From Safety-I to Safety-II: A White Paper

Professor Erik Hollnagel
University of Southern Denmark, Institute for Regional Health Research (IRS), Denmark
Center for Quality, Region of Southern Denmark

Professor Robert L Wears
University of Florida Health Science Center Jacksonville, United States of America

Professor Jeffrey Braithwaite
Australian Institute of Health Innovation, Macquarie University, Australia

First published in 2015 by The Authors

Printed and bound by:

National Library of Congress

Cataloguing-in-Publication data:

Suggested citation:
Hollnagel E., Wears R.L. and Braithwaite J. From Safety-I to Safety-II: A White Paper. The Resilient Health Care Net: Published simultaneously by the University of Southern Denmark, University of Florida, USA, and Macquarie University, Australia.

ISBN: TBA

From Safety-I to Safety-II: A White Paper

Executive summary

The publication of the IOM report *To Err is Human* in 2000 served as a catalyst for a growing interest in improving the safety of health care. Yet despite decades of attention, activity and investment, improvement has been glacially slow. Although the rate of harm seems stable, increasing demand for health services, and the increasing intensity and complexity of those services (people are living longer, with more complex co-morbidities, and expecting higher levels of more advanced care) imply that the number of patients harmed while receiving care will only increase, unless we find new and better ways to improve safety.

Most people think of safety as the absence of accidents and incidents (or as an acceptable level of risk). In this perspective, which we term Safety-I, safety is defined as a state where as few things as possible go wrong. A Safety-I approach presumes that things go wrong because of identifiable failures or malfunctions of specific components: technology, procedures, the human workers and the organisations in which they are embedded. Humans—acting alone or collectively—are therefore viewed predominantly as a liability or hazard, principally because they are the most variable of these components. The purpose of accident investigation in Safety-I is to identify the causes and contributory factors of adverse outcomes, while risk assessment aims to determine their likelihood. The safety management principle is to respond when something happens or is categorised as an unacceptable risk, usually by trying to eliminate causes or improve barriers, or both.

This view of safety became widespread in the safety critical industries (nuclear, aviation, etc.) between the 1960s and 1980s. At that time performance demands were significantly lower than today and systems simpler and less interdependent.

It was tacitly assumed then that systems could be decomposed and that the components of the system functioned in a bimodal manner—either working correctly or incorrectly. These assumptions led to detailed and stable system descriptions that enabled a search for causes and fixes for malfunctions. But these assumptions do not fit today's world, neither in industries nor in health care. In health care, systems such as an intensive care or emergency setting cannot be decomposed in a meaningful way and the functions are not bimodal, neither in detail nor for the system as a whole. On the contrary, everyday clinical work is—and must be—variable and flexible.

Crucially, the Safety-I view does not stop to consider why human performance practically always goes right. Things do not go right because people behave as they are supposed to, but because people can and do adjust what they do to match the conditions of work. As systems continue to develop and introduce more complexity, these adjustments become increasingly important to maintain acceptable performance. The challenge for safety improvement is therefore to understand these adjustments—in other words, to understand how performance usually goes right in spite of the uncertainties, ambiguities, and goal conflicts that pervade complex work situations. Despite the obvious importance of things going right, traditional safety management has paid little attention to this.

Safety management should therefore move from ensuring that 'as few things as possible go wrong' to ensuring that 'as many things as possible go right'. We call this perspective Safety-II; it relates to the system's ability to succeed under varying conditions. A Safety-II approach assumes that everyday performance variability provides the adaptations that are needed to respond to varying conditions, and hence is the reason why things go right. Humans are consequently seen as a resource necessary for system flexibility and resilience. In Safety-II the purpose of investigations changes to become an understanding of how things usually go right, since that is the basis for explaining how things occasionally go wrong. Risk assessment tries to understand the conditions where performance variability can become difficult or impossible to monitor and control. The safety management principle is to facilitate everyday work, to anticipate developments and events, and to maintain the adaptive capacity to respond effectively to the inevitable surprises (Finkel 2011).

In light of increasing demands and growing system complexity, we must therefore adjust our approach to safety. While many adverse events may still be treated by a Safety-I

approach without serious consequences, there is a growing number of cases where this approach will not work and will leave us unaware of how everyday actions achieve safety. This may have unintended consequences because it unintentionally degrades the resources and procedures needed to make things go right.

The way forward therefore lies in combining the two ways of thinking. While many of the existing methods and techniques can continue to be used, the assimilation of a Safety-II view will also require new practices to look for what goes right, to focus on frequent events, to maintain a sensitivity to the possibility of failure, to wisely balance thoroughness and efficiency, and to view an investment in safety as an investment in productivity. This White Paper helps explains the key differences between, and implications of, the two ways of thinking about safety.

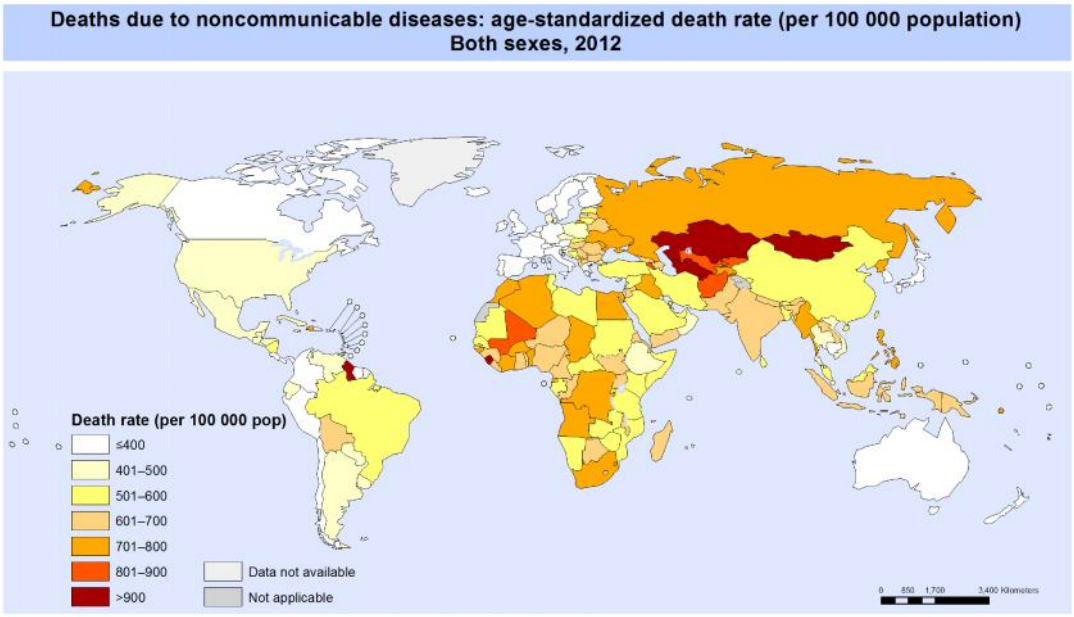
Background: The World Has Changed

To say that the world has changed is not just a phrase. It explains the intention of this White Paper and is also a teaser for the reader's thoughts.

It is a truism that the world we live in has become more complex and interdependent and that this development continues to accelerate. It applies to the ways we work and to how we live our daily lives. This is perhaps most easily seen in the ways we communicate, both in the development from bulky telephones to elegant smartphones and in the change from person-to-person interaction to social networks and media.

Similar changes have taken place in health care in the past 40 years. The World Health Organization (WHO) indicates that worldwide, non-communicable diseases (NCDs) have now become the leading causes of mortality compared with earlier eras.

NCDs include heart disease, stroke, cancer, chronic respiratory diseases, and diabetes. The map below shows the deaths due to non-communicable diseases, worldwide per 100,000 population, age-standardised between 2000 and 2012. This epidemic is a huge burden on patients, their families and communities. The number of emergency visits, GP attendances, general and ICU admissions has grown internationally in both absolute numbers and on a per capita basis to treat these diseases. There seems no end in sight to the increasing trend. At the same time, new threats (surprises) emerge (eg, Ebola, Marburg, etc), and ramify throughout the networked world in unexpected and unpredictable ways.



Source: WHO 2014 at

http://gamapserver.who.int/gho/interactive_charts/ncd/mortality/total/atlas.html

By way of response, the use of high-technology diagnostic and therapeutic interventions (such as CT or MRI scanning, ultrasound, minimally invasive surgery, joint replacements, and open heart surgery) has gone from being experimental and used only in tertiary or quaternary centres for the most difficult of cases, to become routine components in the armamentarium of major hospitals worldwide. The sheer numbers of patients, and the increasingly complex socio-technical environment in which care takes place, constitute a considerable challenge for stakeholders, patients, clinicians, managers, policymakers, regulators, and politicians.

The costs of health care associated with this technological capacity has grown even faster, to the point that it is typically the largest single component of GDP in most western countries, and the fastest growing in virtually all countries. This rate of growth is widely considered to be unsustainable.

In the early days of this revolution in health care, adverse events were considered the unfortunate but inevitable price to be paid for medical advances. When safety became a cause célèbre around 2000, there were therefore few established approaches to deal with patient safety issues. The obvious response was to adopt apparently successful solutions from other industries. These focused largely on component failures, and the human

component—the front-line health care worker—was considered just another fallible element. Thus, the common model that informed early patient safety efforts, and that has settled into the current 'orthodoxy' of patient safety, was based on linear cause-and-effect, component failure models. Just as any disease must have a cause that can be diagnosed and treated, so will any adverse event have a cause that can be found and fixed. Simple linear models, such as Heinrich's (1931) Domino Model that is at the heart of Root Cause Analysis, later supplemented by composite linear models such as Reason's Swiss Cheese Model, were soon adopted as the basic safety tools in health care. Few people noticed that the very same models were being progressively challenged by industrial safety outside healthcare as inadequate to the newer, more complex working environments.

During the second half of the 20th century the focus of industrial safety efforts shifted from technological problems to human factors problems and finally to problems with organisations and safety culture. Unfortunately, few of the models used to analyse and explain accidents and failures developed in a similar way. The result is that safety thinking and safety practices in many ways have reached an impasse. This was the primary driver for the development of resilience engineering in the first decade of this century (e.g., Hollnagel, Woods & Leveson, 2006). Resilience engineering acknowledges that the world has become more complex, and that explanations of unwanted outcomes of system performance therefore can no longer be limited to an understanding of cause-effect relations described by linear models.

Safety-I

To most people safety means the absence of unwanted outcomes such as incidents or accidents. Because the term 'safety' is used and recognised by almost everyone, we take for granted that others understand it the same way that we do and therefore rarely bother to define it more precisely. The purpose of this White Paper is to do just that; and to explore the implications of two different interpretations of safety.

Safety is generically defined as the system quality that is necessary and sufficient to ensure that the number of events that can be harmful to workers, the public, or the environment is acceptably low. The WHO, for instance, defines patient safety as "the prevention of errors and adverse effects to patients associated with health care."

Historically speaking, the starting point for safety concerns has been the occurrence of accidents (actual adverse outcomes) or recognised risks (potential adverse outcomes). Adverse outcomes—things that go wrong—have usually been explained by pointing to their presumed causes, and the response has been to either eliminate or contain them. New types of accidents have similarly been accounted for by introducing new types of causes—either relating to technology (e.g., metal fatigue), to human factors (e.g., workload, 'human error'), or to the organisation (e.g., safety culture). Because this has been effective in providing short-term solutions, we have through the centuries become so accustomed to explaining accidents in terms of cause-effect relations, that we no longer notice it. And we cling tenaciously to this tradition, although it has become increasingly difficult to reconcile with reality. Unfortunately, seeing deficiencies in hindsight does nothing to explain the generation or persistence of those deficiencies.

To illustrate the consequences of defining safety by what goes wrong, consider Figure 1. Here the thin red line represents the case where the (statistical) probability of a failure is 1 out of 10,000. But this also means that one should expect things to go right 9,999 times out of 10,000—corresponding to the green area. (In health care, the failure rate is in the order of a few percent, up to 10 percent, in hospitalized patients, depending on how they are counted; but the principle is the same—things go right much more often than they go wrong.)

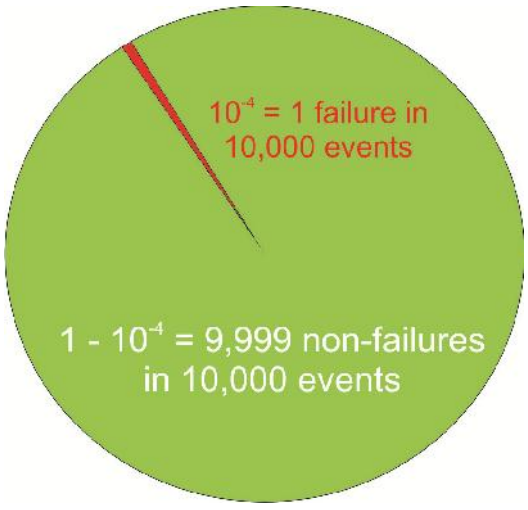


Figure 1: The imbalance between things that go right and things that go wrong

Safety-I efforts focus on what goes wrong, and this focus is reinforced in many ways. Regulators and authorities require detailed reports on accidents, incidents, and even so-

called unintended events, and special agencies, departments, and organisational roles are dedicated to scrutinising adverse outcomes. Numerous models claim they can explain how things go wrong and a considerable number of methods are offered to find the failed component and address the causes. Adverse event and incident data are collected in large databases. Adverse events and incidents are described and explained in thousands of papers, books, and debated in specialised national and international conferences. The net result is a deluge of information both about how things go wrong and about what must be done to prevent this from happening. The general solution is known as 'find and fix': look for failures and malfunctions, try to find their causes, and then eliminate those causes or introduce barriers, or both.

The situation is quite different for the events that go right. Despite their crucial importance, they usually receive little attention in safety management activities such as risk identification, safety assurance and safety promotion. There are no requirements from authorities and regulators to look at what works well and therefore few agencies and departments do that. Possible exceptions are audits and surveys, which may include a focus on strengths, and the occasional 'good news' reviews commissioned by politicians or CEOs to spin positive media stories. However, on the whole, data are difficult to find, there are few models, even fewer methods, and the vocabulary is scant in comparison to that for what goes wrong. There are few books and papers, and practically no meetings. Looking at how things go right also clashes with the traditional focus on failures, and therefore receives little encouragement. This creates a serious problem because we cannot make sure things go right just by preventing them from going wrong. Patently, we also need to know how they go right.

Safety-I promotes a bimodal view of work and activities, according to which acceptable and adverse outcomes are due to different modes of functioning. When things go right it is because the system functions as it should and because people work-as-imagined; when things go wrong it is because something has malfunctioned or failed. The two modes are assumed to be distinctly different, and the purpose of safety management is naturally to ensure that the system remains in the first mode and never ventures into the second (see Figure 2).

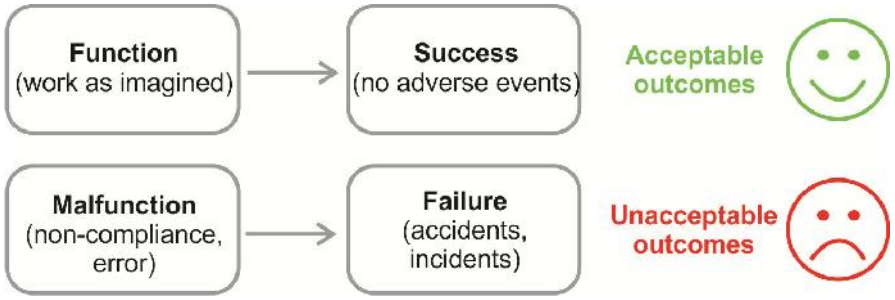


Figure 2: Safety-I assumes that things that go right and things that go wrong happen in different ways

In Safety-I, the starting point for safety management is either that something has gone wrong or that something has been identified as a risk. Both cases use the 'find and fix' approach: in the first case, by finding the causes and then developing an appropriate response, and in the second, by identifying the hazards in order to eliminate or contain them. Another solution is to prevent a transition from a 'normal' to an 'abnormal' state (or malfunction), regardless of whether this is due to a sudden transition or a gradual 'drift into failure'. This is accomplished by constraining performance in the 'normal' state, by reinforcing compliance and by eliminating variability (see Figure 3). A final step is to check whether the number of adverse outcomes (hospital infections, medication errors, or medical device failures, etc.) becomes smaller. If they are, it is taken as proof that the efforts worked as intended.

It is not only wise but also necessary to assess just how effective this mode of safety has been. In the following, Safety-I will be characterised by looking at its manifestations (phenomenology), its underlying mechanisms (aetiology), and its theoretical foundations (ontology).

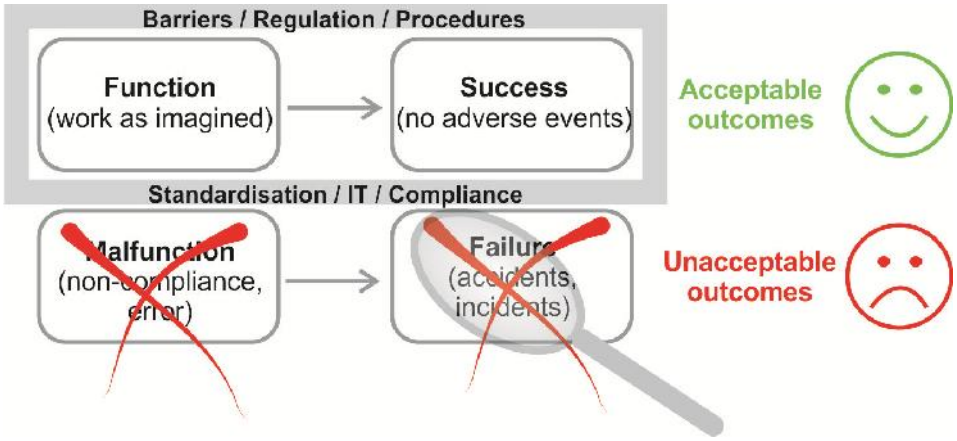


Figure 3: Safety by elimination and prevention

The Manifestations of Safety-I: Looking at what goes wrong

The definition of Safety-I means that the manifestations of safety are the adverse outcomes. A system (e.g., a general practice, a pharmacy, a care facility, or a hospital) is said to be unsafe if there is more than the occasional adverse outcome or if the risk is seen as unacceptable; similarly, it is said to be safe if such outcomes occur rarely or not at all, or if the risk is seen as acceptable. This is, however, an indirect definition because safety is being defined by its opposite, by what happens when it is absent rather than when it is present. A curious consequence is that we analyse and try to learn from situations where, by definition, there was a lack of safety.

Another curious consequence is that the level of safety is inversely related to the number of adverse outcomes. If many things go wrong, the level of safety is said to be low; but if few things go wrong, the level of safety is said to be high. In other words, the more manifestations there are, the less safety there is and vice versa. A perfect level of safety means that there are no adverse outcomes, hence nothing to measure. This unfortunately makes it very difficult, if not impossible, to demonstrate that efforts to improve safety have worked, hence very difficult to argue for continued resources.

To help describe the manifestations, various error typologies of adverse outcomes are available, ranging from the simple (omission-commission) to the elaborate (various forms of 'cognitive error' and violations or non-compliance). Note that these typologies often hide a troublesome confusion between error as outcome (manifestation) and error as cause.

The 'Mechanisms' of Safety-I

The mechanisms of Safety-I are underpinned by the assumptions about how things happen that are used to explain or make sense of the manifestations. The generic mechanism of Safety-I is the *causality credo*—a globally predominant belief that adverse outcomes (accidents, incidents) happen because something goes wrong, hence that they have causes that can be found and treated. While it is obviously reasonable to assume that consequences are preceded by causes, it is a mistake to assume that the causes are trivial or that they can always be found.

The *causality credo* has through the years been expressed by many different accident models. The strong version of the *causality credo* is the assumption about root causes, as expressed by root cause analysis. While this kind of simple linear thinking was probably adequate for the first part of the 20th century, the increasingly complicated and intractable

socio-technical systems that developed in the last half—and especially since the 1970s—required more intricate and more powerful mechanisms. The best of these is the Swiss Cheese Model, which explains adverse outcomes as the result of a combination of active failures and latent conditions. Other examples are TRIPOD (Reason et al., 1989), AcciMap (Rasmussen & Svedung, 2000), and STAMP (Leveson, 2004). Yet in all cases the *causality credo* allows the analysis to reason backwards from the consequences to the underlying causes. But as Reason (1997) noted, "the pendulum may have swung too far in our present attempts to track down possible errors and accident contributions that are widely separated in both time and place from the events themselves." The increasing complexity of these models has led to the somewhat puckish thought that the 'Swiss Cheese Model has passed its sell-by date' (Reason, Hollnagel & Paries 2006).

The Foundation of Safety-I

The foundation of Safety-I represents the assumptions about the nature of the world that are necessary and sufficient for the mechanisms to work. The foundation of Safety-I implies two important assumptions. One is that systems are decomposable into their constituent parts. The other is that systems and their parts either function correctly, or not—that they are bimodal.

Systems are Decomposable

We know that we can build systems by putting things together (e.g., complicated instruments such as a CT scanner or a surgical robot, or complicated socio-technical systems such as a hospital populated with people and equipment) and carefully combining and organising their components. That's the normal way we create systems.

The first assumption is that this process can be reversed and that we can understand systems by decomposing them into meaningful constituents (see Figure 4). We do have some success with decomposing technological systems to find the causes of accidents—medical device failures in the operating theatre, for example. We also assume that we can decompose 'soft systems' (people in organisations) into their constituents (departments, agents, roles, stakeholders, groups, teams). And we finally assume that the same can be done for tasks and for events, partly because of the seductive simplicity of the time-line (this event happened after that event, and thus the first event 'caused' it). But we are wrong in all cases.

Functioning is Bimodal

It is also assumed that the 'components' of a system can be in one of two modes, either functioning correctly or failing (malfunctioning), possibly embellished by including various degraded modes of operation. System components are usually designed or engineered to provide a specific function and when that does not happen, they are said to have failed, malfunctioned, or become degraded. While this reasoning is valid for technological systems and their components, it is not valid for socio-technical systems—and definitely not for human and organisational components, to the extent that it is even meaningless to use it.

While the two assumptions (decomposability and bimodality) make it convenient to look for causes and to respond by 'fixing' them, they also lead to system descriptions and scenarios with illusory tractability and specificity, and quantification with illusory precision. They are therefore insufficient as a basis for safety management in the world of today.

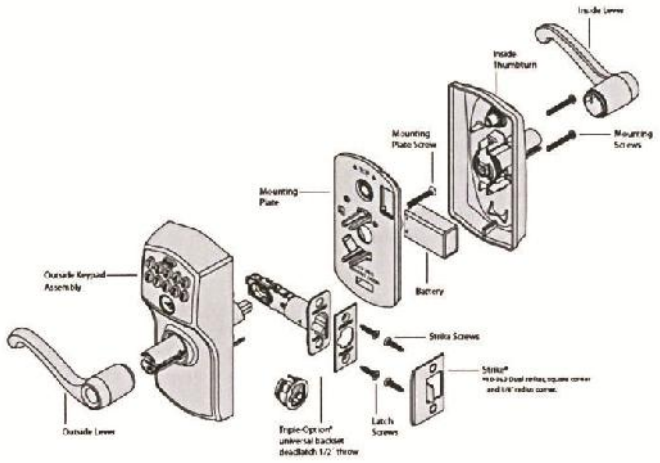


Figure 4: A decomposable system

The Changing World of Health Care

The Ever-Changing Demands on Work, Safety and Productivity

Safety-I is based on a view of safety that was developed roughly between 1965 and 1985 in industrial safety and imported into patient safety years later. Industrial systems in the 1970s were relatively simple when compared with today's world. The dependence on information technology was limited (mainly due to the size and the immaturity of IT itself), which meant that support functions were relatively few, relatively simple, and mostly independent of one another. The level of integration (e.g., across sub-systems and sectors) was low, and

it was generally possible to understand and follow what went on. Support systems were loosely coupled (independent) rather than tightly coupled (interdependent). Safety thinking therefore developed with the following assumptions:

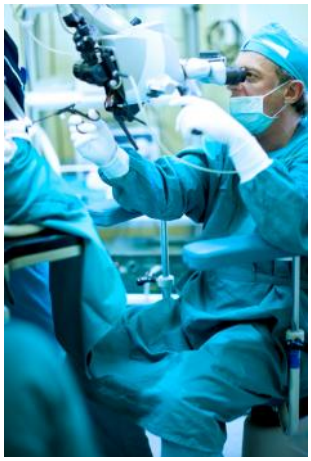
- Systems and places of work are well-designed and correctly maintained.
- Procedures are comprehensive, complete, and correct.
- People at the sharp end (in health care, those on the clinical front line) behave as they are expected to, and as they have been trained to. (They work as they are supposed or imagined to.)
- Designers have foreseen every contingency and have provided the system with appropriate response capabilities. Should things go completely wrong, the systems can degrade gracefully because the sharp end staff can understand and manage the contingencies—even those the designers could not.

While these assumptions were probably never completely correct, they were considered reasonable in the 1970s. But they are not reasonable today, and safety based on these premises is inappropriate for the world as it is in the 2010s.

Health care has since the 1990s regrettably adopted these assumptions rather uncritically, even though health care in 1990 showed little resemblance to industrial workplaces in the 1970s. The situation has by no means improved, since health care in 2015 is vastly different from health care in 1990. Despite that, the assumptions can still be found in the basis for current patient safety efforts.

The 1970's

The 1990's



Present day

Rampant Technological Developments

Like most industries, health care is subject to a tsunami of diverse changes and improvements. Some changes come from well-meant attempts to replace 'fallible' humans with 'infallible' technology, while others are a response to increased performance demands or political expediency. In most countries, ambitious safety targets have been set by national administrations with little concern for whether the targets are meaningful or even practically possible. For example in the US, President Clinton endorsed the IOM's 2000 goal of a 50% reduction in errors in five years, saying anything less would be irresponsible. (Such safety targets also raise the interesting question of whether one can measure an increase in safety by counting how many fewer things go wrong.)

Another disturbing trend is the growing number of cases where problems are selected based on just one criterion: whether they are 'solvable' with a nice and clean technological solution at our disposal. This has two major consequences. One is that problems are attacked and solved one by one, as if they could be dealt with in isolation. The other is that the preferred solution is technological rather than socio-technical, probably because non-technical solutions are rarely 'nice and clean'.

The bottom line of these developments is that few activities today are independent of each other—in health care and elsewhere—and that these mutual dependencies are only going to increase. Functions, purposes, and services are already tightly coupled and the couplings will only become tighter. Consider, for instance, the key WHO action areas targeting patient safety: hand hygiene, and safe surgery using checklists; and others, involving reporting and learning systems, implementing 'solutions', spreading best practice change models ('High 5s'), knowledge management, eliminating central line-associated bloodstream infections, and designing and implementing new checklist applications. While each target may seem plausible, pursuing them as individual strategies risks the emergence of unintended consequences. A change to these will affect others in ways which are non-trivial, not necessarily salutary, and therefore difficult to comprehend. This clashes with the assumptions of Safety-I, which means that any solution based on Safety-I thinking can make things worse.

In consequence of rampant technological developments, of the widespread faith in nice and clean technological solutions, and of the general unwillingness to be sufficiently thorough up-front in order to be efficient later, our ideas about the nature of work and the nature of safety must be revised. We must accept that systems today are increasingly

intractable. This means that the principles of functioning are only partly known (or in an increasing number of cases, completely unknown), that descriptions are elaborate with many details, and that systems are likely to change before descriptions can be completed, which means that descriptions will always be incomplete.

The consequences are that predictability is limited during both design and operation, and that it is impossible precisely to prescribe or even describe how work should be done. Technological systems can function autonomously as long as their environment is completely specified and as long as there is no unexpected variability. But these conditions cannot be established for socio-technical systems. Indeed, in order for the technology to work, humans (and organisations) must provide buffer functionality to absorb excessive variability. People are not a problem to be solved or standardised: they are the adaptive solution.

The Reasons Why Things Work—Again

Because the health systems of today are increasingly intractable, it is impossible to provide a complete description of them or to specify what clinicians should do even for commonly occurring situations. Since performance cannot be completely prescribed, some degrees of variability, flexibility, or adaptivity are required for the system to work. People who contribute such intelligent adjustments are therefore an asset without which the proper functioning would be impossible.

Performance adjustments and performance variability are thus both normal and necessary, and are the reason for both acceptable and unacceptable outcomes. Trying to achieve safety by constraining performance variability will inevitably affect the ability to achieve desired outcomes as well and therefore be counterproductive. For example, standardising approaches by insisting that a clinical guideline on a common medical complaint such as headache or asthma—all fifty or more pages of them—must be slavishly read and everything in them adopted on every occasion when a patient with that condition presents in the Emergency Department, is not just impossible, but leaves almost no time for the actual care to be provided.

Similarly, mandating over 2,000 health department policies (the number that are technically in operation in some publicly funded health systems) and asserting they must be used continuously to guide people's everyday work would lead to systems shut-down. Thus rather than looking for ways in which something can fail or malfunction and documenting detailed procedures, we should try to understand the characteristics of everyday performance variability.

Work-As-Imagined and Work-As-Done

It is an unspoken assumption that work can be completely analysed and prescribed and that Work-As-Imagined therefore will correspond to Work-As-Done. But Work-As-Imagined is an idealized view of the formal task environment that disregards how task performance must be adjusted to match the constantly changing conditions of work and of the world. Work-As-Imagined describes what should happen under normal working conditions. Work-As-Done, on the other hand, describes what actually happens, how work unfolds over time in complex contexts.

One reason for the popularity of the concept of Work-As-Imagined is the undisputed success of Scientific Management Theory (Taylor, 1911). Introduced at the beginning of the 20th century, Scientific Management had by the 1930s established time-and-motion studies as a practical technique and demonstrated how a breakdown of tasks and activities could be used to improve work efficiency. It culminated in the factory production line.

Scientific Management used time and motion studies combined with rational analysis and synthesis to find the best method for performing any particular task that workers then would carry out with proper inducement. Scientific Management thus provided the theoretical and practical foundation for the notion that Work-As-Imagined was a necessary and sufficient basis for Work-As-Done. (Safety was, however, not an issue considered by Scientific Management.) This had consequences both for how adverse events were studied and for how safety could be improved. Adverse events could be understood by looking at the components, to find those that had failed, such as in root cause analysis. And safety could be improved by carefully planning work in combination with detailed instructions and training. These beliefs can be found in the widespread tenets held about the efficacy of

procedures and the emphasis on compliance. In short, safety can be achieved by ensuring that Work-As-Done is made identical to Work-As-Imagined.

But the more intractable environments that we have today means that Work-As-Done differs significantly from Work-As-Imagined. Since Work-As-Done by definition reflects the reality that people have to deal with, the unavoidable conclusion is that our notions about Work-As-Imagined are inadequate if not directly wrong. This constitutes a challenge to the models and methods that comprise the mainstream of safety engineering, human factors, and ergonomics. It also challenges traditional managerial authority. A practical implication of this is that we can only improve safety if we get out from behind our desk, out of meetings, and into operational and clinical environments with operational and clinical people.

Today's work environments require that we look at everyday clinical work or Work-As-Done rather than Work-As-Imagined, hence at systems that are real rather than ideal (Wears, Hollnagel & Braithwaite, 2015). Such systems perform reliably because people are flexible and adaptive, rather than because the systems have been perfectly thought out and designed or because people do precisely what has been prescribed.

Humans are therefore no longer a liability and performance variability is not a threat. On the contrary, the variability of everyday performance is necessary for the system to function, and is the reason for both acceptable and adverse outcomes. Because all outcomes depend on performance variability, failures cannot be prevented by eliminating it; in other words, safety cannot be managed by imposing constraints on normal work.

The way we think of safety must correspond to Work- As-Done and not rely on Work-As-Imagined. Safety-I begins by asking why things go wrong and then tries to find the assumed causes to make sure that it does not happen again—it tries to re-establish Work-As-Imagined. The alternative is to ask why things go right (or why nothing went wrong), and then try to make sure that this happens again.

Safety-II

In the normal course of clinical work, doctors, nurses and allied health staff perform safely because they are able to adjust their work so that it matches the conditions. In tractable and well-engineered systems (such as aviation, mining and manufacturing—but also, e.g., pharmaceutical production), the need for adjustments will be small. In many cases there is also the option of deferring or delaying operations when circumstances become

unfavourable, such as in cases where flights get cancelled due to weather or a mechanical problem can temporarily close the company. Sometimes, the entire system can shut down, as it did after 9/11 in 2001 and when the Icelandic volcano Eyjafjällajökull erupted in April and May, 2010.

Health care is by its very nature often intractable, which means that performance adjustments are needed for the system to function. In many health care situations, the precariousness of the circumstances also make it impossible to delay or defer treatment of patients, even if working conditions are bad (Wears & Perry, 2006).

Given the uncertainty, intractability, and complexity of health care work, the surprise is not that things occasionally go wrong but that they go right so often. Yet as we have seen, when we try to manage safety, we focus on the few cases that go wrong rather than the many that go right. But attending to rare cases of failure attributed to 'human error' does not explain why human performance practically always goes right and how it helps to meet health care goals. Focusing on the lack of safety does not show us which direction to take to improve safety.

The solution to this is surprisingly simple: instead of only looking at the few cases where things go wrong, we should look at the many cases where things go right and try to understand how that happens. We should acknowledge that things go right because clinicians are able to adjust their work to conditions rather than because they work as imagined. Resilience engineering acknowledges that acceptable outcomes and adverse outcomes have a common basis, namely everyday performance adjustments (see Figure 5).

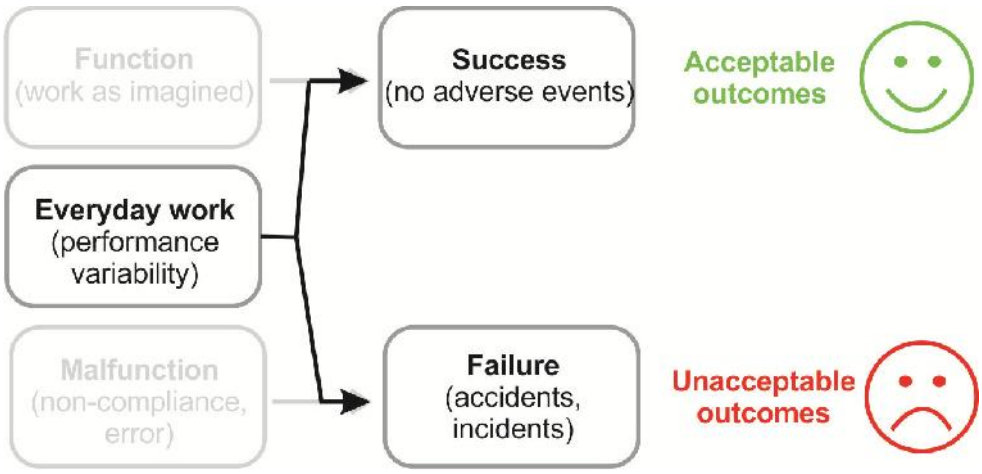


Figure 5: Things that go right and things that go wrong happen in the same way

Because many different work situations today are intractable, it is impossible to prescribe what should be done in any detail except for the most trivial situations. The reason why people nevertheless are able to work effectively is that they continually adjust their work to current conditions—including what others do or are likely to do. As health care systems continue to expand both vertically and horizontally and as their intractability continues to grow, these adjustments become increasingly important for effective performance and therefore present both a challenge and an opportunity for safety management.

According to this view we should avoid treating failures as unique, individual events, and rather see them as an expression of everyday performance variability. Excluding exceptional activities, it is a safe bet that something that goes wrong will have gone right many times before—and will go right many times again in the future. Understanding how acceptable outcomes occur is the necessary basis for understanding how adverse outcomes happen. In other words, when something goes wrong, we should begin by understanding how it (otherwise) usually goes right, instead of searching for specific causes that only explain the failure (see Figure 6). Adverse outcomes are more often due to combinations of known performance variability that usually is seen as irrelevant for safety, than to distinct failures and malfunctions.

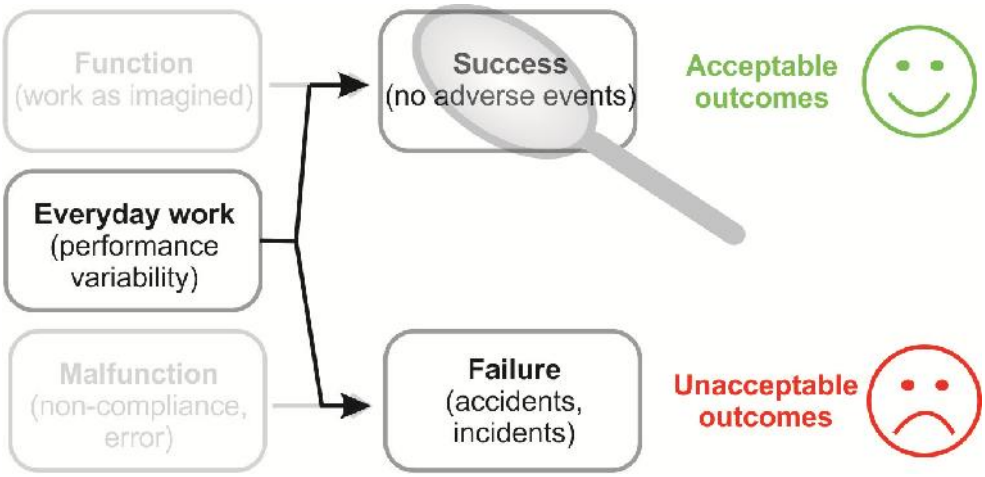


Figure 6: The basis for safety is understanding the variability of everyday performance

Work situations are increasingly intractable, despite our best intention to avoid that. One of the reasons for this is ironically our limited ability to anticipate the consequences of design changes or other interventions—both the intended consequences and the unintended side effects. This problem was addressed many years ago in a discussion of automation, where Bainbridge (1983) pointed out that "the designer who tries to eliminate the operator still leaves the operator to do the tasks which the designer cannot think how to automate". This argument applies not only to automation design but also to work specification and workplace design in health care in general. The more complicated a work situation is, the larger the uncertainty about details will be. And clinical work is hugely complex, and requires high levels of discretion and professional judgement to tailor appropriate care to the circumstances of patients with multiple morbidities.

The premises for safety management in today's complex clinical settings, then, can be summarised as follows:

- Systems and clinical work cannot be decomposed in a meaningful way (there are no natural 'elements' or 'components').
- System functions are not bimodal, separated into 'functioning' or 'malfunctioning,' but everyday performance is—and must be—flexible and variable.
- Outcomes emerge from human performance variability, which is the source of both acceptable and adverse outcomes.
- While some adverse outcomes can be attributed to failures and malfunctions, others are best understood as the result of coupled performance variability.

In consequence of this, the definition of safety should be changed from 'avoiding that something goes wrong' to 'ensuring that everything goes right'. Safety-II is the system's ability to function as required under varying conditions, so that the number of intended and acceptable outcomes (in other words, everyday activities) is as high as possible. The basis for safety and safety management must therefore be an understanding of why things go right, which means an understanding of everyday activities.

Ensuring that as much as possible goes right, in the sense that everyday clinical work achieves its stated purposes, cannot rely on responding to failures since that will only correct what has already happened. Safety management must also be proactive, so that interventions are made before something happens and can affect how it will happen or even prevent something from happening. A main advantage is that early responses, on the whole, require a smaller effort because the consequences of the event will have had less time to develop and spread. And early responses can obviously save valuable time.

In the following, we will characterise Safety-II in more detail. We will first look at its theoretical foundations, then its underlying mechanisms, and finally its manifestations.

The Foundation of Safety-II: Performance Variability rather than Bimodality

In contrast to Safety-I, Safety-II is based on the principle that performance adjustments are ubiquitous and that performance not only always *is* variable but that it *must* be so. This means that it is impossible as well as meaningless to characterise components in terms of whether they succeed or fail, or function or malfunction. The variability should, however, not be interpreted negatively, as in 'performance deviations', 'violations', and 'non-compliance'. On the contrary, the ability to make performance adjustments is an essential human contribution to work, without which only the most trivial activity would be possible.

The 'Mechanisms' of Safety-II: Emergence rather than Causality

Since performance adjustments and performance variability constitute the foundation of Safety-II, it follows that the mechanisms cannot rely on causality and linear propagations of causes and effects. Although it is still common to attribute a majority of adverse outcomes to a breakdown or malfunctioning of components and normal system functions, there is a growing number of cases where that is not possible. In such cases the outcome is said to be emergent rather than resultant. This does not make it impossible to explain what happened, but the explanation will be of a different nature. The meaning of emergence is

not that something happens 'magically,' but that it happens in a way that cannot be explained using the principles of decomposition and causality. This is typically the case for systems that in part or in whole are intractable.

The way we usually explain how something has happened is by tracing back from effect to cause, until we reach the root cause—or run out of time and money. This can be illustrated by a representation such as the fish bone diagram shown in Figure 7.

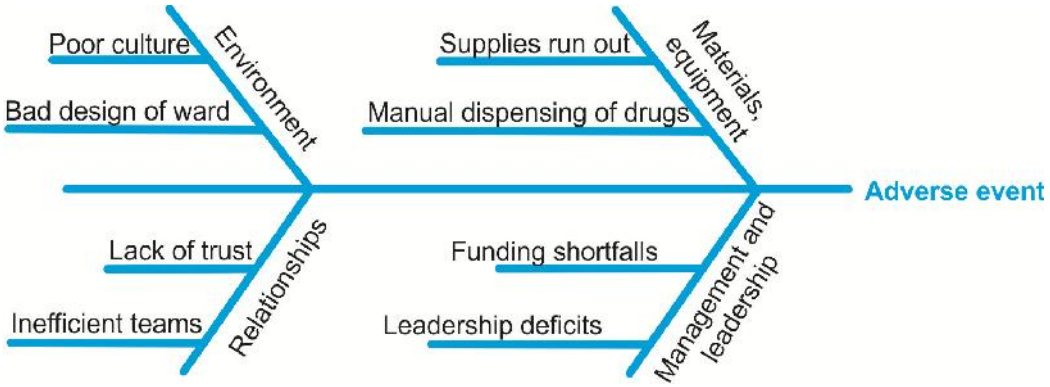


Figure 7: Fish bone diagram using linear logic to track an adverse event

When something goes wrong, there will be an observable change of something. (Otherwise we could not know that anything had happened.) The outcome may be a wrong site surgery, a surgical infection, or a diagnostic failure. Safety-I assumes that the causes are real and the purpose of accident and incident investigation is to trace the developments backwards from the observable outcome to the efficient cause. The causes are also 'real' in the sense that they can be associated with components or functions that in some way have 'failed,' where the 'failure' is either visible after the fact or can be deduced from the facts. Similarly, risk assessment projects the developments forward from the efficient cause(s) to the possible outcomes. They often start with a database of incidents, and assess the risk of another similar thing happening now.

In the case of emergence, the observed (final) outcomes are of course also observable or 'real', but the same is not necessarily true for what brought them about. The outcomes may, for instance, be due to transient phenomena or conditions that only existed at a particular point in time and space. The nurse had a headache; or a doctor's daughter was getting married and everyone was celebrating the event; or local politics were antagonistic that day because two adjoining departments were arguing over resource allocations. These conditions may, in turn, have emerged from other transient phenomena. (see Figure 8). The

'causes' are thus reconstructed (or inferred) rather than found. They may therefore be impossible to eliminate or contain in the usual manner, but it may still be possible to control the conditions that brought them into existence, provided we understand how work normally is done.

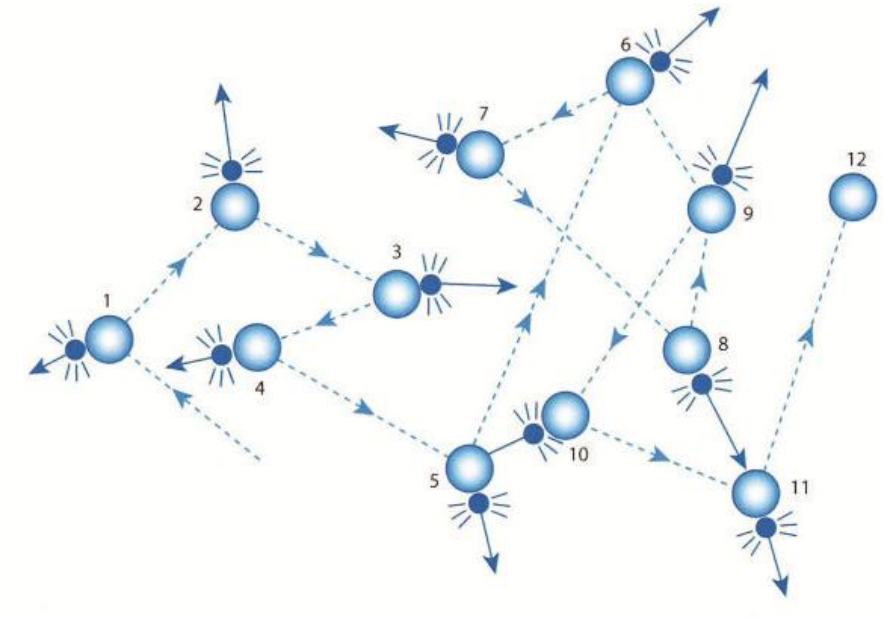


Figure 8: Transient phenomena and emergence

The Manifestations of Safety-II: Things that go right

The definition of Safety-II means that the manifestations are all the possible outcomes, as illustrated by Figure 9, and especially the typical or high frequency outcomes that are usually ignored by safety management. A system is still deemed to be unsafe if adverse outcomes occur yet it is more important to understand how it is safe when they do not occur: safety is consequently defined by what happens when it is present, rather than by what happens when it is absent, and is thus directly related to the high frequency, acceptable outcomes. In other words, the more of these manifestations there are, the higher the level of safety is and vice versa. This makes it possible to demonstrate that efforts to improve safety have worked, hence easier to argue for continued resources. (It also resolves the possible conflict between safety and productivity, but that is another matter.)

To help describe the manifestations of Safety-II, few typologies are currently available. Even though things go right all the time, we fail to notice this because we become used to it. Psychologically, we take it for granted. But since everyday performance is unexceptional, it can be explained in relatively simple terms. For instance everyday performance can be described as performance adjustments that serve to create or maintain required working conditions, that compensate for a lack of time, materials, information, etc., and that try to avoid conditions that are known to be harmful to work. And because everyday performance variability is ubiquitous, it is easier to monitor and manage.

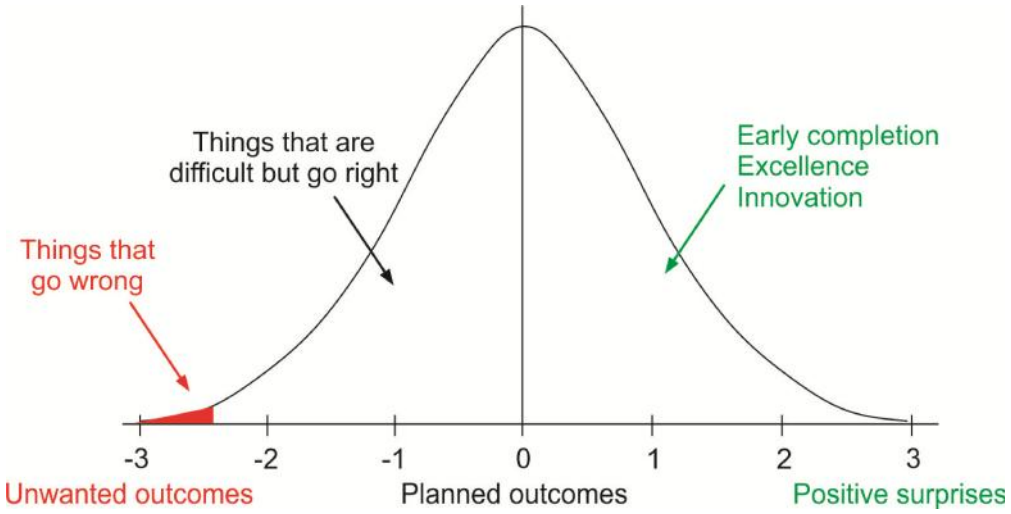


Figure 9: Event probability and safety focus

The Way Ahead

The main reason for juxtaposing Safety-I and Safety-II is to draw attention to the consequences of basing safety management on one or the other. The essential differences are summarised in the following table.

Table 1: Overview of Safety-I and Safety-II

	Safety-I	Safety-II
Definition of safety	That as few things as possible go wrong.	That as many things as possible go right.
Safety management principle	Reactive, respond when something happens or is categorised as an unacceptable risk.	Proactive, continuously trying to anticipate developments and events.

View of the human factor in safety management	Humans are predominantly seen as a liability or hazard. They are a problem to be fixed.	Humans are seen as a resource necessary for system flexibility and resilience. They provide flexible solutions to many potential problems.
Accident investigation	Accidents are caused by failures and malfunctions. The purpose of an investigation is to identify the causes.	Things basically happen in the same way, regardless of the outcome. The purpose of an investigation is to understand how things usually go right as a basis for explaining how things occasionally go wrong.
Risk assessment	Accidents are caused by failures and malfunctions. The purpose of an investigation is to identify causes and contributory factors.	To understand the conditions where performance variability can become difficult or impossible to monitor and control.

What clinicians do in everyday work situations is usually a combination of Safety-I and Safety-II. The specific balance depends on many things, such as the nature of the work, the experience of the people, the organisational climate, management and patient pressures, and patients' disease and other characteristics. Everybody knows that prevention is better than cure, but the conditions may not always allow prevention to play its proper role.

It is a different matter when it comes to the ranks of health care policymakers, and management and regulatory activities. Here the Safety-I view dominates. One reason is that the primary objective of policymakers, managers and regulators historically has been to make sure that patients or the public are not subjected to harm. Another reason is that these levels are removed in time and space from the actual operation of the systems and services, and therefore have limited opportunity to observe or experience how work actually is done. A third reason is that it is much simpler to count the few events that fail than the many that do not—in other words an efficiency-thoroughness trade-off (Hollnagel, 2009). (It is also—wrongly—assumed to be easier to account for the former than for the latter.)

While day-to-day activities at the sharp end rarely are reactive only, the pressure in most work situations is to be efficient rather than thorough. This makes it less legitimate to spend time and efforts to digest and communicate experiences, since this is seen as being non-productive—at least in the short term. Effective safety management nevertheless

requires that some effort is spent up front to think about how work is done, to provide the necessary resources, and to prepare for the unexpected. The pressure towards efficiency—such as the typical hospital's goal, to see more patients for the same cost, by standardising treatments as 'packages', and by shortening average length of stay—makes this more difficult to achieve.

It can be difficult to manage safety proactively for the myriad of small-scale events that constitute everyday work situations. Here, things may develop rapidly and unexpectedly, there are few leading indicators, and resources may often be stretched to the limit. The pace of work leaves little opportunity to reflect on what is happening and to act strategically. Indeed, work pressures and external demands often necessitate opportunistic solutions that force the system into a reactive mode. To get out of this—to switch from a reactive to a proactive mode—requires a deliberate effort. While this may not seem to be affordable in the short term, it is unquestionably a wise investment in the long term.

It is somewhat easier to manage safety proactively for large-scale events because they develop relatively slowly—even though they may begin abruptly. (An example would be a hurricane or major storm that causes multiple injuries and disrupts infrastructures, or a pandemic.) There are often clear indicators for when a response is needed. The appropriate responses are furthermore known, so that preparations can be made ahead of time.

It is important to emphasise that Safety-I and Safety-II represent two complementary views of safety rather than two incompatible or conflicting approaches. Many of the existing practices can therefore continue to be used, although possibly with a different emphasis. But the transition to a Safety-II view will also include some new types of practices, as described in the following.

Transitioning to Safety-II

Look for What Goes Right

A key message is: look at what goes right as well as what goes wrong, and learn from what works as well as from what fails. Indeed, do not wait for something bad to happen but try to understand what actually takes place in situations where nothing out of the ordinary seems to happen. Things do not go well because people simply follow the procedures and work as imagined. Things go well because people make sensible adjustments according to the demands of the situation. Finding out what these adjustments are and trying to learn from them is at least as important as finding the causes of adverse outcomes.

When something goes wrong, such as an infectious outbreak, a communication breakdown, a medication failure, or a wrong patient-wrong procedure problem, it is unlikely to be a unique event. It is rather something that has gone well many times before and that will go well many times again. It is necessary to understand how such everyday activities go well—how they succeed—in order to understand how they might fail. From a Safety-II view they do not fail because of some kind of error or malfunction, but because of unexpected combinations of everyday performance variability.

The difference between a Safety-I and a Safety-II view is illustrated by Figure 10. Safety-I focuses on events at the tails of the normal distribution, and especially events on the left tail that represent accidents. Such events are easy to see because they are rare and because the outcomes differ from the usual. They are, however, difficult to explain—the attractiveness of root causes and linear models notwithstanding. Because they are rare and because they are difficult to understand, they are also difficult to change and manage.

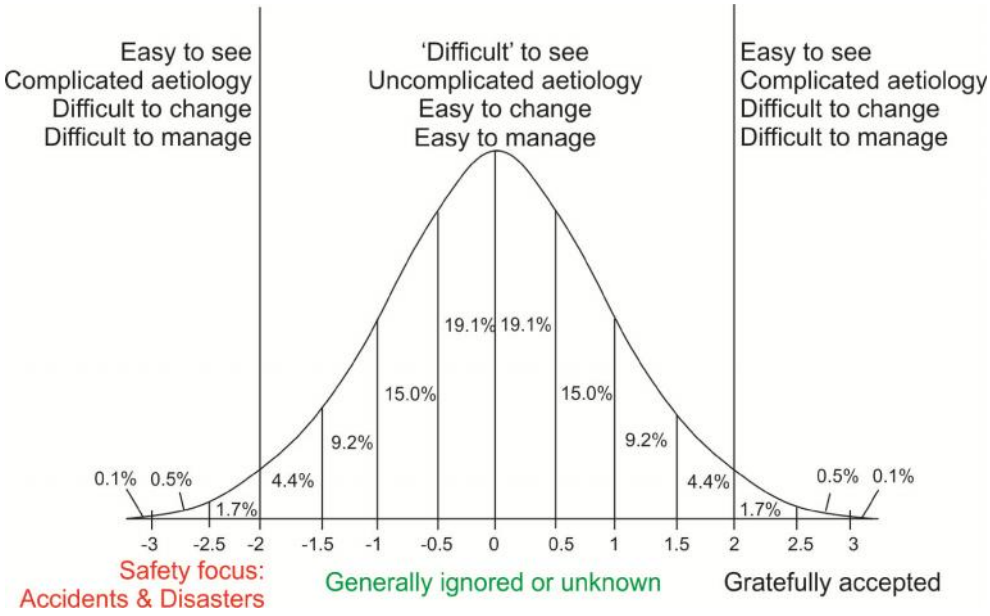


Figure 10: Relation between event probability and ease of perception

Safety-II focuses on events in the middle of the distribution. These are 'difficult' to see, but only because we habitually ignore them in our daily activities. The 'logic' seems to be that if something works, then why spend more time on it? But the fact of the matter is that they usually do not work in the way that we assume, and that Work-As-Done may be significantly different from Work-As-Imagined. The events in the middle can be understood and explained in terms of the mutual performance adjustments that provide

the basis for everyday work. Because they are frequent, because they are small scale, and because we can understand why and how they happen, they are easy to monitor and manage. Interventions are focused and limited in scope (because the subject matter is uncomplicated), and it is therefore also easier—although not necessarily straightforward—to anticipate what both the main and the side effects may be.

There is of course an evolutionary benefit in not paying attention (or too much attention) to the usual as long as it does not harm us and as long as environment is stable. But in our society, the environment is no longer stable and the benefit is therefore illusory.

The work environment, and therefore also work itself, is increasingly unpredictable. This means that the routines that work well today may not work well tomorrow, and that it therefore is important to pay attention to how they work. This is the kind of thoroughness that enables us to be efficient when the time comes to make changes, and to make them rapidly.

Focus on Frequent Events

A second message is: look for what happens regularly and focus on events based on their frequency rather than their severity. Many small improvements of everyday performance may count more than a large improvement of exceptional performance.

The investigation of incidents is often limited by time and resources. There is therefore a tendency to look at incidents that have serious consequences and leave the rest for some other time—that never comes. The unspoken assumption is that the potential for learning is proportional to the severity of the incident or accident.

This is obviously a mistake. While it is correct that more money is saved by avoiding one large scale accident than one small scale accident, it does not mean that the learning potential is greater as well. In addition, the accumulated cost of frequent but small-scale incidents may easily be larger. And since small but frequent events are easier to understand and easier to manage (cf., above), it makes better sense to look to those than to rare events with severe outcomes.

Remain Sensitive to the Possibility of Failure

A third message is: although Safety-II focuses on things that go right, it is still necessary to keep in mind that things can also go wrong and to 'remain sensitive to the possibility of failure'. But the 'possible failure' is not just that something may malfunction as in a Safety-I view, but also that the intended outcomes may not be obtained, i.e., that we fail to ensure that things go right.

Making sure that things go right requires an ongoing concern for whatever works well, not only to ensure that it continues to do so but also to counteract tendencies to employ a confirmation bias or to focus on the most optimistic outlook or outcomes.

In order to remain sensible to the possibility of failure, it is necessary to create and maintain an overall comprehensive view of work—both in the near term and in the long term. This can anticipate and thereby prevent the compounding of small problems or failures by pointing to small adjustments that can dampen potentially harmful combinations of performance variability. Many adverse outcomes stem from the opportunistic aggregation of short-cuts in combination with inadequate process supervision or hazard identification. Being sensible to what happens, to the ways in which it can succeed as well as the ways in which it can fail, is therefore important for the practice of Safety-II.

Be Thorough as well as Efficient

A fourth message is: do not privilege efficiency over thoroughness—or at least, not unduly. If most or all the time is used trying to make ends meet, there will little or no time to consolidate experiences or understand Work-As-Done. It must be legitimate within the organisational culture to allocate resources—especially time—to reflect, to share experiences, and to learn. If that is not the case, then how can anything ever improve?

Efficiency in the present cannot be achieved without thoroughness in the past. And in the same way, efficiency in the future cannot be achieved without thoroughness in the present, i.e., without planning and preparations. While being thorough may be seen as a loss of productivity (efficiency) in the present, it is a necessary condition for efficiency in the future. In order to survive in the long run it is therefore essential to strike some kind of balance.

Investing in Safety, the Gains from Safety

A fifth and final message is: making things go right is an investment in safety and productivity. Spending more time to learn, think, and communicate is usually seen as a cost. Indeed, safety itself is seen as a cost. This reflects the Safety-I view, where an investment in safety is an investment in preventing something from happening. We know the costs, just as when we buy insurance. But we do not know what we are spared, since this is both uncertain and unknown in size. In the risk business, the common adage is 'if you think safety is expensive, try an accident'. And if we calculate the cost of a major accident, such as Betsy

Lehman, the cancer patient who received four times the already-high prescribed dose of the chemotherapy drug cyclophosphamide over a four-day period (Altman 1995), or Willie King, the 51 year old diabetic who had the wrong leg amputated (Clary 1995), almost any investment in safety is cost-effective. However, since we cannot prove that the safety precautions actually are or were the reason why an accident did not happen, and since we cannot say when an accident is likely to happen, the calculation is biased in favour of reducing the investment. (This is something that is typically seen in hard times.)

In Safety-I, safety investments are seen as costs, or are non-productive. Thus if an investment is made and there are no accidents, it is seen as an unnecessary cost. If there are accidents, it is seen as a justified investment. If no investments are made and there are no accidents, it is seen as a justified saving. While if accidents occur, this is seen as bad luck or bad judgement.

In Safety-II, an investment in safety is seen as an investment in productivity, because the definition—and purpose—of Safety-II is to make as many things go right as possible. Thus if an investment is made and there are no accidents, everyday performance will still be improved. If there are accidents, the investment will again be seen as justified. If no investments are made and there are no accidents, performance may remain acceptable but will not improve. While if accidents occur, it is seen as bad judgement.

Conclusion

Since the socio-technical systems on which health care depends continue to become more and more complicated, it seems clear that staying with a Safety-I approach will be inadequate in the long run and in the short run as well. Taking a Safety-II approach should therefore not be a difficult choice to make.

Yet the way ahead lies not in a replacement of Safety-I by Safety-II, but rather in a combination of the two ways of thinking (see Figure 11). It is still the case that the majority of adverse events are relatively simple—or can be treated as relatively simple without serious consequences—and that they therefore can be dealt with in ways that are familiar. But there is a growing number of cases where this approach will not work. For these, it is necessary to adopt a Safety-II view—which essentially means adopting a resilient health care view (Hollnagel, Braithwaite & Wears, 2013).

Safety-II is first and foremost a different way of looking at safety, hence also a different way of applying many of the familiar methods and techniques. In addition to that

it will also require methods on its own, to look at things that go right, to analyse how things work, and to manage performance variability rather than just constraining it (Wears, Hollnagel & Braithwaite, 2015).

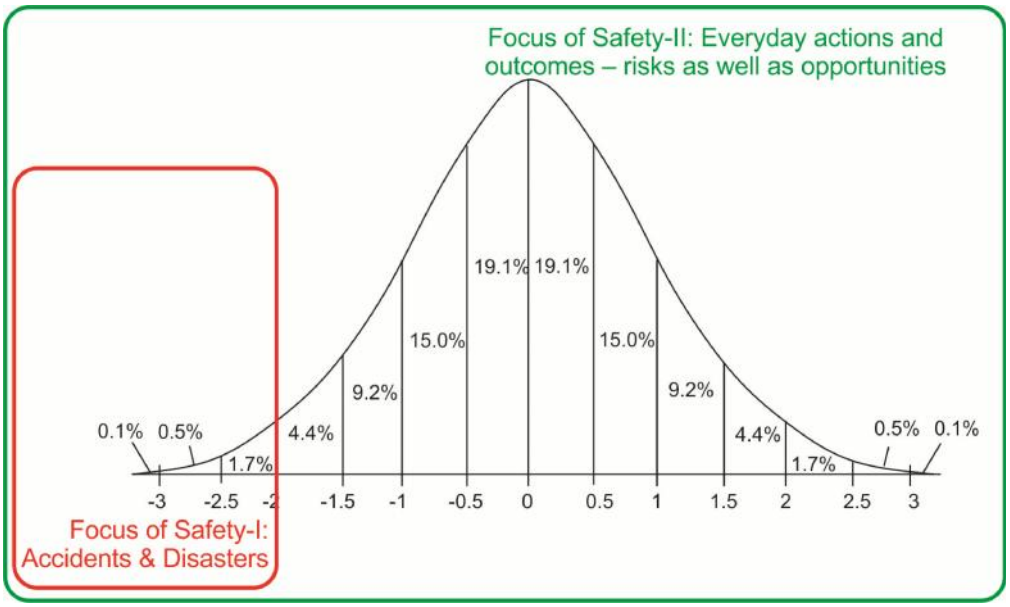


Figure 11: Focus of Safety-I and Safety-II

Epilogue

Introducing a different understanding of today's world and of the systems we work in and depend upon may require something akin to a paradigm shift. The safety community has developed a consensus on how things work and how safety can be ensured, but the increase of knowledge has levelled off, and the wicked problem of adverse events has continued. We must face the fact that the world cannot be explained by cause-effect models. Incidents and accidents do not only happen in a linear manner, but include emergent phenomena stemming from the complexity of the overall health system. Asking for "why and because" does not suffice to explain the system in use and does not lead to an improvement in safety.

As a consequence of the paradigm change, safety experts and safety managers need to leave their 'comfort zone' and explore new opportunities. In that new world, managers as well as practitioners are looking for models and methods to be used. Some methods already are available and have been applied in different settings. For example, the Functional Resonance Analysis Method (FRAM; Hollnagel, 2012) seeks to identify and describe

essential system functions, characterise the potential variability of the functions, define the functional resonance based on dependencies and couplings among functions and identify ways to monitor the development of resonance either to dampen variability that may lead to unwanted outcomes or to amplify variability that may lead to wanted outcomes (http://www.functionalresonance.com/).

The new paradigm also means that the priorities of safety management must change. Instead of conducting investigations after the event or striving to reduce adverse outcomes, safety management should allocate some resources to look at the events that go right and try to learn from them. Instead of learning from events based on their severity, people should try to learn from events based on their frequency. And instead of analysing single severe events in depth, people should explore the regularity of the many frequent events in breadth, to understand the patterns in system performance. A good way to start would be to reduce the dependency on 'human error' as a near-universal cause of incidents and instead understand the necessity of performance variability.

References

Altman, L. (1995). 'Big doses of chemotherapy drug killed patient, hurt 2d'. The New York Times, 24 March.

Bainbridge, L. (1983). Ironies of automation. Automatica, 19(6), 775-779.

Clary, M. (1995). 'String of Errors Put Florida Hospital on the Critical List'. Los Angeles Times, 14 April. Accessed 11 December, 2014: http://articles.latimes.com/1995-04-14/news/mn-54645_1_american-hospital.

EUROCONTROL (2009). A white paper on resilience engineering for ATM. Brussels: EUROCONTROL.

Finkel, M. (2011). On Flexibility: Recovery from Technological and Doctrinal Surprise on the Battlefield. Stanford, CA: Stanford University Press.

Heinrich, H. W. (1931). Industrial accident prevention: A scientific approach. New York: McGraw-Hill.

Hollnagel, E. (2009). The ETTO principle: Efficiency-thoroughness trade-off. Why things that go right sometimes go wrong. Farnham, UK: Ashgate.

Hollnagel, E. (2012). FRAM: The Functional Resonance Analysis Method. Farnham, UK: Ashgate.

Hollnagel, E., Braithwaite, J. & Wears, R. L. (2013) Resilient health care. Farnham, UK: Ashgate.

Hollnagel, E., Woods, D. D. & Leveson, N. G. (2006). Resilience engineering: Concepts and precepts. Aldershot, UK: Ashgate.

Leveson, N. G. (2004). A new accident model for engineering safer systems. Safety Science, 42(4), 237-270.

Rasmussen, J. & Svedung, I. (2000). Proactive risk management in a dynamic society. Karlstad, Sweden: Swedish Rescue Services Agency.

Reason, J., Shotton, R., Wagenaar, W. A., Hudson, P. T. W. & Groeneweg, J. (1989). Tripod: A principled basis for safer operations. The Hague: Shell Internationale Petroleum Maatschappij.

Reason, J. T. (1997). Managing the risks of organizational accidents. Aldershot, UK: Ashgate Publishing Limited.

Reason, J., Hollnagel, E., & Paries, J. (2006). *Revisiting the Swiss Cheese Model of Accidents*. EUROCONTROL. Brétigny-sur-Orge, FR. Retrieved 6 October 2008, from

http://www.eurocontrol.int/eec/gallery/content/public/documents/EEC_notes/2006/EEC_note_2006_13.pdf.

Shorrock, S. and Licu, T. (2013). Target culture: lessons in unintended consequences. HindSight 17. Brussels: EUROCONTROL.

Taylor, F. W. (1911). The principles of scientific management. New York: Harper.

Wears, R. L. and S. J. Perry (2006). Free fall - a case study of resilience, its degradation, and recovery, in an emergency department. 2nd International Symposium on Resilience Engineering, Juan-les-Pins, France, Mines Paris Les Presses.

Wears, R. L., Hollnagel, E. & Braithwaite, J. (2015) The resilience of everyday clinical work. Farnham, UK: Ashgate.

WHO (2014). NCD* death rate, age standardized (per 100 0000 population, 200-2012. Accessed 11 December, 2014. http://gamapserver.who.int/gho/interactive_charts/ncd/mortality/total/atlas.html.

Glossary

Adverse events: The undesirable effects of harm resulting from a health care intervention, treatment or prescription is often called an adverse event. Related terms include incidents, errors, undesirable side effects, or iatrogenic harm. Under Safety-I, a proportion of adverse events is deemed preventable.

(Approximate) Adjustments: When working conditions are underspecified or when time or resources are limited, it is necessary to adjust performance to match the conditions. This is a main reason for performance variability. But the very conditions that make performance adjustments necessary also mean that the adjustments will be approximate rather than perfect. The approximations are, however, under most conditions good enough to ensure the intended performance.

Bimodality: Technological components and systems function in a bimodal manner. Strictly speaking this means that for every element of a system, the element being anything from a component to the system itself, the element will either function or it will not. In the latter case the element is said to have failed. The bimodal principle does, however, not apply to humans and organisations. Humans and organisations are instead multi-modal, in the sense that their performance is variable—sometimes better and sometimes worse but never failing completely. A human 'component' cannot stop functioning and be replaced in the same way as a technological component can.

Causality credo: There is a widely prevailing assumption that adverse events occur because something has gone wrong. Once the cause is found, the situation can be resolved. All accidents and errors, under this logic, can be prevented: the causality credo. However, according to resilient health care principles, successes and failures spring from the same normal activities.

Decomposition: When a problem, process or system can be broken down into parts for the purpose of conceptualising or understanding it, it is decomposable.

Domino model: Heinrich's original domino theory published early in the history of safety, in 1931, depicts the cumulative chain or sequence of events which are triggered by an initial stimulus, metaphorically like a line of dominoes falling over.

Efficiency-thoroughness trade-off: The efficiency-thoroughness trade-off (ETTO) describes the fact that people (and organisations) as part of their activities practically always must make a trade-off between the resources (time and effort) they spend on

preparing an activity and the resources (time, effort and materials) they spend on doing it.

Emergence: In a growing number of cases it is difficult or impossible to explain what happens as a result of known processes or developments. The outcomes are said to be emergent rather than resultant. Emergent outcomes are not additive, not predictable from knowledge of their components, and not decomposable into those components.

Intractable systems: Systems are called intractable if it is difficult or impossible to follow and understand how they function. This typically means that the performance is irregular, that descriptions are complicated in terms of parts and relations, and that it is difficult to understand the details of how the system works. Intractable systems are also underspecified, meaning that it is impossible to provide a complete description of how work should be carried out for a sufficiently large set of situations.

Performance variability: The contemporary approach to safety (Safety-II), is based on the principle of equivalence of 'successes' and 'failures' and the principle of approximate adjustments. Performance is therefore in practice always variable. The performance variability may propagate from one function to others, and thereby lead to non-linear or emergent effects.

Resilience: The performance of a system is said to be resilient if it can adjust its functioning prior to, during, or following events (changes, disturbances, and opportunities), and thereby sustain required operations under both expected and unexpected conditions.

Resilience engineering: The scientific discipline that focuses on developing the principles and practices that are necessary to enable systems to perform resiliently.

Root Cause Analysis (RCA): In Safety-I thinking, safety breaches, errors and adverse events manifest regularly. Linear models suggest that getting to the ultimate source of a problem, and fixing it, can prevent recurrence. Hence: root cause analysis. Critics suggest that this is reactive at best, and few ultimate sources of problems are amenable to simplistic fixes.

Safety-I: Safety is described here as the condition where the number of adverse outcomes (e.g., accidents, incidents and near misses) is as low as possible. Safety-I is achieved by trying to make sure that things do not go wrong, either by eliminating the causes of malfunctions and hazards, or by containing their effects.

Safety-II: Safety is described here as a condition where the number of acceptable outcomes

is as high as possible. It is the ability to succeed under varying conditions. Safety-II is achieved by trying to make sure that things go right, rather than by preventing them from going wrong.

Scientific management theory: Frederick W Taylor was an early management theorist who studied work and its constituent tasks, aiming to simplify it and optimise efficiency. Also known as Taylorism, scientific management conducts time and motion studies to help determine the most efficient way of performing tasks. Management should plan and train, and workers should take instructions, work hard, and perform efficiently. This denies worker autonomy and is ill-suited to modern professional workplaces including health care.

Socio-technical systems: Originally coined by Trist, Bamforth and Emery from their work in English coal mines, sociotechnical systems theory focuses on the relationships between workers and technology. More recently, the emphasis has been to look at society's and organisations' complex infrastructures and human behaviour. According to this perspective, society itself, and its organisations and institutions, are complex socio-technical systems.

Swiss Cheese Model: All socio-technical systems include barriers and defences to prevent that accidents happen and that harm results. The Swiss cheese model of accident causation suggests that multiple barriers are similar to layers of Swiss cheese, stacked one after the other. While the layers mitigate the risk of an accident taking place, the barriers may sometimes have 'holes' in them and therefore not work as intended. When the holes line up, all defences are defeated, making accidents much more likely.

System flexibility: A flexible system is one that can adapt in response to internal or external changes. Responsiveness and adaptability are key to sustaining performance over time.

Tractable systems: Systems are called tractable if it is possible to follow and understand how they function. This typically means that the performance is highly regular, that descriptions are relatively simple in terms of parts and relations, and that it is easy to understand the details of how the system works.

Work-As-Done: What actually happens. Those providing care or services—doctors, nurses and allied health professionals—do clinical work on the front line. They appreciate the fine details of how clinical work is accomplished, but they do not always have responsibility for the standards, policies and procedures that govern their work.

Work-As-Imagined: What designers, managers, regulators, and authorities believe happens

or should happen. Those remote from the clinical front line receive second-or third-order accounts of how work is done, and there is always a lag between everyday clinical work and the information managers and policymakers receive about it. The basis for developing standards, policies and procedures will therefore always be incomplete and often incorrect.

About the authors

Professor Erik Hollnagel

Dr Erik Hollnagel, M.SC., PhD, is Professor at the Institute of Regional Health Research, University of Southern Denmark (DK), Chief Consultant at the Centre for Quality, Region of Southern Denmark, Visiting Professor at the Centre for Healthcare Resilience and Implementation Science, Macquarie University (Australia), and Professor Emeritus at the Department of Computer Science, University of Linköping (S). He has through his career worked at universities, research centres, and industries in several countries and with problems from many domains including nuclear power generation, aerospace and aviation, software engineering, land-based traffic, and healthcare.

His professional interests include industrial safety, resilience engineering, patient safety, accident investigation, and modelling large-scale socio-technical systems. He has published widely and is the author or editor of 22 books, including five books on resilience engineering, as well as a large number of papers and book chapters. The latest titles, from Ashgate, are "Safety-I and Safety-II: The past and future of safety management", "Resilient Health Care", "FRAM – the Functional Resonance Analysis Method", and "Resilience engineering in practice: A guidebook". Professor Hollnagel also coordinates the Resilient Health Care net (www.resilienthealthcare.net) and the FRAMily (www.functionalresonance.com).

Professor Robert Wears

Dr Bob Wears, M.D., PhD, M.S., is an emergency physician, Professor of Emergency Medicine at the University of Florida, and Visiting Professor in the Clinical Safety Research Unit at Imperial College London. His further training includes a Master's degree in computer science, a 1 year research sabbatical focused on psychology and human factors in safety at Imperial College, followed by a PhD in industrial safety from Mines ParisTech (Ecole

Nationale Supérieure des Mines de Paris). He serves on the board of directors of the Emergency Medicine Patient Safety Foundation, and multiple editorial boards, including *Annals of Emergency Medicine, Human Factors and Ergonomics,* the *Journal of Patient Safety,* and the *International Journal of Risk and Safety in Medicine.*

Professor Wears has co-edited two books, *Patient Safety in Emergency Medicine*, and *Resilient Health Care,* and he is working on two more. His research interests include technical work studies, resilience engineering, and patient safety as a social movement. His research papers and commentaries have appeared in *JAMA*, *Annals of Emergency Medicine*, *Safety Science*, *BMJ Quality & Safety*, *Cognition Technology & Work, Applied Ergonomics,* and *Reliability Engineering & Safety Science.*

Professor Jeffrey Braithwaite

Dr Jeffrey Braithwaite, BA, MIR (Hons), MBA, DipLR, PhD, FAIM, FCHSM, **FFPHRCP (UK)** is Foundation Director, Australian Institute of Health Innovation, Director, Centre for Healthcare Resilience and Implementation Science and Professor of Health Systems Research, Faculty of Medicine and Health Sciences, Macquarie University, Australia. His research examines the changing nature of health systems, particularly patient safety, standards and accreditation, leadership and management, the structure and culture of organisations and their network characteristics, attracting funding of more than AUD$60 million. He holds visiting professorial appointments at the University of Birmingham, UK; Newcastle University, UK; The University of Southern Denmark; the University of New South Wales, Australia; and the Canon Institute of Global Studies, Tokyo, Japan.

Professor Braithwaite has published extensively (more than 300 total publications) and has presented at international and national conferences on more than 600 occasions, including over 60 keynote addresses. His research appears in journals such as the *British Medical Journal*, *The Lancet, Social Science & Medicine, BMJ Quality and Safety, International Journal of Quality in Health Care, Journal of Managerial Psychology, Journal of the American Medical Informatics Association*, and many other prestigious journals. Professor Braithwaite has received numerous national and international awards for his teaching and research.

Acknowledgements

We would like to pay tribute to our treasured colleagues in the Resilient Health Care Network who are a constant source of ideas and inspiration. Sincere thanks go to Dr Brette Blakely, Ms Jackie Mullins, and Ms Sue Christian-Hayes, who researched documents, sourced photos and proofed and formatted the White Paper. Errors remaining are ours alone. We also gratefully acknowledge the initiative of Tony Licu to commission the Eurocontrol White Paper on Safety-I and Safety-II (Hollnagel, E., Leonhardt, J., Licu, T. & Shorrock, S. (2013). *From Safety-I to Safety-II: A White Paper*. Bruxelles, BE: Eurocontrol.) has been the inspiration for this document.