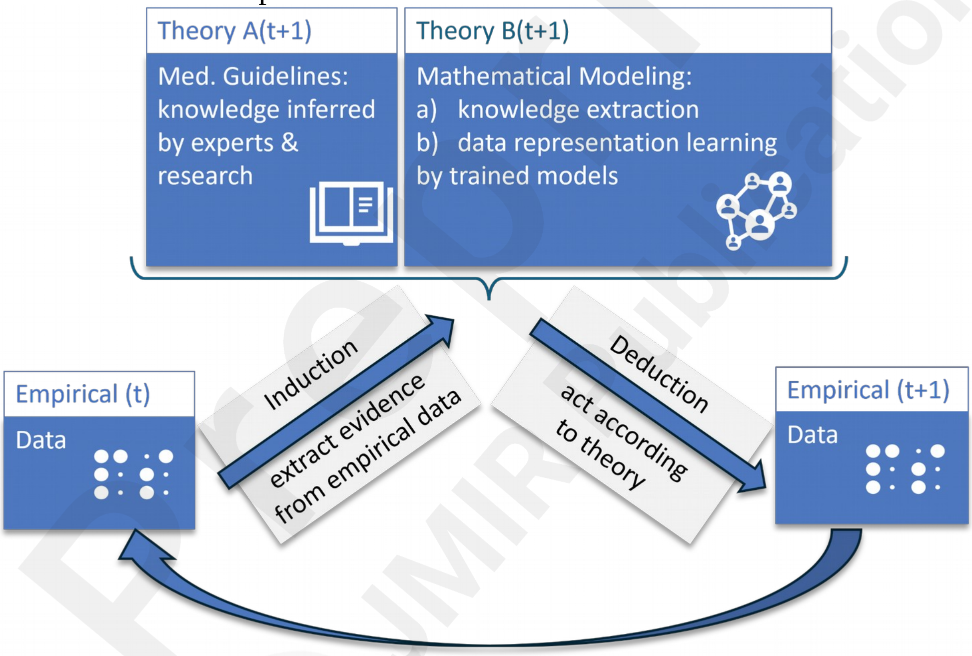

Figure 6 **Lifecycle of the data and models in relation to the clinical guidelines.** Concept induction, as the extraction of evidence from empirical data (t) and deduction, is the generation of new empirical data (t+1) by action according to a given general theory. The theory is, on the one hand, given by the medical guidelines – Theory A(t+1) and, on the other hand, by the mathematical models trained on the given empirical data – Theory B(t+1).

Figure 6 illustrates the context in which we see the most appropriate use and application of ML model support systems. We understand mathematical modeling as a subsection of medical guidelines. Therefore, in Figure 6, we now want to visualize the role of mathematical modeling in evidence generation and thus in improving health care. Figure 6 shows how, from time to time, empirical observations stored as data are evaluated to develop and adopt improvements to medical guidelines adopted by health professionals.

In the present work, we reviewed some works where KGs and GNNs have been successfully applied to problems closely related to quality management. One shortcoming is that almost all reviewed methods apply to inpatient outcomes, and few models explore procedure efficacy, which is often based on a multilabel classification. This is because multilabel problems are difficult to implement and achieve a high level of accuracy.

Graph integration is an appealing idea given that graph models are essentially an integration of data and information. Furthermore, when time is considered, it is possible to generate a digital representation of the patient, i.e., an HDS. Furthermore, this concept has considerable potential for both inpatients and outpatients. This assumes that data generated by portable devices can track the patient's condition at home. Furthermore, this concept can be coupled with digitally stored patient journeys, i.e., the digital cohort [87].

Currently, HDS essentially involves the integration of different models within a process. We believe that in the future, the application of KGs will contribute to the rapid development of HDS. For example, recently, some authors have suggested that the combination of KGs and LNNs could be the key to the implementation of productive avatars in medicine [88].

Additionally, novel decentralized data storage concepts may help train and define accurate models with sufficient data volumes. For example, by integrating different and disparate information sources, some of which are interconnected through the semantic web (according to the Worldwide Web 3.0 standards), it is possible to create models not only for industrial production but also for the solution of critical and strategic problems by generating universal avatars [67].

From this perspective, solutions built on the FAIR principle (Findability, Accessibility, Interoperability, and Reusability [89]) to store data can be helpful for extracting data from patients and citizens in real environments and not just data in controlled environments such as clinical data. On the basis of these principles, protocols based, for example, on semantic webs, such as Solid [90], could help accelerate the implementation of fully decentralized or semidecentralized data gathering from individuals. Several recent initiatives are developing such ideas to implement digital twins on the basis of semantic web principles (see, for instance, the EDITH-CSA project [70]). Remarkable in such initiatives is not only the definition and exploration of models but also the definition of data standards required for data interoperability, most of which are based on fast healthcare interoperability resources (FHIR) standards.

However, HDS faces other challenges. In the event of poor data safety standards, criminals could gain access to medical records and HDS for blackmailing or ransom demands. Companies such as health insurance or disability insurance might also demand the patient's HDS prognosis before enrollment. On the basis of the HDS, patients could be denied insurance, insurance rates could be increased, or existing insurance conditions could be adapted regularly on the basis of the HDS. This could ultimately lead to a selection of people who will not receive insurance at all or who will receive only reduced coverage. Institutions and employers could also demand an HDS model before hiring people to select healthy people.

Another relevant issue is data contamination by ML models. For example, if a machine catches early signs of sepsis and doctors treat it, “this creates a ‘contaminated association’ in the data. Furthermore, an intervention triggered by one model can quietly disrupt another, even if they are focused on entirely different outcomes[91]. This can be particularly critical for HDTs. Here, a potential solution is to preserve clean data from datasets where predictive models are not used.

The level of mandatory treatment recommendations based on HDS must be discussed and assessed. Patient wishes are crucial. What happens if the physician or patient deviates from HDS's recommendations? In this case, the insurance company may not cover these alternative treatments since HDS does not predict them to be most effective. This question is related to the overall reliability of the HDS. In retrospect, the accuracy of the HDS recommendation can be assessed, e.g., by checking procedure efficacy and tolerability.

6 Conclusion

The integration of KGs and DTs into healthcare holds transformative potential for improving clinical outcomes and the quality of care. The following are the key conclusions of this study:

- Patient outcomes are not abstract quality terms, since they can be formalized and linked to measurable clinical processes:
- Medical guidelines are essential but often underutilized in quality monitoring
- Patient safety can be quantified and predicted
- Real-world clinical decisions are complex and require personalization
- Knowledge graphs and GNNs help bridge fragmented clinical data
- Health digital shadows (HDSs) offer a vision for prospective, patient-specific quality evaluation
- Quality management should shift from static checklists to dynamic, data-driven guidance

However, this optimistic view is feasible only when the technologies are implemented in an ethical way, avoiding blind customer overconfidence in the delivered results. Furthermore, implementation requires accounting from the stakeholders' perspective: what is considered a desirable outcome from a patient-centered perspective (e.g., symptom relief, preserved quality of life) may diverge from institutional or systemic goals such as shorter hospital stays or cost efficiency. As such, mathematical models for healthcare quality must clearly define their optimization target—whether the primary focus lies in maximizing individual patient well-being, the operational performance of the healthcare system, or economic sustainability. These dimensions are not mutually exclusive, but the prioritization among them fundamentally shapes the modeling framework and interpretation of safety-related events.

The combination of patient outcomes, procedures and guidelines is not only interesting from a technical point of view but also extremely relevant for improving the quality of hospitals and health centers. The integration of guidelines implies the application of ML in different ways, from alerting and predicting critical conditions to the application of similar methods employed in fraud detection to recognize potential deviations from guidelines.

The topic of quality management in medicine is expected to become as relevant as diagnosis or image recognition are currently relevant, and we expect an increasing interest in increasing ML applications for quality management in the next few years.

Acknowledgments

We would like to thank Matthias Weidemüller for insightful and inspiring discussions, as well as for the invaluable support provided throughout this work. JGDO would like to thank Faizan E Mustafa for inspiring discussions and support in the research about graph knowledge and Elena Ramirez for the relationship of medical decisions in the framework of the theory of rational decision making. JGDO is thankful for Felix Weil's constant support.

Funding

This project was supported by the Ministry for Economics, Labor and Tourism from Baden-Württemberg, Germany via grant agreement number BW1_1456 (AI4MedCode).

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2181/1 - 390900948 (the Heidelberg STRUCTURES Excellence Cluster).

This project was also supported by broad funding to the Klinikum Stuttgart, Stuttgart Cancer Center – Tumorzentrum Eva Mayr-Stihl, from the Eva Mayr-Stihl Foundation, Waiblingen, Germany.

Contributions

The first idea was proposed and drafted by AN and JGDO. JGDO contributed to the overview of the knowledge graphs. SN, AN, and JGDO contributed to the digital twin overview. SN and MK contributed to the medical background, examples, and critical evaluation of the various methods described in this article. All the authors contributed to the final version of this article.

Declaration, use of generative AI

The authors declare that no generative artificial intelligence has been used for data analysis, figure generation, or text composition. The texts are all generated by the authors. The text was scanned with Rubiq [https://www.aje.com/rubriq/] to detect spelling and typographical errors.

List of Abbreviations

Abbreviation	Meaning
BERT	Bidirectional Encoder Representations from Transformers
DL	Deep Learning
DT	Digital Twin
DS	Digital Shadow
EHR	Electronic Health Record
GNN	Graph Neural Network
HDT	Health Digital Twin
HPC	Health Care Practitioners
ICD	International Disease Classification
ICU	Intensive Care Unity
IQI	Inpatient Quality Indicator
IMC	Intermediate Medical Care
KG	Knowledge Graph
HDS	Health Digital Shadow
LOS	Length of ICU stay
LIMS	Laboratory Information Management Systems
MD	Medical Decision
ML	Machine Learning
MP	Medical Path
MTS	Multivariate Time Series
MI	Medical Intervention
PBPK	Physiologically Based Pharmacokinetic Modeling
QM	Quality Metrics
QI	Quality indicator

Bibliography

[1] S. McLachlan, E. Kyrimi, K. Dube, G. Hitman, J. Simmonds, and N. Fenton, “Towards standardisation of evidence-based clinical care process specifications,” *Health Informatics J.*, vol. 26, no. 4, pp. 2512–2537, Dec. 2020, doi: 10.1177/1460458220906069.
[2] S. Timmermans and A. Mauck, “The promises and pitfalls of evidence-based medicine,” *Health Aff. Proj. Hope*, vol. 24, no. 1, pp. 18–28, 2005, doi: 10.1377/hlthaff.24.1.18.
[3] C. A. Sinsky, H. Bavafa, R. G. Roberts, and J. W. Beasley, “Standardization vs Customization: Finding the Right Balance,” *Ann. Fam. Med.*, vol. 19, no. 2, pp. 171–177, 2021, doi: 10.1370/afm.2654.
[4] K. Alexopoulos and G. Chryssolouris, “Process,” in *CIRP Encyclopedia of Production Engineering*, S. Chatti, L. Laperrière, G. Reinhart, and T. Tolio, Eds., Berlin, Heidelberg: Springer, 2019, pp. 1349–1352. doi: 10.1007/978-3-662-53120-4_6567.
[5] N. Da Silva, “The Industrialization of Healthcare and Its Critiques,” in *Handbook of Economics and Sociology of Conventions*, R. Diaz-Bone and G. de Larquier, Eds., Cham: Springer International Publishing, 2020, pp. 1–25. doi: 10.1007/978-3-030-52130-1_45-2.
[6] S. Dodwad, “Quality management in healthcare,” *Indian J. Public Health*, vol. 57, pp. 138–143, Oct. 2013, doi: 10.4103/0019-557X.119814.
[7] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro, “Process mining in healthcare: A literature review,” *J. Biomed. Inform.*, vol. 61, pp. 224–236, Jun. 2016, doi: 10.1016/j.jbi.2016.04.007.
[8] M. May, “Eight ways machine learning is assisting medicine,” *Nat. Med.*, vol. 27, no. 1, pp. 2–3, Jan. 2021, doi: 10.1038/s41591-020-01197-2.
[9] J.-F. Bach, “Causality in medicine,” *C. R. Biol.*, vol. 342, no. 3, pp. 55–57, Mar. 2019, doi: 10.1016/j.crvi.2019.03.001.
[10] J. G. Diaz Ochoa, “A unified method for assessing the observability of dynamic complex systems,” *Comput. Biol. Med.*, vol. 160, p. 107012, Jun. 2023, doi: 10.1016/j.compbiomed.2023.107012.
[11] UNO, “Department of Economic and Social Affairs Sustainable Development,” Ensure healthy lives and promote well-being for all at all ages. [Online]. Available: https://sdgs.un.org/goals/goal3
[12] K. Steele and H. O. Stefánsson, “Decision Theory,” in *The Stanford Encyclopedia of Philosophy*, Winter 2020., E. N. Zalta, Ed., Metaphysics Research Lab, Stanford University, 2020. Accessed: Mar. 12, 2024. [Online]. Available: https://plato.stanford.edu/archives/win2020/entries/decision-theory/
[13] R. M. Epstein and R. L. Street, “The Values and Value of Patient-Centered Care,” *Ann. Fam. Med.*, vol. 9, no. 2, pp. 100–103, Mar. 2011, doi: 10.1370/afm.1239.
[14] C. Kersting, M. Kneer, and A. Barzel, “Patient-relevant outcomes: what are we talking about? A scoping review to improve conceptual clarity,” *BMC Health Serv. Res.*, vol. 20, p. 596, Jun. 2020, doi: 10.1186/s12913-020-05442-9.
[15] Y. Liu, K. C. Avant, Y. Aungsuroch, X.-Y. Zhang, and P. Jiang, “Patient outcomes in the field of nursing: A concept analysis,” *Int. J. Nurs. Sci.*, vol. 1, no. 1, pp. 69–74, Mar. 2014, doi: 10.1016/j.ijnss.2014.02.006.
[16] S. A. Hadian, R. Rezayatmand, N. Shaarbafchizadeh, S. Ketabi, and A. R. Pourghaderi, “Hospital performance evaluation indicators: a scoping review,” *BMC Health Serv. Res.*, vol. 24, no. 1, Art. no. 1, Dec. 2024, doi: 10.1186/s12913-024-10940-1.
[17] X. S. Wang and V. Gottumukkala, “Patient-reported outcomes: Is this the missing link in patient-centered perioperative care?,” *Best Pract. Res. Clin. Anaesthesiol.*, vol. 35, no. 4, pp.

565–573, Dec. 2021, doi: 10.1016/j.bpa.2020.10.006.
[18] R. Busse, N. Klazinga, D. Panteli, and W. Quentin, Eds., *Improving healthcare quality in Europe: Characteristics, effectiveness and implementation of different strategies*. in European Observatory Health Policy Series. Copenhagen (Denmark): European Observatory on Health Systems and Policies, 2019. Accessed: Mar. 06, 2024. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK549276/
[19] Chloe Garnham, “Patient outcomes: overview, measures & ways to improve.” [Online]. Available: https://dovetail.com/patient-experience/what-are-patient-outcomes/
[20] C. McLeod, R. Norman, E. Litton, B. R. Saville, S. Webb, and T. L. Snelling, “Choosing primary endpoints for clinical trials of health care interventions,” *Contemp. Clin. Trials Commun.*, vol. 16, p. 100486, Dec. 2019, doi: 10.1016/j.conctc.2019.100486.
[21] S. Yanamadala, D. Morrison, C. Curtin, K. McDonald, and T. Hernandez-Boussard, “Electronic Health Records and Quality of Care,” *Medicine (Baltimore)*, vol. 95, no. 19, p. e3332, May 2016, doi: 10.1097/MD.0000000000003332.
[22] “Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V.” [Online]. Available: https://www.awmf.org/
[23] T. Langer and M. Follmann, “Das Leitlinienprogramm Onkologie (OL): Nukleus einer evidenzbasierten, patientenorientieren, interdisziplinären Onkologie?,” *Z. Für Evidenz Fortbild. Qual. Im Gesundheitswesen*, vol. 109, no. 6, pp. 437–444, Jan. 2015, doi: 10.1016/j.zefq.2015.09.007.
[24] Jens Schick, “IQM Positionspapier Ergebnisqualität.” [Online]. Available: https://www.initiative-qualitaetsmedizin.de/positionspapier
[25] D. Ä. G. Ärzteblatt Redaktion Deutsches, “Initial Cancer Treatment in Certified Versus Non-Certified Hospitals (29.09.2023),” Deutsches Ärzteblatt. Accessed: Mar. 06, 2024. [Online]. Available: https://www.aerzteblatt.de/int/archive/article?id=233578
[26] N. Segnan and A. Ponti, “Artificial intelligence for breast cancer screening: breathtaking results and a word of caution,” *Lancet Oncol.*, vol. 24, no. 8, pp. 830–832, Aug. 2023, doi: 10.1016/S1470-2045(23)00336-4.
[27] J. Rudnitckaia, H. S. Venkatachalam, R. Essmann, T. Hruška, and A. W. Colombo, “Screening Process Mining and Value Stream Techniques on Industrial Manufacturing Processes: Process Modelling and Bottleneck Analysis,” *IEEE Access*, vol. 10, pp. 24203–24214, 2022, doi: 10.1109/ACCESS.2022.3152211.
[28] E. Katsoulakis *et al.*, “Digital twins for health: a scoping review,” *Npj Digit. Med.*, vol. 7, no. 1, p. 77, Mar. 2024, doi: 10.1038/s41746-024-01073-0.
[29] G. Gigerenzer and J. A. M. Gray, *Bessere Ärzte, bessere Patienten, bessere Medizin. Aufbruch in ein transparentes Gesundheitswesen: Mit einem Vorwort von Günther Jonitz. Strüngmann Forum Reports*, 1st ed. Berlin: MWV Medizinisch Wissenschaftliche Verlagsgesellschaft, 2013.
[30] J. Peng, E. C. Jury, P. Dönnes, and C. Ciurtin, “Machine Learning Techniques for Personalised Medicine Approaches in Immune-Mediated Chronic Inflammatory Diseases: Applications and Challenges,” *Front. Pharmacol.*, vol. 12, Sep. 2021, doi: 10.3389/fphar.2021.720694.
[31] L. Emanuel *et al.*, “What Exactly Is Patient Safety?,” in *Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 1: Assessment)*, K. Henriksen, J. B. Battles, M. A. Keyes, and M. L. Grady, Eds., in Advances in Patient Safety. , Rockville (MD): Agency for Healthcare Research and Quality, 2008. Accessed: Mar. 06, 2024. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK43629/
[32] U. Nimptsch and T. Mansky, *G-IQI – German Inpatient Quality Indicators Version 5.4*. 2022. Accessed: Aug. 02, 2023. [Online]. Available: https://depositonce.tu-berlin.de/handle/11303/17090
[33] “Agency for Healthcare Research and Quality.” [Online]. Available: https://www.ahrq.gov/
[34] B. Walliser, Ed., “Decision-making as reasoning,” in *Cognitive Economics*, Berlin, Heidelberg:

Springer, 2008, pp. 49–67. doi: 10.1007/978-3-540-71347-0_4.
[35] M. Lafonte, J. Cai, and M. E. Lissauer, “Failure to rescue in the surgical patient: a review,” *Curr. Opin. Crit. Care*, vol. 25, no. 6, pp. 706–711, Dec. 2019, doi: 10.1097/MCC.0000000000000667.
[36] R. Greenes, *Clinical Decision Support: The Road to Broad Adoption*, 2nd ed. Amsterdam: Academic Press, 2014.
[37] J. H. Barth *et al.*, “Why are clinical practice guidelines not followed?,” *Clin. Chem. Lab. Med.*, vol. 54, no. 7, pp. 1133–1139, Jul. 2016, doi: 10.1515/cclm-2015-0871.
[38] P. Börm, “Leitlinienbasierter Clinical Decision Support – Anforderungen an evidenzbasierte Entscheidungsunterstützungssysteme,” *OP-J.*, vol. 37, no. 01, pp. 28–35, Apr. 2021, doi: 10.1055/a-1284-3193.
[39] “AWMF Onlie, Prostatakarzinom.” [Online]. Available: https://register.awmf.org/de/leitlinien/detail/043-022OL
[40] J. G. Diaz Ochoa and F. E. Mustafa, “Graph neural network modelling as a potentially effective method for predicting and analyzing procedures based on patients’ diagnoses,” *Artif. Intell. Med.*, vol. 131, p. 102359, Sep. 2022, doi: 10.1016/j.artmed.2022.102359.
[41] M. Becker, S. Kasper, B. Böckmann, K.-H. Jöckel, and I. Virchow, “Natural language processing of German clinical colorectal cancer notes for guideline-based treatment evaluation,” *Int. J. Med. Inf.*, vol. 127, pp. 141–146, Jul. 2019, doi: 10.1016/j.ijmedinf.2019.04.022.
[42] N. Alharbe, M. A. Rakrouki, and A. Aljohani, “A Healthcare Quality Assessment Model Based on Outlier Detection Algorithm,” *Processes*, vol. 10, no. 6, Art. no. 6, Jun. 2022, doi: 10.3390/pr10061199.
[43] S. A. Alowais *et al.*, “Revolutionizing healthcare: the role of artificial intelligence in clinical practice,” *BMC Med. Educ.*, vol. 23, p. 689, Sep. 2023, doi: 10.1186/s12909-023-04698-z.
[44] A. Zaiss and H.-P. Dauben, “ICHI – International Classification of Health Interventions,” *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, vol. 61, no. 7, pp. 778–786, Jul. 2018, doi: 10.1007/s00103-018-2747-6.
[45] S. S. Lynch, “Drug Efficacy and Safety.” 2022. [Online]. Available: https://www.msdmanuals.com/professional/clinical-pharmacology/concepts-in-pharmacotherapy/drug-efficacy-and-safety
[46] A. Kline *et al.*, “Multimodal machine learning in precision health: A scoping review,” *Npj Digit. Med.*, vol. 5, no. 1, p. 171, Nov. 2022, doi: 10.1038/s41746-022-00712-8.
[47] Y. Si *et al.*, “Deep representation learning of patient data from Electronic Health Records (EHR): A systematic review,” *J. Biomed. Inform.*, vol. 115, p. 103671, Mar. 2021, doi: 10.1016/j.jbi.2020.103671.
[48] M. Liu, F. Zhang, P. Huang, S. Niu, F. Ma, and J. Zhang, “Learning the Satisfiability of Pseudo-Boolean Problem with Graph Neural Networks,” in *Principles and Practice of Constraint Programming*, H. Simonis, Ed., Cham: Springer International Publishing, 2020, pp. 885–898. doi: 10.1007/978-3-030-58475-7_51.
[49] Z. Zheng, C. Guo, J. Chen, and J. Li, “Graph Neural Network-Based Representation Learning for Medical Time Series,” in *Artificial Neural Networks and Machine Learning – ICANN 2023*, L. Iliadis, A. Papaleonidas, P. Angelov, and C. Jayne, Eds., Cham: Springer Nature Switzerland, 2023, pp. 194–205. doi: 10.1007/978-3-031-44223-0_16.
[50] “S3-Leitlinie Prävention, Diagnostik, Therapie und Nachsorge des Lungenkarzinoms.” [Online]. Available: https://www.leitlinienprogramm-onkologie.de/fileadmin/user_upload/Downloads/Leitlinien/Lungenkarzinom/Version_3/LL_Lungenkarzinom_Langversion_3.0.pdf
[51] E. Rocheteau, C. Tong, P. Veličković, N. Lane, and P. Liò, “Predicting Patient Outcomes with Graph Representation Learning,” Jan. 11, 2021, *arXiv*: arXiv:2101.03940. doi: 10.48550/arXiv.2101.03940.

[52] C. Tong, E. Rocheteau, P. Veličković, N. Lane, and P. Liò, “Predicting Patient Outcomes with Graph Representation Learning,” in *AI for Disease Surveillance and Pandemic Intelligence: Intelligent Disease Detection in Action*, A. Shaban-Nejad, M. Michalowski, and S. Bianco, Eds., in Studies in Computational Intelligence. , Cham: Springer International Publishing, 2022, pp. 281–293. doi: 10.1007/978-3-030-93080-6_20.
[53] S. Tang *et al.*, “Predicting 30-day all-cause hospital readmission using multimodal spatiotemporal graph neural networks,” *IEEE J. Biomed. Health Inform.*, vol. PP, Jan. 2023, doi: 10.1109/JBHI.2023.3236888.
[54] C. Theodoropoulos, N. Mulligan, T. Stappenbeck, and J. Bettencourt-Silva, “Representation Learning for Person or Entity-centric Knowledge Graphs: An Application in Healthcare,” May 10, 2023, *arXiv*: arXiv:2305.05640. doi: 10.48550/arXiv.2305.05640.
[55] S. N. Golmaei and X. Luo, “DeepNote-GNN: predicting hospital readmission using clinical notes and patient network,” in *Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, in BCB ’21. New York, NY, USA: Association for Computing Machinery, Aug. 2021, pp. 1–9. doi: 10.1145/3459930.3469547.
[56] J. Shang, T. Ma, C. Xiao, and J. Sun, “Pre-training of Graph Augmented Transformers for Medication Recommendation,” Nov. 26, 2019, *arXiv*: arXiv:1906.00346. doi: 10.48550/arXiv.1906.00346.
[57] J. Zhu *et al.*, “Geometric graph neural networks on multi-omics data to predict cancer survival outcomes,” *Comput. Biol. Med.*, vol. 163, p. 107117, Sep. 2023, doi: 10.1016/j.compbiomed.2023.107117.
[58] Z. Bai *et al.*, “Predicting response to neoadjuvant chemotherapy in muscle-invasive bladder cancer via interpretable multimodal deep learning,” *Npj Digit. Med.*, vol. 8, no. 1, p. 174, Mar. 2025, doi: 10.1038/s41746-025-01560-y.
[59] M. Ye *et al.*, “Integration of graph neural networks and transcriptomics analysis identify key pathways and gene signature for immunotherapy response and prognosis of skin melanoma,” *BMC Cancer*, vol. 25, no. 1, Art. no. 1, Dec. 2025, doi: 10.1186/s12885-025-13611-4.
[60] R. Sato, “A Survey on The Expressive Power of Graph Neural Networks,” Oct. 16, 2020, *arXiv*: arXiv:2003.04078. doi: 10.48550/arXiv.2003.04078.
[61] U. Alon and E. Yahav, “On the Bottleneck of Graph Neural Networks and its Practical Implications,” presented at the International Conference on Learning Representations, Feb. 2022. Accessed: Sep. 07, 2022. [Online]. Available: https://openreview.net/forum?id=i80OPhOCVH2
[62] J. Topping, F. Di Giovanni, B. P. Chamberlain, X. Dong, and M. M. Bronstein, “Understanding over-squashing and bottlenecks on graphs via curvature,” arXiv.org. Accessed: Oct. 04, 2024. [Online]. Available: https://arxiv.org/abs/2111.14522v3
[63] Z. Zhang, B. Jiang, J. Tang, J. Tang, and B. Luo, “Incomplete Graph Learning via Partial Graph Convolutional Network,” *IEEE Trans. Artif. Intell.*, vol. 5, no. 9, pp. 4315–4321, Sep. 2024, doi: 10.1109/TAI.2024.3386499.
[64] Z. Lu, Y. Liu, G. Wen, B. Zhou, W. Zhang, and J. Zhang, “Noise-resistant graph neural networks with manifold consistency and label consistency,” *Expert Syst. Appl.*, vol. 245, p. 123120, Jul. 2024, doi: 10.1016/j.eswa.2023.123120.
[65] C. Kulkarni *et al.*, “Hybrid disease prediction approach leveraging digital twin and metaverse technologies for health consumer,” *BMC Med. Inform. Decis. Mak.*, vol. 24, no. 1, Art. no. 1, Dec. 2024, doi: 10.1186/s12911-024-02495-2.
[66] A. Vallée, “Digital twin for healthcare systems,” *Front. Digit. Health*, vol. 5, p. 1253050, Sep. 2023, doi: 10.3389/fdgth.2023.1253050.
[67] J. Akroyd, S. Mosbach, A. Bhave, and M. Kraft, “Universal Digital Twin - A Dynamic Knowledge Graph,” *Data-Centric Eng.*, vol. 2, p. e14, Jan. 2021, doi: 10.1017/dce.2021.10.
[68] J. Richter, F. Lange, T. Scheper, D. Solle, and S. Beutel, “Digitale Zwillinge in der

Bioprozesstechnik – Chancen und Möglichkeiten," *Chem. Ing. Tech.*, vol. 95, no. 4, pp. 498–510, 2023, doi: 10.1002/cite.202200166.
[69] S. A. Peters, *Physiologically-based pharmacokinetic (PBPK) modeling and simulations principles, methods, and applications in the pharmaceutical industry*. Hoboken, N.J.: Wiley, 2012. Accessed: Oct. 15, 2013. [Online]. Available: http://onlinelibrary.wiley.com/book/10.1002/9781118140291
[70] "EDITH: Building the European Virtual Human Twin." [Online]. Available: https://www.edith-csa.eu/wp-content/uploads/2023/04/EDITH_D3.1_vision_roadmap_outline_final.pdf
[71] S. Mei and K. Zhang, "A machine learning framework for predicting drug–drug interactions," *Sci. Rep.*, vol. 11, no. 1, p. 17619, Sep. 2021, doi: 10.1038/s41598-021-97193-8.
[72] R. B. Diasio and S. M. Offer, "Testing for Dihydropyrimidine Dehydrogenase Deficiency to Individualize 5-Fluorouracil Therapy," *Cancers*, vol. 14, no. 13, p. 3207, Jun. 2022, doi: 10.3390/cancers14133207.
[73] J. G. Diaz Ochoa, J. Bucher, J. Niklas, and K. Mauch, "Multi-scale modeling for individualized spatiotemporal prediction of drug effects," *Toxicol. Lett.*, vol. 221, Supplement, pp. S83–S84, Aug. 2013, doi: 10.1016/j.toxlet.2013.05.093.
[74] J. G. Diaz Ochoa, J. Bucher, and J. Niklas, "Prediction of Human Toxicity Thresholds from In Vitro ADMET Data using ab-initio pharmacokinetics," Stuttgart, 2013.
[75] S. T. Mulder *et al.*, "Dynamic Digital Twin: Diagnosis, Treatment, Prediction, and Prevention of Disease During the Life Course," *J. Med. Internet Res.*, vol. 24, no. 9, p. e35675, Sep. 2022, doi: 10.2196/35675.
[76] T. Sun, X. He, and Z. Li, "Digital twin in healthcare: Recent updates and challenges," *Digit. Health*, vol. 9, p. 20552076221149651, Jan. 2023, doi: 10.1177/20552076221149651.
[77] J. G. Diaz Ochoa, J. Bucher, A. Pery, J. Zaldivar Comenges, J. Niklas, and K. Mauch, "A Multi-Scale Modeling Framework for Individualized, Spatiotemporal Prediction of Drug Effects and Toxicological Risk," *Front. Pharmacol.*, vol. 3, 2013, Accessed: Nov. 13, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fphar.2012.00204
[78] H.-C. Chang, A. M. Gitau, S. Kothapalli, D. R. Welch, M. E. Sardiu, and M. D. McCoy, "Understanding the need for digital twins' data in patient advocacy and forecasting oncology," *Front. Artif. Intell.*, vol. 6, 2023, Accessed: Dec. 08, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frai.2023.1260361
[79] T. Margaria and S. Ryan, "Data and Data Management in the Context of Digital Twins," in *The Digital Twin*, N. Crespi, A. T. Drobot, and R. Minerva, Eds., Cham: Springer International Publishing, 2023, pp. 253–278. doi: 10.1007/978-3-031-21343-4_10.
[80] B. J. Awrahman, C. Aziz Fatah, and M. Y. Hamaamin, "A Review of the Role and Challenges of Big Data in Healthcare Informatics and Analytics," *Comput. Intell. Neurosci.*, vol. 2022, p. 5317760, Sep. 2022, doi: 10.1155/2022/5317760.
[81] K. P. Venkatesh, M. M. Raza, and J. C. Kvedar, "Health digital twins as tools for precision medicine: Considerations for computation, implementation, and regulation," *Npj Digit. Med.*, vol. 5, no. 1, Art. no. 1, Sep. 2022, doi: 10.1038/s41746-022-00694-7.
[82] F. Tao, H. Zhang, and C. Zhang, "Advancements and challenges of digital twins in industry," *Nat. Comput. Sci.*, vol. 4, no. 3, pp. 169–177, Mar. 2024, doi: 10.1038/s43588-024-00603-w.
[83] X. Zhai *et al.*, "Heterogeneous System-on-Chip-Based Lattice-Boltzmann Visual Simulation System," 2020, doi: 10.1109/JSYST.2019.2952459.
[84] H. Endo *et al.*, "Development and validation of the predictive risk of death model for adult patients admitted to intensive care units in Japan: an approach to improve the accuracy of healthcare quality measures," *J. Intensive Care*, vol. 9, p. 18, Feb. 2021, doi: 10.1186/s40560-021-00533-z.
[85] J. Reiss and R. A. Ankeny, "Philosophy of Medicine," in *The Stanford Encyclopedia of Philosophy*, Spring 2022., E. N. Zalta, Ed., Metaphysics Research Lab, Stanford University,

2022. Accessed: Aug. 02, 2023. [Online]. Available: https://plato.stanford.edu/archives/spr2022/entries/medicine/

[86] C. Abettan and J. V. M. Welie, “The impact of twenty-first century personalized medicine versus twenty-first century medicine’s impact on personalization,” *Philos. Ethics Humanit. Med.*, vol. 15, no. 1, Art. no. 1, Dec. 2020, doi: 10.1186/s13010-020-00095-2.

[87] D. Böhler and M. Friebe, “(Digital) Patient Journey and Empowerment: Digital Twin,” in *Novel Innovation Design for the Future of Health: Entrepreneurial Concepts for Patient Empowerment and Health Democratization*, M. Friebe, Ed., Cham: Springer International Publishing, 2022, pp. 169–178. doi: 10.1007/978-3-031-08191-0_17.

[88] L. Nye, “Digital Twins for Patient Care via Knowledge Graphs and Closed-Form Continuous-Time Liquid Neural Networks,” Jul. 08, 2023, *arXiv*: arXiv:2307.04772. doi: 10.48550/arXiv.2307.04772.

[89] M. D. Wilkinson *et al.*, “The FAIR Guiding Principles for scientific data management and stewardship,” *Sci. Data*, vol. 3, p. 160018, Mar. 2016, doi: 10.1038/sdata.2016.18.

[90] “About Solid.” [Online]. Available: https://solidproject.org/about

[91] A. Vaid, “Medical AI can transform medicine — but only if we carefully track the data it touches,” *Nature*, vol. 642, no. 8069, pp. 864–866, Jun. 2025, doi: 10.1038/d41586-025-01946-8.