Institute *for* Healthcare Improvement

Moving from Safety I to Safety II in the Measurement of Patient Safety

*Moving Measurement into Action: Designing Global Principles for Measuring Patient Safety | Pre-Seminar Briefing Document on Safety I and Safety II*

**Prepared by:**
Erik Hollnagel, Ph.D.
Senior professor, Patient Safety, Jönköping University, School of Health and Welfare, Jönköping, Sweden
Visiting Professorial Fellow, Macquarie University, Sydney, Australia

## A Changing Landscape

Safety is traditionally defined as a condition where the number of adverse outcomes *is as low as possible* (Safety-I). The purpose of safety management is therefore to prevent the occurrence of such outcomes as far as possible. This leads to a reactive approach based on responding to what goes wrong or what is identified as a risk, where safety is measured by counting the number of adverse or unwanted outcomes.

The alternative is to define safety as a condition where the number of intended and acceptable outcomes is *as high as possible* (Safety-II). From this perspective the purpose of safety management is to ensure that everyday work achieves its objectives so that as much as possible goes well. Safety is managed by focusing on what it achieves and is likewise measured in terms of acceptable outcomes. In order to do this, safety management cannot just be reactive to prevent adverse outcomes, but must also be proactive to facilitate the work that is the basis for everyday acceptable performance (Braithwaite, Wears and Hollnagel, 2015).

A main challenge to safety management is that technical and socio-technical systems alike have become gradually harder to understand and control. This has required a corresponding development in methods, although usually with a significant delay. Briefly told, the understanding has gone from single factor models over simple and composite linear models to complicated multi-linear models. Similarly, new types of causes have been introduced whenever the 'usual suspects' were unable to explain why a major accident or disaster happened. The genealogy of causes goes from 'acts of god', to technical malfunctions and failures, to 'human errors', to organisational failures, to safety culture, and to complex systems – which at the moment represents the pinnacle of safety thinking. In each case the new causes have unfortunately been introduced without challenging the unspoken assumption that outcomes can be understood as effects that follow from prior causes, a.k.a. the *causality credo*. While it may be comforting to think within the same paradigm, it is not viable in the long run (Hollnagel, 2014).

## Safety as Absence ("without") or Safety as Presence ("with")

Reason (2000) astutely pointed out that "safety is defined and measured more by its absence than by its presence". One unfortunate consequence of this is that safety measurements are

inversely proportional to the level of safety and that "perfect" safety therefore means there is nothing to measure. Another is that we try to learn from events where safety was absent rather than present, which makes learning a tad difficult. But adopting a Safety-II perspective avoids both problems. Measuring safety in terms of the acceptable outcomes may at first appear difficult because it clashes with traditional safety management. Yet it is quite natural to do so for other aspects such as productivity, quality, customer satisfaction, etc. And it also means that we do not have to wait for an accident to happen in order to learn anything.

## Safety Measurements and Safety Management

In order to manage patient safety, and more generally in order to manage any process or activity, three things are needed: knowing where we want to be – the goal or target; knowing where we are right now – the current position or status, hence the difference between the current position and the goal; and finally knowing how to change the current position to get closer to the goal. All three require information or meaningful measurements of some kind.

All three also require that measurements are compared to something. Indeed, a measurement is not meaningful unless it can put in relation, implicitly or explicitly, to something else. It can be an *absolute* comparison to some kind of norm or standard, such as "zero accidents". It can be a *relative* comparison some similar system or organisation, such as a ranking of hospitals using HSMR. And it can be an *autologous* comparison of a system to itself over time, in order to show how it has changed.

A comparison of measurements only makes sense it if the two – or more - "somethings" that are compared are similar, specifically that they work or function in the same manner. This is rarely, if ever, the case for absolute and relative comparisons. But it is the case for autologous comparisons, at least if the system has not been subject to major changes between measurements.[1]

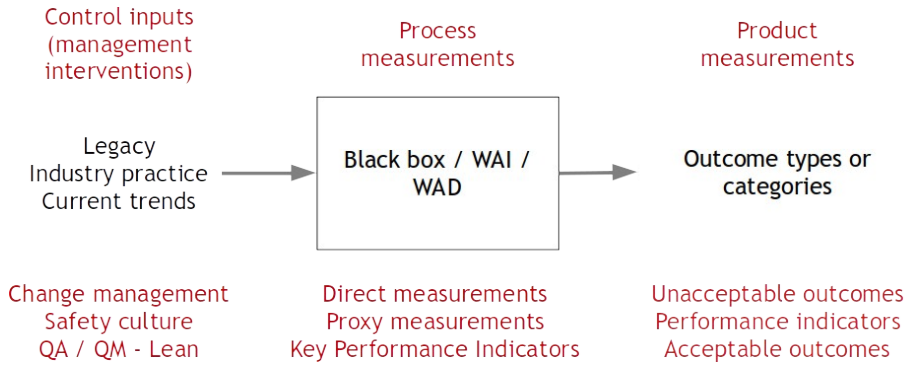
## Process and Product Measures

It is impossible to manage an organisation, hence to manage safety, unless we have a reasonably accurate understanding of how it works and unless we can measure how well it works. In relation to measurements a distinction is usually made between *process* and *product* measures. The former represent how the organisation works and the current status, while the latter represent the results. While measures of recognisable outcomes, acceptable and unacceptable alike, can be convenient they are nevertheless not the best basis for safety management. A further problem is that they always are lagging, sometimes with a considerable delay.

Safety management assumes that the outcomes in some way are the consequence of how an organisation functions and also that it is possible to control that. The assumptions can be relatively simple, for instance the belief in linear causality that is part of all Safety-I models and of most organisational models as well. The assumptions can also be more

1 Comparing the current state to the goal is an autologous rather than an absolute comparison because it refers to the intended future state of the system itself.

elaborate, such as in high reliability organisations or even non-linear as in complex adaptive systems or resilience engineering.




In order to manage a process, to manage patient safety, we must be able to measure the the process in order to know the current state or position. This can be by direct or by proxy measures which both depend on a good understanding of how it works. But how much do we actually know about what happens in a health system, or in any large-scale socio-technical system? In the worst case the system is a black box, which means that we know little if anything. In the slightly better case we assume that what goes on corresponds to Work-as-Imagined, although that rarely is so in practice. In the best case our understanding corresponds to Work-as-Done (Braithwaite, Wears & Hollnagel, 2017).

Considering the outputs, the choice of what to measure also requires a good understanding of the process. The outputs are assumed to represent the consequences of how the chosen control inputs have changed the process – assuming that the system otherwise is reasonably stable. A crucial but often overlook side of that is *when* – in the sense of how soon or how late – the expected outputs should be recognisable.

Finally, the control inputs also depend on accurate knowledge of the process, both how it takes place and what the current status is (process measurements). Inputs are chosen to manage the process, which essentially means to change the current position to get closer to the goal. They should therefore be based on an understanding of the process actually works rather than rely on institutionalised practices or "industry standards".

## Measurement Challenges

Measurements that are based on outcomes or products are usually simple to make but their meaningfulness depends on the underlying model or assumptions about how an organisation functions. Since the relationship between the organisation (the process) and the outcomes (the products) either is underspecified or expressed in very general terms, safety culture being an example of that, relying on outcome measures alone is not advisable.

A better solution is to base safety management on process measures, where a distinction can be made between direct measures and proxy measures. In the case of technical systems it is usually simple both to define relevant process measures (because the processes are designed and well known) and to make the measurements. In the case of socio-technical systems it is far more difficult, verging on the impossible. This is not for a lack of trying,

such as the various forms of SPC, six sigma, Balanced Score Cards, etc., illustrate – even though some of them strictly speaking they are outcome measures in disguise.

## The Practical Consequences

A short-hand characterisation of the differences between Safety-I and Safety-II is to note that Safety-I leads to *protective* safety while Safety-II leads to *productive* safety. Safety-I and Safety-II both want as little as possible to go wrong, but they do it in different ways. Protective safety means that the number of adverse outcomes, the number of things that go wrong, is reduced by *preventing* things from going wrong - by elimination, barriers, and protection. Productive safety means that the number of successful outcomes, the number of things that go well, is increased as much as possible. This can be done by *understanding* how work goes well and find ways to *support* and *facilitate* that. And since something cannot go well and go wrong at the same time, an increase in the number of things that go well will obviously lead to a decrease in the number of things that go wrong.

## References

Braithwaite, J. Wears, R. L. and Hollnagel, E. (2015). Resilient health care: turning patient safety on its head. *International Journal for Quality in Health Care*, 1–3. doi: 10.1093/intqhc/mzv063

Braithwaite, J., Wears, R. L. & Hollnagel, E. (Eds.) (2017). *Resilient Health Care, Volume 3: Reconciling Work-as-Imagined and Work-as-Done*. CRC Press.

Hollnagel, E. (2014). *Safety-I and Safety-II: The past and future of safety management*. Farnham, UK: Ashgate.

Reason, J. T. (2000). Safety paradoxes and safety culture. *Injury Control & Safety Promotion*, *7*(1), 3-14.