



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/243227/>

Version: Accepted Version

Article:

Sujan, Mark Alexander, Lounsbury, Olivia, Pickup, Laura et al. (2026) Resilience Engineering Concepts, Safety-II Language, FRAM Practice, and Co-Creating Communities: How Hollnagel Reshaped Patient Safety. Safety science. 107210. ISSN: 0925-7535

<https://doi.org/10.1016/j.ssci.2026.107210>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Resilience Engineering Concepts, Safety-II Language, FRAM Practice, and Co-Creating Communities: How Hollnagel Reshaped Patient Safety

M. Sujan¹, O. Lounsbury², L. Pickup³, K. Preston¹, R. Patriarca⁴

¹ Department of Computer Science, University of York, UK

² OxSTaR, University of Oxford, UK

³ LP Human Factors Ltd., UK

⁴ Department of Mechanical and Aerospace Engineering, Sapienza University of Rome, Italy

Corresponding Author: M Sujan, Department of Computer Science, University of York, UK
mark.sujan@york.ac.uk

Manuscript type: Original Article

Word count: 5,986

Resilience Engineering Concepts, Safety-II Language, FRAM Practice, and Co-Creating Communities: How Hollnagel Reshaped Patient Safety

1. INTRODUCTION

Back in 2011, one of the authors of this paper submitted a manuscript in collaboration with Erik Hollnagel to a leading patient safety journal. The manuscript argued the case for resilient healthcare systems and acknowledged in its title that *“performance variability is ubiquitous, inevitable but also useful”*. The manuscript was rejected. Arguably, because it was not as strong as we had intended. However, one comment from the editor in chief is worth highlighting: *“it [Resilience Engineering applied to healthcare] may be counterproductive since it may encourage exactly the type of thinking we have tried to avoid in [the] past 10 – 15 years, namely a reversion to thinking that ‘every patient is different’ justifies huge amounts of variation and lack of standardization”*.

This comment reflected the dominant mindset at the time [1]. The emphasis was on quality improvement (QI) approaches designed to reduce failures in care, with a strong focus on bringing industry-style standardisation to areas of care perceived to suffer from too much unjustified variation [2]. Further comments reinforced this perspective: *“There is too much variation even in aspects of care for which there is an established or ‘right’ approach, and even for areas where there are uncertainties, it would often be better to standardize things”*.

This feedback was indicative of a broader unease with the core premise of Hollnagel’s work in Resilience Engineering and Resilient Health Care, namely, that performance variability is not inherently a problem, but a necessary and usually beneficial feature of work in complex settings [3]. Following Hollnagel’s perspective, performance variability is the source of both success and failure [4]. At the time, this claim appeared outlandish and threatening to many leaders in patient safety, who were concerned with control, compliance and reduction of variation through “evidence-based” interventions, such as checklists and care bundles, as primary strategies for improvement [5].

Fast forward to 2025, and, thankfully, patient safety as a multi-disciplinary field is in a different, and arguably more mature place. Concepts, such as performance variability, adaptations and trade-offs, the need to understand work-as-done and to learn from what goes well, and, more broadly, Safety-II as a different lens on patient safety have been integrated into patient safety

frameworks, policies and practice [6-9]. These ideas now feature prominently in an increasing number of publications, both in that particular journal, which rejected the manuscript in 2011, as well as across most patient safety literature, including a special issue of *Safety Science* dedicated to Resilient Health Care [10-13]. Healthcare professionals and researchers now routinely and confidently talk about the realities of working in complex and dynamic environments, recognising that work-as-done typically differs from work-as-imagined by managers and policy makers [14-17].

This shift owes much to the contributions of Erik Hollnagel, who provided not only theories and methods, but also a vocabulary and a community through which healthcare practitioners could engage with the complexity of their work and begin to make sense of patient safety in a more proactive way.

In this paper, we reflect on Erik Hollnagel's distinct contribution to the field of patient safety. We argue that his influence goes beyond reframing core safety concepts, such as the role of performance variability and the focus on what goes well. Equally as important is his development of language, methods and communities that have enabled practitioners to engage meaningfully with complex systems.

The next section (Section 2) traces the early emphasis in patient safety on control and standardisation driven by quality improvement approaches. Then, we explore how Hollnagel's work in Resilience Engineering (Section 3) and Safety-II (Section 4) challenged these mainstream approaches by introducing new ways of thinking and talking about safety in complex settings. Section 5 looks at the role of the Functional Resonance Analysis Method (FRAM) as a practical tool to enable stakeholders to interrogate, visualise and discuss performance variability and work-as-done. In Section 6, we reflect on Hollnagel's legacy in action across three dimensions: first, the reframing of patient safety practice from compliance monitoring to systemic understanding; second, the relevance of his concepts and methods for governing emerging technologies such as artificial intelligence; and third, the role of professional and practitioner communities as co-creators in translating and extending his ideas. We conclude (Section 7) by considering the enduring significance of Hollnagel's contribution for future patient safety, emphasising patient safety not as a static state where nothing goes wrong to be maintained, but as something that is actively achieved: in short, a shift from managing safety to managing safely [18].

2. THE MAINSTREAM PATIENT SAFETY MOVEMENT FROM 2000 ONWARDS: RELIABILITY, ERROR AND STANDARDISATION

2.1 Linear logic and failure focus

The evolution of system safety theories and accident models has informed the methods and trends adopted in patient and wider healthcare safety improvement. However, only a small proportion of these theories is widely applied in practice. This is illustrated by the widespread adoption of Root Cause Analysis (RCA), which aims to apply systems thinking and safety science to identify incident causal factors and support organisational learning [19]. The language and search for a small number of “causes”, coupled with the visualisation of findings through an Ishikawa “fishbone” diagram, have contributed to a reductionist approach to interpreting healthcare incidents. This fails to recognise the role of interacting system components, which are acknowledged as essential for fully representing how events unfold within the complex context of healthcare systems [20]. The impact of this oversimplification has resulted in a person and behaviour-based approach to investigations and their subsequent recommendations for improvement, which usually focus on training or writing of procedures and safety notices. While such interventions can reflect legitimate strategies for managing risk, their effectiveness in healthcare is inherently limited when they are treated as stand-alone solutions, rather than as elements that must be embedded within broader socio-technical system change [21]. Consequently, they might degenerate to what is now referred to as “safety clutter” [22, 23]. These approaches came with their own language: error, deviation, and non-compliance. While powerful, this vocabulary often rendered the adaptive and relational aspects of clinical work invisible.

RCA continues to be relied upon within the QI domain, with ongoing encouragement for individual organisations to focus on QI approaches to address safety [24]. QI has provided a strong compass, guiding healthcare staff on how to analyse their systems. This has provided a set of methods and skills to encourage clinicians to improve their own systems. Although widely adopted by many enthusiasts, the concepts of QI have frequently been selectively applied and oversimplified in patient safety contexts [25]. This might be because the methods again can encourage a reductionist approach to describing healthcare systems as linear systems and treating them akin to a tractable manufacturing production line. This approach has subsequently led to actions that seek to standardise processes and behaviours, and to

reduce waste, as QI and Lean principles have become conflated. This is not to suggest that standardisation has no place in improving patient safety, but that its application is but one strategy, and that it requires careful consideration of context and how standardised practices, such as standardised communication protocols or the surgical safety checklist, interact with other elements of the work system [26, 27].

The simplistic application of QI provides another example where the temptation for a reductionist approach has misrepresented the original intention of Donabedian's model of quality in healthcare, which emphasises the interplay of system structure, processes, and outcomes [28], yet patient safety applications of QI have often foregrounded measurable process compliance at the expense of understanding how care is actually organised and adapted in practice.

Darzi differentiates access to care from the quality of care, defining access as the timing and convenience of receiving care. He defines quality by three dimensions: safe (do no harm and as error-free as possible), effective (maintaining functionality or improving survival rates), and experience (compassionate and respectful care) [29]. The Institute of Medicine (IOM) defines quality to include six dimensions, one of which is "safe", described as "avoiding harm to patients from care that is intended to help them". These definitions of "safe" describe intended outcomes of a healthcare system but have inadvertently encouraged equating safety with initiatives that seek to count incidents and reduce the variability of one aspect of the system.

Healthcare has, understandably, looked to other safety-critical industries to learn about safety. However, the translation of knowledge has not always fully understood the differences between the properties of these industries' systems and healthcare systems or selected only one aspect of the theory to focus on, while neglecting other underpinning principles [30].

The IOM publication "To Err is Human" [31] was instrumental in educating healthcare professionals on the concept of High Reliability Organisations (HROs). However, the language and definition of reliability used within healthcare has been misleading. In the context of HRO, reliability is not equal to a lack of variance in performance. The concept behind HRO suggests that for a system to be reliable, it must be able to accommodate unpredictable and variable working conditions [32]. Healthcare appears to have tried to shoehorn HRO into a QI mindset. This approach equates routines and standardisation with reliability, which is challenged in

testing circumstances where unvarying procedures cannot cope with unanticipated situations. This is at odds with the intention of HROs, which emphasises not only anticipation of and preparedness to known risks, but also highlights the importance of the social and relational foundations that underpin an organisation's ability to recognise weak signals and to respond to errors flexibly and successfully [30, 33, 34].

Another impactful visualisation of safety is James Reason's "Swiss Cheese" model of organisational accidents [35]. This model is widely used in healthcare and has encouraged consideration of both active and latent failures [36]. The "Swiss Cheese" model emphasises a focus on the failure of a system through the alignment of "holes" or weaknesses in the controls intended to prevent failure. The model has served healthcare well to redirect focus purely from the person in the system to organisational (latent) contributory factors. However, again the language and visualisation position the "Swiss Cheese" model as fundamentally a failure-oriented model, structured around defences-in-depth, unsafe acts and latent failures (or: conditions). Conceptually, it does not have a place for adaptation as a source of success, and therefore offers limited support for understanding how healthcare staff regularly achieve safe work despite inadequate resources or controls [37].

2.2 A shift towards complexity

Le Coze has discussed how visualisations have contributed to communicating complex concepts in safety research [38]. He highlights the groundbreaking work of Perrow and Rasmussen in redefining how we think about safety, but these contributions remained largely peripheral to the mainstream patient safety community [39, 40].

Perrow's work captures in the graphical representation in an easy way how different systems compare in terms of the complexity of system interactions and the degree of coupling, and how this influences the likelihood of systemic accidents across different high-risk industries. This suggests that an industry's inherent properties determine the relationship between coupling (the dependency between system parts) and complexity, which in turn affects the predictability of accidents.

Rasmussen's illustrations highlight the dynamic nature of risk management within sociotechnical systems. This changed the understanding of "human error" and accident causation, recognising that at every system level, trade-offs are made to balance competing

goals like safety, economics, and efficiency. He challenged the idea of static safety barriers, instead visualising them as a dynamic “envelope of boundaries” that reflect organisational goals. This powerful image explains how safety is created through the constant adjustments of actors across the system. Rasmussen's model helps reframe “human error” by showing that these balancing acts can either succeed or fail, and both are the result of actors trying to balance multiple goals.

2.3 A conceptual impasse in patient safety

These developments reveal a growing conceptual tension within patient safety. On the one hand, dominant improvement approaches continued to rely on linear models of causation and interventions that emphasised control and the elimination of error. On the other hand, work by Perrow and Rasmussen introduced different ways of understanding safety as an emergent property of sociotechnical systems characterised by interactions and continual adjustment. While influential within safety science, the perspectives of Perrow and Rasmussen remained only partially integrated into mainstream patient safety practice. As a result, patient safety found itself at a conceptual impasse: increasingly aware of complexity yet largely constrained to the logics and approaches shaped by linear assumptions about failure and control.

Erik Hollnagel's work can be understood as a deliberate attempt to transcend this impasse. He recognised that moving beyond a focus on failure would require more than new concepts alone, but a supporting architecture for the field that could reshape how safety is understood, discussed, managed and improved. The sections that follow explore this new architecture in terms of its conceptual foundations, imagery and language, analytical tools, and co-creating communities of practice.

3. RESILIENCE ENGINEERING: A NEW CONCEPTUAL FOUNDATION FOR PATIENT SAFETY

Resilience Engineering (RE) entered the safety field with the 2006 volume “Resilience Engineering: Concepts and Precepts”, edited by Erik Hollnagel, David Woods, and Nancy Leveson [41]. The book distilled discussions from an October 2004 workshop in Söderköping, Sweden, and called for a new way of thinking about safety-critical work. Its core insight was that in modern sociotechnical systems accidents seldom stem from a single component failure. They tend to emerge when otherwise normally functioning system elements interact in unforeseen, dysfunctional ways. Failure is therefore a property of the whole system, the

dark mirror of everyday success. In the same book, Dekker recognises RE as a “confused consensus”, out of which several initiatives would be developed.

From this initial volume, several research streams unfolded. Leveson extended her system-theoretic work into STAMP (System-Theoretic Accident Model and Processes), viewing safety as the maintenance of constraints across hierarchical levels [42]. David Woods later gave RE its central vocabulary of “graceful extensibility”, demonstrating through field studies and theoretical models how human-machine systems stretch performance boundaries following certain patterns [43]. Dekker advanced a more human-centred perspective that emphasises supporting practitioners rather than controlling and constraining them [44].

Hollnagel’s line of inquiry is just as distinctive, focusing on performance variability, i.e., the small adaptations people across a system make to keep operations flowing. The same variability that lets staff cope with complexity can, under different conditions, seed failure. What matters is understanding how functions couple in time and space as situations change. As Le Coze noted, these perspectives converge in the “New View” of safety, which studies systems as integrated wholes rather than collections of parts [45].

One of Hollnagel’s most influential concepts is the Efficiency-Thoroughness Trade-Off (ETTO), which captures the constant tension between working quickly and working carefully [3]. Organisations slide along this continuum in response to resource pressures and competing goals. Building on ETTO, Hollnagel articulated four resilience potentials - responding, monitoring, learning, and anticipating - that together describe a system’s capacity to maintain essential performance in both expected and unexpected conditions [46]. He translated these ideas into practical tools such as the Resilience Assessment Grid (later renamed Systemic Potentials Management) and the Functional Resonance Analysis Method (FRAM), which shows how everyday variability in separate functions can interact and amplify into unwanted outcomes [47]. Other more recent RE tools took inspiration from ETTO, too, for example Structured Exploration of Complex Adaptations (SECA) [48].

Despite its conceptual promise, Hollnagel’s RE vision saw limited early uptake in healthcare. Patient safety improvement programmes largely remained anchored in linear, causal models of failure, with improvement interventions focused on reducing error and enforcing compliance. Moreover, technical terms like “ETTO” or “resilience potentials” lacked resonance with the everyday language of clinical work on hospital wards, during handovers,

and more generally in care pathways. This made it more challenging to communicate the practical relevance of RE.

To close this gap, Hollnagel helped seed early communities of practice, such as the annual “Safety-II in Practice” workshop and the Resilient Health Care Network. Later, these communities of practice would become central to translating abstract principles into concrete, clinically meaningful insights (see Section 6).

RE provided the theoretical groundwork. As described in the next Section, Safety-II provided the accessible language, the strategic framing and a growing community to allow these ideas to reshape how patient safety is talked about and practised in healthcare.

4. SAFETY-II: LANGUAGE AS A TOOL FOR REIMAGINING PATIENT SAFETY

The introduction of Safety-II marked a pivotal moment in the development of patient safety thinking [49]. Not because it introduced entirely new ideas, but because it offered new terminology and language that made those ideas intelligible, speakable and actionable within patient safety and healthcare more broadly. Earlier concepts associated with RE had limited uptake in healthcare up to that point, arguably due to the use of language and metaphors more familiar to traditional safety-critical domains. Safety-II, on the other hand, succeeded in part because it gave practitioners a positive vocabulary for aspects of clinical work that are essential for safety, but which had long been invisible or inexpressible.

Safety-II represents a reframing: from understanding and operationalising safety as the absence of something, e.g., freedom from unacceptable levels of risks of physical injury or damage (defence and other high-risk sectors), or preventable adverse events (healthcare sector), to seeing safety as the presence of abilities that enable things to go right under varying conditions, Hollnagel’s point being that it is more meaningful to study the presence of something rather than an absence. This shift is accompanied by a focus on understanding and managing these abilities or systemic resilience potentials by learning from everyday work and treating performance variability as inevitable and generally useful [50].

This is not just a conceptual shift, but a linguistic one. Language, in this context, serves not only as a medium for communication, but also, importantly, as a tool for reorienting perception and discourse. For example, as Bruner has argued, language and the stories we tell shape what we see as meaningful and actionable [51]. Safety-II offered not just new labels

but a new narrative, shifting the focus from human error and compliance with protocols to necessary adaptations and trade-offs in everyday work to achieve successful outcomes in complex systems.

At this point, it is worth acknowledging that, strictly speaking, there is no such thing as “Safety-I”. What we have is a rich diversity of concepts, models and methods within safety science, which have been developed over the past 100 years [52]. It would be overly reductionist and potentially counterproductive to group all these approaches under the Safety-I label and treat Safety-II as a wholesale replacement. Indeed, such criticism has been expressed in the literature [53, 54]. The same caution also applies to Safety-II, which should not be seen as a unitary framework or method, but rather as a set of concepts and perspectives that extend existing traditions. In fact, key concepts of Safety-II are, to a certain extent, extensions of thinking that started in the 1970s and 1980s, including the pioneering work of Jens Rasmussen as well as Erik Hollnagel’s collaboration with David Woods [40, 55, 56].

However, the linguistic juxtaposition of Safety-I and Safety-II offered a powerful rhetorical device [6]. It provided people with a simple, albeit simplistic, frame of reference and an accessible starting point for thinking about safety through a different lens. The contrast, while open to critique, was strategic. It served as a provocation, challenging entrenched assumptions and opening conceptual space for practitioners to reframe what counts as safety. This has been challenged conceptually, for example by Reason, who cautioned against abandoning the language of error, which in his view remained indispensable for learning and accountability [57], as well as practically by current approaches to learning from incidents, accidents and adverse events [58].

The vocabulary associated with Safety-II played a crucial role in this shift. Terms such as work-as-done being distinct from work-as-imagined, trade-offs and performance variability, and learning from what goes well enabled healthcare professionals to describe, in a positive and constructive way, the realities of their practice and everyday work, which were previously either difficult to articulate or framed negatively in the dominant language of human error, lack of compliance and violation of standards. By naming and positively framing these concepts, Safety-II made these aspects of clinical work not only discussable but legitimate sources of learning and levers for improvement in healthcare organisations.

Yet, the very accessibility of this language has also drawn critique. As Safety-II terminology gained traction across research, policy and practice, its core concepts were at times adopted in superficial ways and reduced to slogans or misinterpreted as merely affirming existing practices [45]. This risk of dilution is inherent to any language-driven shift. The easier terms are to adopt, the easier they are to co-opt without critical engagement.

This new language also enabled new practices. As Wenger argues, shared language within a community of practice does more than enable communication, it shapes how people understand their roles, actions, and shared purpose [59]. The notion of “learning from what goes well”, for example, has been influential in reshaping how incidents are investigated in the NHS in England. The Patient Safety Incident Response Framework (PSIRF), adopted in 2022, continues to focus on patient safety incidents in name, but it emphasises that investigations and learning responses should go beyond the incident itself to examine how the system functions on an everyday basis. PSIRF uses the term “learning response” to denote different systems-based approaches that can be utilised following a patient safety incident, e.g., patient safety incident investigation, After Action Review, Swarm Huddle and Thematic Review. The focus of PSIRF is on learning and improvement, not on attributing accountability and blame. NHS England notes explicitly in the PSIRF guidance that responding to patient safety incidents is no longer a search for root causes, but a broader effort to understand how outcomes are produced within a complex system [60]. Similarly, the new Learning from Patient Safety Events Service (LfPSE), which replaces previous national incident reporting systems, reflects this shift through its emphasis on “events” rather than “incidents”, though, arguably, in practice it still retains a predominantly incident-focused orientation.

The spread of this new language was not accidental. It was supported by an active and growing community of practice, especially the Resilient Health Care Society (formerly the Resilient Health Care Network), founded by Erik Hollnagel in collaboration with patient safety leaders Jeffrey Braithwaite and the late Bob Wears. This interdisciplinary community provided an annual forum for academics and patient safety practitioners to come together and ground Hollnagel’s typically abstract ideas in clinical realities. These efforts have resulted in a series of six edited volumes, which carry titles such as *“The resilience of everyday work”*, *“Reconciling work-as-imagined and work-as-done”*, *“Working across boundaries”* and *“Muddling through*

with purpose” to provide clinically relevant explorations based on the new terminology [61-64].

In this way, Safety-II became more than an abstract conceptual alternative. It enabled a way of talking about patient safety that made a different kind of thinking and doing possible. One might argue that Safety-II reconfigured the conceptual grammar of patient safety and, thereby, expanded the limits of what could be seen, said and valued.

5. The FRAM AS A PRACTICE: MAKING SENSE OF COMPLEX CLINICAL SYSTEM BEHAVIOURS

While Resilience Engineering and Safety-II made new ways of thinking about safety intelligible and speakable, the Functional Resonance Analysis Method (FRAM) gave healthcare practitioners a structured way to operationalise these ideas in investigations, quality improvement, and system design [4]. Hollnagel developed the FRAM before Safety-II and in parallel with early RE initiatives. The origins of the FRAM go back to 2004, where it was published as an accident model (Functional Resonance Accident Model), but it has since evolved into an analysis method after discussions and reflections in FRAM-related workshops, such as the “FRAMily” [65], though it can be debated whether the FRAM meets the criteria of a scientific method, whether it is a technique, or whether it sits somewhere in between as a theory-driven form of inquiry. The FRAM as an accident model suggests that accidents emerge from everyday functional variability across the system, each creating only a weak undetectable signal, but occasionally reinforcing (or resonating with) other variability to create a detectable signal, i.e., the incident or accident. In this way, the logic of the FRAM supports analysts to describe how accidents can emerge from the interaction of normal performance variability even when no individual system element has failed in the traditional sense. The corollary to this statement is that safety management should be concerned with monitoring and managing variability rather than preventing failures, echoing again thinking that was introduced at that time by a nascent RE community.

Initially, the FRAM saw only limited use in healthcare. Early interest tended to come from researchers closely affiliated with Hollnagel and the Resilience Engineering community, while the majority of applications of the FRAM were in aviation [66, 67]. Its wider uptake came later, when the FRAM was reframed as an analysis method for making sense of complex socio-technical system behaviour. This shift coincided with, and was catalysed by, the emergence of Safety-II, which provided an accessible conceptual framing and vocabulary that helped

healthcare practitioners engage with systemic modelling of everyday clinical work. The FRAM was also seen as a practical tool that incorporated foundational RE thinking in its principles.

Since about 2015, there has been an increasing interest in the FRAM in healthcare, making it one of the strongest application domains as evidenced by reviews of the literature [47, 68]. The FRAM as an analysis method started to be used not just to analyse why things go wrong as one would do in a traditional patient safety incident investigation, but to understand how things usually go right under variable and often challenging conditions. For example, a critical reflection on the FRAM applied in healthcare included studies that provided insights into the management of deteriorating patients on a surgical emergency unit, used the FRAM to develop shared understanding of dependencies and variabilities in the process of blood sampling, described the system context for applying sodium fluoride varnish in children's dental care, highlighted the difference between work-as-imagined and work-as-done in cardiothoracic surgery departments, and undertook a proactive risk assessment in a neurosurgical unit [14, 17, 69-72]. Common characteristics across these applications of the FRAM were that the studies address stubborn problems where a different perspective or analytical lens was thought to offer complementary insights, focus on illuminating systems dynamics and work-as-done rather than looking at failures, reflect on the gap between work-as-imagined and work-as-done, and that they are highly participatory in their approach [69]. The flexibility of the FRAM has enabled its application across different levels of the healthcare system, including mapping national-level responses, such as the UK's management of the COVID-19 pandemic [73].

Despite this progress, concepts central to Hollnagel's thinking, such as non-linearity, performance variability, and functional interaction, are at times invoked rhetorically, while the underlying analytical logic of studies remains largely linear. In such cases, FRAM or Safety-II terminology is layered onto studies that still seek discrete causes and corrective interventions characteristic of traditional Safety-I approaches [74]. This tendency risks misrepresenting the intent of Hollnagel's contribution and contributes to the perception that non-linearity is more an abstract argument than an enacted practice. However, this does not invalidate the underlying theory. Rather, it highlights the difficulty of translating non-linear concepts into institutionalised safety practices, given the continued dominance of linear assumptions about

causality and evaluation that shape how improvement is expected to occur. Importantly, there are published examples where non-linearity is enacted more faithfully in practice.

One illustration of how non-linearity is enacted in patient safety practice can be seen in FRAM-based analyses of the management of the deteriorating patient [17]. In this work, decisions to escalate care were not triggered deterministically by early warning scores or predefined thresholds. Instead, they developed over time through ongoing assessments by nurses and junior (resident) doctors, shaped by issues such as their previous experiences of escalation, the perceived availability and responsiveness of senior staff, as well as competing clinical demands on the ward. Staff frequently adjusted how and when they raised concerns, anticipating whether escalation would be appreciated, and weighing the potential consequences of interrupting already stretched teams. These locally developed ways of managing deterioration were often effective in maintaining patient safety, yet under different circumstances they could also contribute to delayed intervention. As a result, no stable or linear relationship could be established between specific actions and outcomes. Rather than producing isolated corrective actions, the FRAM supported recommendations aimed at changing how concerns about deterioration were surfaced and taken forward across professional boundaries, including how escalation decisions were negotiated in practice in light of workload pressures and access to senior staff. In this sense, non-linearity is enacted not through different forms of measurement, but through a different way of reasoning about why things happen and what it means to intervene.

While in terms of uptake and the ability to provide a systems-based lens that shifts the focus of analysis from “what went wrong / could go wrong” to “how does the system work / how does the system produce outcomes” the FRAM can be considered highly successful, practical challenges remain [75]. First, even where FRAM analyses align with a non-linear and systems-based analysis logic, recommendations for systems improvement derived from the FRAM need to be translated into coherent and actionable improvements. This translation process risks diluting the systemic insights if recommendations are simplified into isolated actions. Second, because the FRAM typically models everyday functioning rather than specific incidents, its outputs may appear less urgent to decision makers focused on immediate risk or harm. Third, FRAM analyses frequently generate multi-faceted recommendations that

necessitate coordination across disparate stakeholders, which can be difficult to achieve in fragmented healthcare systems, working with disjointed mindsets.

Addressing these challenges requires careful communication. Some of the challenges have been considered in the most recent FRAM book [76]. The book aims to convey both challenges and practical solutions, as well as critical reflections on the method's use. A recurring documented issue is that the FRAM is often deployed as a fashionable badge: researchers map the functions and couplings but stop short of showing how the resulting insights changed design decisions, informed improvement interventions, or measurably improved outcomes. In such cases the model becomes an elegant diagram rather than a vehicle for organisational learning: the FRAM for the FRAM's sake, leaving its real value unproven and its potential under-realised, with the risk to generate scepticism in the long run outside the niche community of users.

Following Hollnagel's intuition to design the initial FRAM, nowadays FRAM analysts and the FRAM community at large need to consider how to convey both the analytical value and the practical implications of their analyses to decision makers. Here again, the language introduced by Safety-II plays an important role. It enables analysts to explain the systemic nature of their findings in terms that are accessible, relevant and resonant within the different cultures of healthcare organisations.

6. LEGACY IN ACTION

6.1 Reframing Safety Practice from Compliance to Systemic Understanding and Improvement

The evolution of the theoretical principles (RE), language (Safety-II) and methods (the FRAM) associated with Hollnagel's work has often sparked debate between scholars and practitioners. Productively so: these debates have matured the theory, built a community, and, importantly, changed the practice. For safety practitioners in hospitals and health systems, the legacy is deeply practical and is made clear in the questions they ask, the conversations they have, and the actions they prioritise.

Take, for example, a personal story from one of the authors who was serving as a patient safety professional for a large health system. The author was tasked to ensure compliance with actions resulting from a Root Cause Analysis after delayed intervention for a deteriorating

patient on the ward. As was common at the time, most of the resulting actions focused on training and reminders to staff about the warning signs of deterioration [77]. The standard approach would have been to track compliance with the required actions and follow up individually with deviants. Instead, Resilience Engineering concepts were imported into the discussions with staff from the author's external role as a researcher. Quickly, it became clear that the patient deterioration challenges were not driven by knowledge gaps. Instead, staff had developed compensatory mechanisms to overcome fragmented workflows, inadequate structures for team-based sensemaking, and prioritisation trade-offs that varied from individual to individual. These were the very mechanisms that prevented unrecognised deterioration for as long as they had. Instead of correcting what was originally perceived as a "deviation", the team was now equipped with a non-normative language to discuss the system behaviour and revised the action items to better understand what supported these adaptations and where they were at risk of breaking down.

This example illustrates a broader shift in how patient safety itself can be understood. Recent reflections on so-called "new view" safety theories, including Hollnagel's work, suggest that they are best seen not as safety frameworks in a narrow sense, but as principles of effective organisational management, with safety serving as a particularly revealing application domain rather than the exclusive focus [78]. This interpretation is consistent with Hollnagel's long-standing position that safety cannot be separated from the everyday management of complex systems and the trade-offs organisations must continuously navigate. In healthcare, this perspective is especially consequential. Patient safety is typically concerned with high-frequency events that are embedded in routine care, under conditions of persistent mismatches in demand and capacity, and competing priorities such as access to care, efficiency, equity, and patient experience. This is very different from the traditional focus in safety-critical industries on low-probability, high-severity accidents. In healthcare, treating safety as a discrete managerial objective is neither analytically nor practically tenable. Hollnagel's contribution was therefore not simply to apply management thinking to patient safety, but to articulate a reframing of patient safety as a problem of managing healthcare systems safely. This reframing has been taken up and further developed by sections of the patient safety community familiar with and supportive of his work, offering an additional lens through which safety can be understood and acted upon. It does not replace established

approaches such as quality improvement or evidence-based safety intervention paradigms, but sits alongside them, sometimes in tension and sometimes in complement, expanding the conceptual repertoire available for grappling with safety in complex healthcare systems.

For the patient safety manager, the shift in framing opened up a broader repertoire of possible actions and alleviated artificial constraints driven by narrow framing. Hollnagel's work "legitimises" insight that comes from patterns, rather than isolated incidents, and provides safety professionals with vocabulary that previously felt anecdotal or difficult to articulate meaningfully. Not only will moving from the world of assurance to the world of learning and support happen slowly, but the most effective practitioners will be able to speak both languages. Hollnagel's legacy offers more nuance than just an unhelpful binary choice. It allows safety practitioners to honour the complexity of care and discern when a protocol, re-education, or auditing is and isn't helpful for staff, patients, and families.

6.2 Applying Hollnagel's Legacy to Emerging Technologies

Looking forward, Hollnagel's legacy offers tools and mindsets that are well-suited to modern, rapidly evolving socio-technical healthcare systems, including those integrating artificial intelligence (AI) technologies. The development of AI for healthcare is increasing and becoming prominent in policy and guidance across health systems worldwide [79]. Virtually all areas of healthcare will be affected by AI, including those related to consultations, note taking, diagnosing, medical research and clinical trials, public health initiatives and medical education [80-82]. These technologies are not simple changes to an otherwise stable system, but they are disrupters to what is, already, a complex socio-technical endeavour [83, 84].

Existing safety approaches, developed for bounded technologies in specific contexts, struggle when applied to emerging and frontier AI systems [85]. Generative AI and, potentially, artificial general intelligence (AGI) introduce new classes of hazards, not only at the level of individual devices or workflows, but also through systemic effects arising from their widespread adoption. These risks span operational, organisational, societal, and even existential domains [86]. As noted in a recent scoping review of emerging technology risks, such threats extend well beyond conventional safety hazards to include legal, ethical, socio-economic, malicious-use, and geopolitical dimensions [87]. This creates a profound challenge for traditional risk models, which tend to be narrowly scoped and linear in focus.

Hollnagel's legacy, particularly the theoretical principles of Resilience Engineering, the reframing language of Safety-II, and the systemic mapping afforded by the FRAM, offers a starting point for addressing these challenges. The Safety-II vocabulary of work-as-done versus work-as-imagined, performance variability, and trade-offs can help reframe AI safety discussions to focus on how these technologies will interact with real work systems, rather than on performance metrics in isolation [88]. The Safety-II mindset, with its emphasis on understanding "what goes right" in everyday operations, also has value in identifying unanticipated benefits and resilience pathways that might otherwise be overlooked in a risk-dominated discourse.

The FRAM could be adapted to model not just localised workflows but whole health systems, enabling analysts to explore how generative AI might alter interdependencies, shift resilience potentials, and create new couplings across functions. This could support anticipatory sensemaking and more informed governance of AI-enabled healthcare systems [89, 90]. Such approaches would complement other systemic risk assessment methods, helping safety professionals and policy makers move from piecemeal hazard analysis to a richer, whole system understanding. While extensions to Hollnagel's methods will likely be needed to fully capture the dynamics of frontier AI, his work provides a conceptual and practical foundation for grappling with technologies whose risks and benefits will evolve at a pace unprecedented in healthcare.

While these principles and tools can meaningfully reframe how we approach AI-enabled healthcare systems, there are limits. At the frontier of AI safety, adaptation and resilience alone may not suffice. Some hazards, such as those involving malicious use or systemic destabilisation, may require the imposition of hard constraints, global coordination, or entirely new governance frameworks. Hollnagel's legacy thus offers both a conceptual foundation and a caution: while his methods can help us understand and manage the integration of AI into complex sociotechnical systems, they will need to be extended and complemented by approaches capable of addressing the unprecedented systemic and existential risks that frontier technologies may bring.

6.3 Communities as Co-Creators of Legacy

A distinctive element of Hollnagel's contribution to patient safety is not only his theoretical, linguistic, and methodological innovations, but also his role in cultivating communities capable

of sustaining and evolving them. The Resilient Health Care Society, the *Safety-II in Practice* workshops, and the highly active FRAMily meetings exemplify this dimension of his legacy. These communities provide more than opportunities for presenting papers. They act as interpretive spaces where concepts are stress-tested against the realities of practice, adapted to diverse contexts, and enriched through dialogue across professional and sectoral boundaries – in the format of “long discussions interrupted by short presentations” (a phrase Erik Hollnagel likes to use).

These communities bring together clinicians, engineers, human factors specialists, policymakers, and academics, and enable cross-pollination that has been crucial for diffusion beyond the original Resilience Engineering circle. Safety-II terminology, for example, has gained traction partly because practitioners could discuss concrete examples of work-as-done, performance variability, or trade-offs in settings far removed from the theory’s origins. Through shared cases and collective problem-solving, the abstract vocabulary is translated into actionable insights that feel relevant to local priorities.

Such communities also serve as a safeguard against dilution. As discussed earlier, language-driven shifts risk being co-opted superficially, stripped of their conceptual depth. Within these forums, members can challenge misuse, reassert the underlying principles, and document richer, more nuanced applications. This custodial role helps preserve the integrity of the concepts while allowing necessary adaptation.

Finally, these communities provide the social infrastructure for addressing emerging challenges. The foundational ideas of Resilience Engineering, Safety-II, and the FRAM are not static, but must be reinterpreted in light of shifting technologies, organisational forms, and societal expectations. Communities offer a mechanism for this reinterpretation to happen quickly and collaboratively, whether in response to the integration of generative AI into health systems, or the need to rethink safety under conditions of climate stress, pandemics, or other systemic disruptions.

In this sense, the communities Hollnagel nurtured are not only carriers of his legacy but active agents in its evolution.

7. CONCLUSION

Erik Hollnagel's contribution to patient safety extends far beyond the introduction of new theories. Through Resilience Engineering, Safety-II, and the FRAM, he provided conceptual foundations, practical tools, and a vocabulary that have enabled healthcare professionals to see, talk about, and act on safety in new ways. His work has reoriented patient safety from a focus on error elimination and compliance towards understanding and supporting the adaptive capacities that make everyday success possible in complex systems.

This legacy is not only conceptual. Hollnagel created the social infrastructure through which ideas could be tested, adapted, and embedded in practice by nurturing communities such as the Resilient Health Care Society, Safety-II in Practice workshops, and the FRAMily. These communities have been instrumental in translating abstract concepts into meaningful action and in preserving their integrity as they diffuse into policy and organisational life.

Hollnagel's ideas have not been without criticism. They have been described as polarising, susceptible to superficial adoption, and lacking a rigorous empirical evidence base [11, 53, 54]. Viewed more broadly, however, these criticisms do not just apply to Hollnagel's work but reflect a persistent problem in safety science regarding the lack of rigorous evidence production [45, 91]. Scholars have often developed compelling theories and analytical tools without specifying testable mechanisms through which improvements in safety are expected to occur. Similar critiques have therefore been directed at other influential approaches, including systems-theoretic models (e.g., STAMP), safety cases, and, particularly in healthcare, the widespread use of root cause analysis as a means of learning from incidents [91-94].

While some of this criticism is justified, it is also partly based on assumptions about system stability and linear causal pathways from intervention to outcome, which rarely hold in complex systems such as healthcare. Outcomes are shaped by multiple interacting and continually shifting conditions, including ageing populations, increasing multimorbidity, more technologically complex interventions, workforce shortages, and the rapid introduction of disruptive digital and AI technologies. Moreover, Hollnagel's contributions, like many safety science ideas, are rarely implemented in isolation, but are embedded within ongoing change processes characterised by negotiation, adaptation, and the absence of clearly bounded start and end points.

To date, the most tangible benefits of Hollnagel's contribution to patient safety have not been reductions in adverse events, but changes in how safety is framed, studied and acted on in

practice. These shifts have started to produce richer understandings of how healthcare systems function, have sparked discussions about potentially superficial and unhelpful safety improvement interventions, and have strengthened the capacity of organisations and healthcare professionals to engage in learning and improvement aligned more closely to the realities of complex clinical work.

Against this backdrop, more appropriate approaches to evaluation might, therefore, be found in realist evaluation, ethnographic or mixed-methods process evaluations, and theory-driven frameworks such as Normalisation Process Theory, which seek to explain how complex interventions become enacted and sustained in practice [95-97]. In healthcare in particular, improvement efforts are often undertaken through small, local initiatives with limited methodological rigour and weak theorisation, making it difficult to determine whether an intervention worked, why it worked, or whether it could reasonably be expected to work elsewhere [25].

The principles and methods associated with Hollnagel remain highly relevant in a healthcare landscape shaped by increasing complexity, interdependence, and technological disruption. They offer valuable perspectives for addressing emerging challenges such as the integration of artificial intelligence into care systems, while also reminding us of the limits of any single approach in the face of systemic and existential risks.

Hollnagel's enduring legacy is not a fixed doctrine, but lies in the combination of conceptual clarity, practical methodology, and community engagement. Together, these have expanded the repertoire of patient safety practice beyond managing safety as a static end state, towards managing safely as a dynamic, collective achievement.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used ChatGPT (OpenAI) and Gemini (Google) to refine the narrative structure and logical flow of the manuscript, assist in the synthesis of theoretical concepts, and help interpret specific peer-review feedback. After using this tool, the authors critically reviewed and edited all content and take full responsibility for the content of the published article.

REFERENCES

- [1] Wears R, Sutcliffe K. Still Not Safe: Patient Safety and the Middle-Managing of American Medicine. Oxford: Oxford University Press; 2019.
- [2] Berwick DM. Controlling Variation in Health Care: A Consultation from Walter Shewhart. Medical care. 1991;29.
- [3] Hollnagel E. The ETTO Principle: Efficiency-Thoroughness Trade-Off. Farnham: Ashgate; 2009.
- [4] Hollnagel E. FRAM, the functional resonance analysis method: modelling complex socio-technical systems: Ashgate Publishing, Ltd.; 2012.
- [5] Shekelle PG, Pronovost PJ, Wachter RM, McDonald KM, Schoelles K, Dy SM, et al. The Top Patient Safety Strategies That Can Be Encouraged for Adoption Now. Annals of internal medicine. 2013;158:365-8.
- [6] Hollnagel E, Wears RL, Braithwaite J. From Safety-I to Safety-II: A White Paper. Published simultaneously by the University of Southern Denmark, University of Florida, USA, Macquarie University, Australia: The Resilient Health Care Net 2015.
- [7] Mannion R, Braithwaite J. False Dawns and New Horizons in Patient Safety Research and Practice. International Journal of Health Policy and Management. 2017;6:685-9.
- [8] Sujan M. A Safety-II Perspective on Organisational Learning in Healthcare Organisations; Comment on “False Dawns and New Horizons in Patient Safety Research and Practice”. International Journal of Health Policy and Management. 2018;7:662-6.
- [9] Lounsbury O, Sujan M. Achieving a restorative Just Culture through the patient safety incident response framework. Journal of Patient Safety and Risk Management. 2023;28:153-5.
- [10] Hollnagel E, Sujan M, Braithwaite J. Resilient Health Care – Making steady progress. Safety Science. 2019;120:781-2.
- [11] Verhagen MJ, de Vos MS, Sujan M, Hamming JF. The problem with making Safety-II work in healthcare. BMJ quality & safety. 2022;31:402-8.
- [12] Leistikow I, Bal RA. Resilience and regulation, an odd couple? Consequences of Safety-II on governmental regulation of healthcare quality. BMJ quality & safety. 2020;29:1-2.
- [13] Wiig S, Aase K, Billett S, Canfield C, Røise O, Njå O, et al. Defining the boundaries and operational concepts of resilience in the resilience in healthcare research program. BMC health services research. 2020;20:330.
- [14] Pickup L, Atkinson S, Hollnagel E, Bowie P, Gray S, Rawlinson S, et al. Blood sampling - Two sides to the story. Applied Ergonomics. 2017;59:234-42.
- [15] Schutijser BCFM, Jongerden IP, Klopotoska JE, Portegijs S, de Bruijne MC, Wagner C. Double checking injectable medication administration: Does the protocol fit clinical practice? Safety Science. 2019;118:853-60.
- [16] Verhagen MJ, de Vos MS, Hamming JF. Taking Morbidity and Mortality Conferences to a Next Level: The Resilience Engineering Concept. Annals of surgery. 2020;272.
- [17] Sujan M, Bilbro N, Ross A, Earl L, Ibrahim M, Bond-Smith G, et al. Failure to rescue following emergency surgery: A FRAM analysis of the management of the deteriorating patient. Applied Ergonomics. 2022;98:103608.
- [18] Hollnagel E. From Safety to Safety: Principles and Practice of Systemic Potentials Management. Abingdon, UK: Routledge; 2025.
- [19] Peerally MF, Carr S, Waring J, Dixon-Woods M. The problem with root cause analysis. BMJ quality & safety. 2016.

- [20] Vincent CA. Analysis of clinical incidents: a window on the system not a search for root causes. *Quality & safety in health care*. 2004;13:242-3.
- [21] Liberati EG, Peerally MF, Dixon-Woods M. Learning from high risk industries may not be straightforward: a qualitative study of the hierarchy of risk controls approach in healthcare. *International Journal for Quality in Health Care*. 2018;30:39-43.
- [22] Halligan D, Janes G, Conner M, Albutt A, Debono D, Carland J, et al. Identifying Safety Practices Perceived as Low Value: An Exploratory Survey of Healthcare Staff in the United Kingdom and Australia. *Journal of Patient Safety*. 2023;19.
- [23] Rae AJ, Provan DJ, Weber DE, Dekker SW. Safety clutter: the accumulation and persistence of 'safety'work that does not contribute to operational safety. *Policy and practice in health and safety*. 2018;16:194-211.
- [24] Lord Darzi A. *Better Health and Care for All: A 10-Point Plan for the 2020s*. London: Institute for Public Policy Research; 2018.
- [25] Dixon-Woods M. How to improve healthcare improvement—an essay by Mary Dixon-Woods. *BMJ (Clinical research ed)*. 2019;367:l5514.
- [26] Vincent C, Amalberti R. *Safer healthcare: Strategies for the real world*. Springer; 2016.
- [27] Bosk CL, Dixon-Woods M, Goeschel CA, Pronovost PJ. Reality check for checklists. *The Lancet*. 2009;374:444-5.
- [28] Berwick D, Fox DM. "Evaluating the Quality of Medical Care": Donabedian's Classic Article 50 Years Later. *The Milbank Quarterly*. 2016;94:237-41.
- [29] Lord Darzi A. *High Quality Care For All: NHS Next Stage Review Final Report*. London 2008.
- [30] Sutcliffe KM, Paine L, Pronovost PJ. Re-examining high reliability: actively organising for safety. *BMJ quality & safety*. 2017;26:248-51.
- [31] Kohn LT, Corrigan JM, Donaldson MS. *To Err Is Human: Building a Safer Health System*. Washington: The National Academies Press; 2000.
- [32] Weick K, Sutcliffe K. *Managing the unexpected: resilient performance in an age of uncertainty*. San Francisco: Jossey Bass; 2007.
- [33] Myers CG, Sutcliffe KM. High reliability organising in healthcare: still a long way left to go. *BMJ quality & safety*. 2022;31:845-8.
- [34] Le Coze JC. Vive la diversité! High Reliability Organisation (HRO) and Resilience Engineering (RE). *Safety Science*. 2019;117:469-78.
- [35] Reason J. *Managing the risks of Organizational Accidents*. Farnham: Ashgate; 1997.
- [36] Reason J. Human error: models and management. *BMJ (Clinical research ed)*. 2000;320:768-70.
- [37] Dekker S. *Patient Safety: A Human Factors Approach*. Boca Raton: CRC Press; 2011.
- [38] Le Coze J-C. Visualising Safety. In: Le Coze JC, editor. *Safety Science Research*. Boca Raton: CRC Press; 2019. p. 151-71.
- [39] Perrow C. *Normal Accidents: Living with High-Risk Technologies*. London: Basic Books; 1984.
- [40] Rasmussen J. Risk management in a dynamic society: a modelling problem. *Safety science*. 1997;27:183-213.
- [41] Hollnagel E, Woods DD, Leveson N. *Resilience Engineering: Concepts and Precepts*. Aldershot: Ashgate; 2006.

- [42] Leveson N. A new accident model for engineering safer systems. *Safety Science*. 2004;42:237-70.
- [43] Woods DD. The theory of graceful extensibility: basic rules that govern adaptive systems. *Environment Systems and Decisions*. 2018;38:433-57.
- [44] Dekker S. *Just culture: balancing safety and accountability*. Farnham: Ashgate; 2012.
- [45] Le Coze JC. The 'new view' of human error. Origins, ambiguities, successes and critiques. *Safety Science*. 2022;154:105853.
- [46] Hollnagel E. RAG - The resilience analysis grid. In: Hollnagel E, Paries J, Woods D, Wreathall J, editors. *Resilience Engineering in Practice*. Farnham: Ashgate; 2011. p. 275-96.
- [47] Patriarca R, Di Gravio G, Woltjer R, Costantino F, Praetorius G, Ferreira P, et al. Framing the FRAM: A literature review on the functional resonance analysis method. *Safety Science*. 2020;129:104827.
- [48] Patriarca R, Leonhardt J, Licu A. Introducing the structured exploration of complex adaptations to learn from operations in an air navigation service provider. In: Leva MC, Patelli E, Podofillini L, Wilson S, editors. *Proceedings of the 32nd European Safety and Reliability Conference, ESREL 2022-Understanding and Managing Risk and Reliability for a Sustainable Future: Research Publishing*; 2022. p. 33-40.
- [49] Hollnagel E. *Safety-I and Safety-II*. Farnham: Ashgate; 2014.
- [50] Hollnagel E. *Safety-II in practice: developing the resilience potentials*. Abingdon: Routledge; 2018.
- [51] Bruner JS. *Making stories: Law, literature, life*: Harvard University Press; 2003.
- [52] Dekker S. *Foundations of safety science: A century of understanding accidents and disasters*: Routledge; 2019.
- [53] Leveson N. *Safety III: A Systems Approach to Safety and Resilience*. Cambridge, MA: MIT; 2020.
- [54] Cooper MD. The Emperor has no clothes: A critique of Safety-II. *Safety Science*. 2022;152:105047.
- [55] Rasmussen J. Human error and the problem of causality in analysis of accidents. *Philosophical Transactions of the Royal Society of London B, Biological Sciences*. 1990;327:449-62.
- [56] Hollnagel E, Woods DD. Cognitive Systems Engineering: New wine in new bottles. *International Journal of Man-Machine Studies*. 1983;18:583-600.
- [57] Reason J. *Managing the risks of organizational accidents*: Routledge; 2016.
- [58] Stanton NA, Margaryan A, Littlejohn A. Editorial: Learning from Incidents. *Safety Science*. 2017;99:1-4.
- [59] Wenger EC, Snyder WM. Communities of practice: The organizational frontier. *Harvard business review*. 2000;78:139-46.
- [60] NHS England. *Patient Safety Incident Response Framework*. London: NHS England; 2022.
- [61] Braithwaite J, Wears R, Hollnagel E. *Resilient Health Care III: Reconciling Work-as-Imagined with Work-as-Done*. Farnham: Ashgate; 2017.
- [62] Wears R, Hollnagel E, Braithwaite J. *The Resilience of Everyday Clinical Work*. Farnham: Ashgate; 2015.
- [63] Braithwaite J, Hollnagel E, Hunte G. *Resilient Health Care: Muddling Through with Purpose, Volume 6*: CRC Press; 2021.

- [64] Braithwaite J, Hollnagel E, Hunte GS. Working Across Boundaries: Resilient Health Care, Volume 5: CRC Press; 2019.
- [65] Hollnagel E. Barriers and Accident Prevention. Aldershot: Ashgate; 2004.
- [66] Alm H, Woltjer R. Patient safety investigation through the lens of FRAM. In: de Waard D, Axelsson A, Berglund M, Peters B, Weikert C, editors. Human Factors: A system view of human, technology and organisation. Maastricht: Shaker Publishing; 2010. p. 153-6.
- [67] Sujan MA, Felici M. Combining Failure Mode and Functional Resonance Analyses in Healthcare Settings. *Computer Safety, Reliability, and Security*. 2012;364-75.
- [68] Salehi V, Veitch B, Smith D. Modeling complex socio-technical systems using the FRAM: A literature review. *Human Factors and Ergonomics in Manufacturing & Service Industries*. 2021;31:118-42.
- [69] Sujan M, Pickup L, de Vos MS, Patriarca R, Konwinski L, Ross A, et al. Operationalising FRAM in Healthcare: A critical reflection on practice. *Safety Science*. 2023;158:105994.
- [70] Damen NL, de Vos MS, Moesker MJ, Braithwaite J, de Lind van Wijngaarden R, Kaplan J, et al. Preoperative Anticoagulation Management in Everyday Clinical Practice: An International Comparative Analysis of Work-as-Done Using the Functional Resonance Analysis Method. *Journal of patient safety*. 2018.
- [71] Patriarca R, Falegnami A, Costantino F, Bilotta F. Resilience engineering for socio-technical risk analysis: Application in neuro-surgery. *Reliability Engineering & System Safety*. 2018;180:321-35.
- [72] Ross A, Sherriff A, Kidd J, Gnich W, Anderson J, Deas L, et al. A systems approach using the functional resonance analysis method to support fluoride varnish application for children attending general dental practice. *Applied Ergonomics*. 2018;68:294-303.
- [73] Slater D, Hollnagel E, MacKinnon R, Sujan M, Carson-Stevens A, Ross A, et al. A systems analysis of the COVID-19 pandemic response in the United Kingdom – Part 1 – The overall context. *Safety Science*. 2022;146:105525.
- [74] Sujan M, Lounsbury O, Pickup L, Kaya GK, Earl L, McCulloch P. What kinds of insights do Safety-I and Safety-II approaches provide? A critical reflection on the use of SHERPA and FRAM in healthcare. *Safety Science*. 2024;173:106450.
- [75] Sujan M, Lounsbury O. Maximising the impact of the FRAM: from analysis to implementation. In: Patriarca R, editor. *Navigating the FRAM: Mastering the Functional Resonance Analysis Method for Modelling Complex Socio-Technical Systems*: Taylor & Francis; 2025.
- [76] Patriarca R. *Navigating the FRAM: Mastering the Functional Resonance Analysis Method for Modelling Complex Socio-Technical Systems*: Taylor & Francis; 2026.
- [77] Burke JR, Downey C, Almoudaris AM. Failure to Rescue Deteriorating Patients: A Systematic Review of Root Causes and Improvement Strategies. *Journal of patient safety*. 2020.
- [78] Karanikas N, Zerguine H. Are the new safety paradigms (only) about safety and sufficient to ensure it? An overview and critical commentary. *Safety Science*. 2024;170:106367.
- [79] Haug CJ, Drazen JM. Artificial Intelligence and Machine Learning in Clinical Medicine. *New England Journal of Medicine*. 2023;388:1201-8.
- [80] Brownstein JS, Rader B, Astley CM, Tian H. Advances in Artificial Intelligence for Infectious-Disease Surveillance. *New England Journal of Medicine*. 2023;388:1597-607.

- [81] Rajpurkar P, Lungren MP. The Current and Future State of AI Interpretation of Medical Images. *New England Journal of Medicine*. 2023;388:1981-90.
- [82] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*. 2019;25:44-56.
- [83] Sujan M, Furniss D, Grundy K, Grundy H, Nelson D, Elliott M, et al. Human factors challenges for the safe use of artificial intelligence in patient care. *BMJ Health & Care Informatics*. 2019;26:e100081.
- [84] Elish MC, Watkins EA. Repairing innovation: a study of integrating AI in clinical care. New York: Data & Society Research Institute; 2020.
- [85] Hilton B, Buhl MD, Korbak T, Irving G. Safety Cases: A Scalable Approach to Frontier AI Safety. *arXiv preprint arXiv:250304744*. 2025.
- [86] Salmon PM, Carden T, Hancock PA. Putting the humanity into inhuman systems: How human factors and ergonomics can be used to manage the risks associated with artificial general intelligence. *Human factors and ergonomics in manufacturing & service industries*. 2021;31:223-36.
- [87] Salmon PM, King BJ, Elstak I, McLean S, Read GJ. Tomorrow's demons: a scoping review of the risks associated with emerging technologies. *Ergonomics*. 2024;1-17.
- [88] Sujan M, Pool R, Salmon P. Eight human factors and ergonomics principles for healthcare artificial intelligence. *BMJ Health & Care Informatics*. 2022;29:e100516.
- [89] Stilgoe J, Owen R, Macnaghten P. Developing a framework for responsible innovation. *Research Policy*. 2013;42:1568-80.
- [90] Shneiderman B. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human-Computer Interaction*. 2020;36:495-504.
- [91] Rae A, Provan D, Aboelssaad H, Alexander R. A manifesto for Reality-based Safety Science. *Safety Science*. 2020;126:104654.
- [92] Kellogg KM, Hettinger Z, Shah M, Wears RL, Sellers CR, Squires M, et al. Our current approach to root cause analysis: is it contributing to our failure to improve patient safety? *BMJ quality & safety*. 2017;26:381-7.
- [93] Leveson N. The use of safety cases in certification and regulation. *Journal of System Safety*. 2011;47.
- [94] Sujan M, Habli I. Safety cases for digital health innovations: can they work? *BMJ quality & safety*. 2021;30:1047-50.
- [95] Dixon-Woods M, Bosk CL, Aveling EL, Goeschel CA, Pronovost PJ. Explaining Michigan: developing an ex post theory of a quality improvement program. *The Milbank quarterly*. 2011;89:167-205.
- [96] May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. *Sociology*. 2009;43:535-54.
- [97] Pawson R, Tilley N. *Realistic Evaluation*. London: Sage; 1997.