The problem with making Safety-II work in healthcare

Merel J Verhagen, Marit S de Vos, Mark Sujan, Jaap F Hamming

INTRODUCTION

Patient safety is typically assessed by the frequency of adverse events or incidents, which means we seek to determine safety by its absence rather than its presence. The Safety-II perspective aspires to overcome this paradox by bringing into focus situations where safety is actually present, that is, in everyday work that usually goes well. Central to Safety-II is the notion that, in complex systems such as healthcare, safety is a consequence of collective efforts to adapt to dynamic conditions and uncertainty, rather than the natural state of a system where nothing untoward happens. This type of thinking has been met with significant interest and enthusiasm in healthcare, because it feeds increased appreciation for the fact that healthcare workers continuously ensure that most patients receive safe and high-quality care in challenging circumstances. However, despite its appeal and potential, significant challenges remain for the potential enrichment of quality and safety practices in healthcare.

WHAT SAFETY-II BRINGS TO THE TABLE

Healthcare as a complex sociotechnical system requiring non-linear approaches

The Safety-II perspective was developed in recognition of the complexity and inherent variability in modern systems such as healthcare. This is opposed to the more traditional view that systems are largely deterministic and predictable, with simple cause-and-effect relationships. In this view, systems are usually decomposed into their constituent components, and safety is thought to arise from the reliability of those components, and from barriers put in place to mitigate potential failures. Methods such as root cause analysis (RCA) or failure mode and effects analysis are examples aligned to this type of thinking. Yet, experiences in healthcare with such traditional methods often show oversimplification, limiting the practical utility and the subsequent contribution to quality improvement.

New directions in safety science started challenging these assumptions, arguing that risk arises in complex systems not necessarily from the failure of individual components, but from the structure of such systems and their functional interconnectedness. Building on this, the field of resilience engineering (and resilience healthcare) was developed as a paradigm to understand how people cope with complexity and uncertainty to achieve success in dynamic conditions.

The notion of Safety-II is based on resilience engineering and was introduced as a term to distinguish and contrast the two perspectives on safety (ie, Safety-I and Safety-II), along with their underlying assumptions. Healthcare is characterised as a complex adaptive system, with emergent properties resulting from a labyrinth of interactions, making it non-linear, dynamic and largely intractable. For instance, the workflow of an emergency department (ED) is designed through work instructions that are based on the assumption that there is adequate capacity to meet the demand. In practice,
however, the ED functions more like an open, self-regulating system, which needs to balance mismatches in demand and capacity arising from inter-relationships beyond the ED, such as varying numbers of patients arriving in ambulances, variability in the availability of hospital beds and diagnostics capacity and availability of opportunities to discharge patients into the community. ED clinicians continuously attempt to align demand and capacity in a dynamic fashion by adjusting the way they work, for example, by ordering tests early, by switching between patients as they wait for test results, by prioritising which patients can and cannot wait or by using additional spaces such as an assessment unit to create further capacity. It is largely impossible to specify the behaviour of such dynamic systems in detail for every situation, and unsuccessful outcomes can occur even though none of the system’s elements (eg, equipment or people) failed. This latter notion is illustrated by Woods and Branlat’s description of how complex adaptive systems fail: for example, through decompensation (exhausting the system’s adaptive capacity; eg, during peaks in demand) or working at cross-purposes (adaptations that are meaningful locally, but inadequate at a system level; eg, some referral and gatekeeping behaviours between ED and hospital clinicians). From a Safety-II perspective, systems may fail due to the aggregation and amplification of everyday variability (‘functional resonance’); a non-linear phenomenon. Therefore, Safety-II suggests to move from linear (eg, RCA) to non-linear methods, such as the functional resonance analysis method (FRAM), to study the interactions that make up everyday work processes.

**Safety as an ongoing capacity rather than freedom from error**

Traditional patient safety management is often reactive and failure oriented, responding to events or risks perceived as unacceptable. Interventions then commonly focus on standardisation, checks and barriers to make failures less likely and to guard against their consequences. In the Safety-II view, the focus of safety management shifts from the exclusive consideration of adverse events, failures and ways to prevent these towards understanding and strengthening the abilities that serve to continuously create safety in everyday practice. Among other things, many naturally developed checks and informal working practices are present in healthcare settings, such as in an anecdotal example about an experienced secretary using a personal checklist to guide the patients’ peri-operative anticoagulation management, even though not formally being responsible for this task. With this practice of double checking, the secretary was able to detect treatment plans that deviated from standard practice, in which case a quick phone call could be made to check with the physicians whether this was an unintended oversight or a deliberate decision. Accordingly, Safety-II approaches study in a non-normative way the role of workers and systems in creating and maintaining safety, such as through seemingly hidden acts to support thoroughness. This sustaining of required operations under both expected and unexpected conditions is referred to as resilient performance, which in turn is considered to be enabled by four cornerstone abilities (‘resilience abilities or potentials’): learning from past experience, monitoring the system’s performance and changes in its environment, anticipating potential developments and responding to actual ones. Box 1 provides a simplified example of resilience abilities in the delivery of care during the COVID-19 pandemic. More succinctly, resilience in healthcare could be defined as ‘the capacity to adapt to challenges and changes at different system levels, to maintain high-quality care’. This definition highlights also that resilient performance results from activities across the system (not just the frontline), and it broadens the consideration of the goals of the system to include aspects of quality beyond safety, which have to be managed together rather than in isolation.

**Variability is inevitable in healthcare and the source of both success and failure**

Variable conditions and performance require dynamic trade-offs and adjustments, which will always be approximate rather than precise. Some will turn out to be inadequate in hindsight, especially when interacting in unanticipated ways. Yet, these ‘approximate adjustments’ are essential for everyday work because competing demands and inherent uncertainty cannot be completely designed out. Therefore, these adjustments are considered to be the underlying source for both success and failure. In other words, the belief

---

**Box 1  Simplified example of resilience abilities in a system’s delivery of care during the COVID-19 pandemic**

- **Ability to monitor**: looking out for the number of patients with COVID-19 admitted and the number of staff off sick to apprehend the impact and requirements of the current situation.
- **Ability to respond**: changing the schedules of staff members and moving non-urgent services into the community to create extra in-hospital capacity.
- **Ability to anticipate**: the creation of long-term predictions about the impact of the expected increase in infections on the capacity of a hospital.
- **Ability to learn**: a reflection on the hospital’s response to the preceding waves of infections, and the identification of what worked and why, to subsequently contemplate on plans to deal with future waves and challenge underlying beliefs about what ‘good’ care looks like in these circumstances.
that things go wrong for different reasons than they go right is rejected in the Safety-II perspective. With that, the perspective supports a non-normative approach that helps to overcome inherent limitations of retrospective investigations of events, such as outcome and hindsight bias. Greater attention for good work delivered daily by healthcare workers, as well as their struggles, will serve to identify and empower existing good practices, and will likely provide a novel perspective on the issues driving clinician burnout, and contribute to better morale and positive culture, which have been linked to positive patient outcomes.

Learning from everyday work rather than adverse events
To understand the dynamic circumstances and variability in everyday practice, and the required (resilience) behaviours of our systems to deal with these, these need to be observed when actually present, that is, in everyday work. Looking only at adverse events hinders such learning. In addition, the consideration of everyday work is less controversial and less threatening than investigating situations where patients came to harm or where health professionals became second victims; when analysing harmful situations, there is always the risk that issues of accountability and culpability get interwoven with the desire to learn from the event.

An idealised view of how work should be carried out in practice, referred to as work-as-imagined, will never foresee all local contingencies, and therefore will often be different from actual work-as-done. Work-as-done provides a realistic view of how work is carried out in everyday practice, illustrating that when work-as-imagined is used as the norm when studying an adverse event, this fails to acknowledge the actual habitual activities of clinicians, managers and supporting staff. Therefore, the study of work-as-done becomes central to both safety and quality improvement and to understanding adverse events.

BRINGING SAFETY-II INTO PRACTICE IN HEALTHCARE: CHALLENGES AND OPPORTUNITIES
Challenges for safety management: how to balance Safety-II and Safety-I
A theoretical challenge with Safety-II is brought up by Leveson stating that Safety-II introduces a concept as a strawman that does not actually exist, namely Safety-I. It could be argued that safety science is rich and diverse, with different schools of thought and methods that cannot be lumped together under one label, and doing so therefore represents inappropriate generalisation and oversimplification. Safety science has been informed by insights from disciplines over the decades, such as, for instance, organisation science (eg, high-reliability organisations) or human factors (eg, cognitive systems engineering). Resilience engineering and, by extension, Safety-II draw on concepts introduced with many of these earlier developments. Owing to the relative novelty of the field, the theory underpinning Safety-II lacks sufficient practical guidance on whether and how existing safety models and management approaches might coexist and be used in harmony with Safety-II. Does the Safety-II perspective replace Safety-I, include and extend it, or do both complement one another?

In line with other commentators, it seems prudent to suggest that successful safety management will rely on requisite variety developed from a diversity of insights and perspectives, with different healthcare processes and contexts demanding different strategies. Safety-I and Safety-II represent two distinct, yet complementary views of safety and can coexist alongside each other depending on the situational requirements. Some activities in healthcare are relatively repeatable, linear and controllable, for which approaches concerned with constraining variation are likely to be effective; for example, process-driven approaches such as standardisation and checklists. As seen in super specialised clinics, systems with high volume and relatively low circumstance variation will thrive on well-designed routines and strict work instructions, and tools focusing more on standardisation and protocol adherence. At the same time, the COVID-19 pandemic demonstrated that few areas are free of the need for adaptation. For the many more variable and less predictable processes in healthcare, such as those involving larger multidisciplinary teams or patients with complex comorbidities, application of additional safety barriers tends to add even more complexity, potentially worsening rather than improving safety. The risks of overly constraining performance have long been recognised, including by Reason. Instead, these situations will benefit from a more non-linear approach that is goal oriented, rather than process oriented, thereby providing professionals with flexibility to adjust to meet the dynamic conditions.

Challenges for event analyses: how to move from descriptions to interventions
Much of the applied Safety-II literature has been concerned with describing in detail work-as-done across different healthcare settings. A frequently used method is FRAM, which enables the analyst to study a system based on the interaction of variability between functions. FRAM can be useful for looking at everyday work and the inherent complexity of interactions collaboratively with the people who do the work, and for highlighting to stakeholders that simplistic and reductionist interventions will by themselves likely not be successful in improving safety. Even though this method is not intended to ‘diagnose and fix’ the system, it may help to identify leverage points for system improvement. However, FRAM provides little guidance about how an organisation subsequently moves from the description of variability to monitoring and
controlling it effectively. The language sometimes used is reminiscent of the criticised Safety-I thinking; for example, the FRAM book suggests to ‘dampen’ variability that is getting out of control. How this can be achieved, beyond regress to established methods that reduce variability by constraining behaviour (i.e., Safety-I), is underspecified. Variability is not a means to an end or something to be measured, controlled and managed per se. The in-depth study of variability in everyday work provides insights into how people and systems deal with complexity and uncertainty, how people make dynamic trade-offs and how safety, or success, is created through proactive resilient processes (i.e., the cornerstone abilities). This understanding can then feed the team’s initiatives to foster and strengthen, for instance, informal learning, or multidisciplinary collaboration (e.g., identifying and adapting a treatment strategy depending on the patient’s clinical condition), rather than to focus on barriers and controls. In order to be accepted as a legitimate tool for safety improvement, future studies should work on ways to illustrate how methods, such as FRAM, can be an effective means for quality improvement, by establishing their usefulness in applying Safety-II thinking to work, supporting systems resilience and the conditions for successful work.

Challenges for learning: how to learn from everyday work rather than negative outcomes

At first glance, learning from everyday work (i.e., a seemingly boundless number of situations) might appear challenging for healthcare organisations that already struggle to investigate and learn from the large number of adverse events. In addition, regulatory expectation puts the focus for investigation firmly on adverse events. A solution may be to build on and extend existing practices. As was also indicated by systems’ thinkers such as Reason at the time, the scope of adverse event investigations can be broadened to develop a rich understanding of work-as-done beyond the specifics of an event. The fact that understanding the wider system is a key element of the Safety-II perspective helps to reinforce this view that has been included in some Safety-I theories or models, but remained underexposed in practice. This can be operationalised, for example, by investigating everyday work for urgent topics that have been identified by traditional safety investigations, such as preoperative medication management. Another example may be to focus on tensions, contradictions and trade-offs in a work routine such as intravenous infusions, and subsequently explore in what way a specific infusion incident relates to this. The question becomes ‘what was so ordinary about this case’, instead of ‘what was so extraordinary’, focusing on what this tells us about work-as-done. Using this different approach shifts the focus from what went wrong and what barriers might have failed towards the trade-offs and adaptations people need to make and how these might be supported. Lastly, the importance of reflecting on everyday practice could be emphasised more strongly, seeking ways to ensure that such reflection is routinely carried out in practice. For example, morbidity and mortality conferences can be adapted to also discuss cases with expected outcomes rather than only negative cases, thereby enabling learning from how success is created in these cases despite perhaps similar risks and challenges.

Challenges with lack of validation studies and evidence for effectiveness

A major challenge lies on progressing from conceptual thinking to evidence-based actionable insights on how exactly the Safety-II perspective can improve healthcare and under what conditions. Examples of studies bringing this view to practice effectively, in terms of developing, testing and evaluating interventions, seem scarce. Even though FRAM analyses have been carried out in a wide range of settings, these studies often struggle to demonstrate that the studied processes have actually been improved. In this, a first piece of evidence for the Safety-II perspective could lie in how an understanding of the perspective might actually suggest alternative insights into some widespread patient safety projects, as outlined in box 2. Irrespective of whether or not contributors to these projects were intentionally working with Safety-II, its perspective might suggest other perceptions of the mechanisms by which successful patient safety programmes may work; for instance, through building the resilient performance of teams and systems, or reconciling gaps between work-as-imagined and work-as-done.

The scientific evidence base of the Safety-II perspective might be weakened by the fact that researchers in the safety science domain are traditionally using research designs that are less suited to the production of rigorous evidence of effectiveness. Here, healthcare might have a relatively stronger tradition of underpinning interventions with evidence, for example, frameworks for the evaluation of complex interventions and realist evidence synthesis approaches. The challenge for Safety-II researchers and practitioners is to articulate explicitly the (in)formal theories or logic models for proposed interventions, so that claims about how to improve safety and under what circumstances can be critically appraised. In addition, there is a wider problem about how improvements in safety should be measured or assessed from a Safety-II perspective. It appears paradoxical to use Safety-I measures, such as adverse event or protocol adherence rates, as success indicators for Safety-II approaches. Arguably, if Safety-II is concerned with the strengthening of abilities such as learning or responding, then assessment approaches from, for example, the education...
Box 2 Examples of patient safety initiatives approached from the Safety-II perspective

- **Failure to rescue**: The concept of ‘failure to rescue’ concerns mortality following a major complication after surgical treatment. Acknowledging that a successful rescue depends on a team’s ability to timely recognise and efficiently manage a severe complication, the work in this field led to the development of systems to detect and respond to patient deterioration. These efforts can be seen as ways to support successful rescue rather than prevent failure; supporting the resilient performance of the system through building the capacity to monitor using scoring systems, anticipate through briefings and respond with special medical emergency teams. Rather than focusing on compliance with scoring systems, the focus lies on supporting the staff in dynamic, ambiguous and often time-pressured situations to bring additional resources when needed (‘slack’).

- **Surgical safety checklist**: The surgical safety checklist serves as a means to strengthen perioperative consistency of care and communication. Checking the team members’ conformity to perioperative protocols (eg, the administration of antibiotics or sterility of instruments) seems to align solely with the Safety-I perspective of preventing malfunction by discussing issues known to cause risk. Yet, from a Safety-II perspective, the checklist can be considered a tool to collectively consider multidisciplinary goals, risks to be expected and resources available, thereby serving as a ‘pre-brief’ for resilient performance. These perspectives will affect how people engage with the checklist, that is, as a compliance tool or as a tool that strengthens team performance, which reflects that checklists are a complex social intervention.

- **The Michigan Keystone project**: With the intention to decrease catheter-related bloodstream infections, the successful Michigan Keystone project consisted of a large-scale intervention that included, among other things, the introduction of a central-line insertion cart for supplies, a checklist and a consecutive shift in culture, encompassing a shared sense of mission and empowerment of nurses. While the technical interventions, such as a checklist, seemed central factors for success at first glance, an alternative view is that these interventions required local efforts to align work-as-imagined and work-as-done, which is also a key concept in Safety-II. Teams had to redesign their daily practices or had to customise the checklist to fit their specific context. This also fits within the Safety-II perspective, in that it makes work-as-done central to quality improvement.

or social science domain might be more appropriate. For instance, workplace-based learning is concerned with workplace competencies, which are similar in nature to resilience abilities, in as far as they are also potentials, that is, they can be developed and deployed in changing and uncertain contexts. Competency development includes reflection, and the assessment of competencies is often based on portfolios, which can be used to develop existing competencies and construct new ones. Suitably adapted to the organisational level, such approaches could potentially be used to document, reflect on and improve resilience abilities, for instance learning, by means of a qualitative, reflective and participative process, for example, the ‘change laboratory’ based on Engeström’s theory of expansive learning. Future studies, using qualitative or mixed-methods designs, might develop and use new measurement and formative evaluation frameworks, focusing on the dynamic interplay between everyday work and the resilience abilities, how these abilities could be enhanced and what this ultimately does for practice.

**CONCLUSION**

Safety-II offers a distinct perspective on patient safety by accentuating the importance of understanding the uncertainties and trade-offs in everyday work, with its successes and failures. Its potential value lies in providing a deeper understanding of the abilities that serve to create safety, and ways to strengthen these in closer alignment with the complexity of everyday ‘work-as-done’. However, challenges with the concept’s credibility, practicality and scientific evidence base hamper its adoption in healthcare. A lack of guidance exists for how to balance Safety-II with current safety practices, and how exactly to operationalise its key concepts, such as learning from everyday work. A less antagonistic debate of Safety-I versus Safety-II, further development of practical approaches to specify how to move from description to practical intervention, and research designs that provide robust evidence about the effectiveness of Safety-II, could support the integration of Safety-II thinking into existing safety management practices in healthcare.

**Twitter** Mark Sujan @MarkSujan

**Contributors** MV, MdV, MS and JH conceived and discussed the concept of this perspective. MV wrote the first draft of the manuscript, and all authors contributed to reviewing it. All authors read and approved the final manuscript. JH acted as guarantor.

**Funding** This work was funded in part by the National Institute for Health Research (NIHR) [Programme Grant for Applied Research NIHR200868]. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** There are no data in this work.
REFERENCES

2 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.
41 Lawton R. It Ain’t What You Do (But the Way That You Do It): Will Safety II Transform the Way We Do Patient Safety?
The problem with...


