Good Science, an antidote to Ben Goldacre's "bad science".

Authors:

Julian Simcox – practicing Patient Leader in North Somerset e: **julian@cohear.com**

Terry Weight – retired improvement consultant and statistician e: **terry.weight1@btinternet.com**

Abstract

In his writings about healthcare Ben Goldacre's nemesis is "bad science" – but just what is "good science"? Healthcare is a revealing context when addressing this question because in the wake of the 2013 Francis Report into why in the Mid-Staffordshire hospital in the mid-noughties as many as 1200 patients unexpectedly died, Don Berwick was commissioned by the Prime Minister to assess how best to make the NHS more safe .. in his subsequent report he is very clear about what this is going to take: a wholly new culture, and at its heart: good science. The delivery system that he recommends is a radical departure from the way things have previously been organised – including a comprehensive methodology aimed at generating evidence-based knowledge – in real time, or as close to real time as practically possible – and emanating from team-based learning via structured improvement methods, conducted not by statisticial experts but by healthcare professionals, managers, and staff – indeed even patients and their carers. He generically refers to all this as "Improvement Science"(7) which embodies the scientific method (P-D-S-A) in a way that is simpler and often more appropriate than the Test-Learn-Adapt model recently advocated jointly by Ben Goldacreand the Cabinet Office's behaviour insights team (27). Vitally, improvement science is not some esoteric whim, it has a time-honoured provenance – it comes with an established methodology, one that has been under refinement for over 80 years, in a variety of sectors, *and* across the globe. In short it is an approach that works – a way of making good science accessible to those who most need it: those directly responsible for designing and operating systems that deliver quality care.

Ben Goldacre is additionally advocating a scientific, evidence-based, approach to healthcare – and in government policy-making generally – and that this ought to replace the policy-based evidence that we more commonly see in use. We whole heartedly support him in this albeit, as we will argue, his approach fixates upon just a single method of research – perhaps teeing-up unsuspecting healthcare professionals for failure? If in this paper we can establish what good science is, we hope can yield a more unified approach to evidence-based healthcare – one like Don Berwick's.

If everyone at least in healthcare – especially activists like Ben Goldacre – could get full square behind Don Berwick's proposals, an historic opportunity momentarily exists to create a tsunami of clinicians, managers and citizens/ patients – as well as civil servants, politicians and journalists – who together can make evidence-based scientific thinking an all pervasive cultural norm? (410 words)

Keywords

Good science; Bad science; Evidence-based policy; Policy-based evidence; Improvement science; Knowledge Management; Randomised Controlled Trial; Systems Thinking; Safe healthcare systems; Statistical Process Control; Continual Improvement; World Class; Measurement of healthcare outcomes; System Behaviour Charting; Design Thinking; Idealized Design; Ben Goldacre; Don Berwick; W. Edwards Deming; Russell Ackoff;

Contents

1. Introduction

Ben Goldacre – across the whole of the public sector, but particularly in healthcare (including its many private sector suppliers) – makes "bad science" his arch-enemy. We two, now retired activists in our respective communities, are strong followers of his campaigns and great supporters of his aims. He is to be especially commended for popularising the idea that medicine is only as efficacious as the cummulative knowledge upon which it is based, and that everyone who works in healthcare has a responsibility to contribute to the furtherance of that knowledge i.e. to do "good science". In collaboration with the Cabinet Office's behaviour insights team, he has recently published a polemic (27) advocating evidence-based government policy. For us this too is commendable, yet there is a potentially grave error of ommission in their paper which seems to fixate upon just a single method of research, and risks setting-up the unsuspecting healthcare professional for failure and disappointment – for as Abraham Maslow once famously said *".. it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail"* (43).

Goldacre and his colleagues purvey Randomised Controlled Trials (RCTs) as if in medicine it's the only way to conduct medical research, but in doing so they ignore "Improvement Science" (7) which comes from a separate school of scientific thought and in our long experience has much more to offer the practicing medic – especially now that the government has accepted the Berwick Report (8) on patient safety, a document that has improvement science at its core, and that possesses the potential – across the whole world of medicine – to unify and elevate research practice into something that all healthcare professionals, all staff, all patients/carers and public, can engage in. To be fair, Goldacre and co may be seeing RCTs as *simply* a useful starting point, but as we will argue, from a pragmatic perspective, a Randmised Controlled Trial is rarely the best place to start and making it so might actually hold back moves to make policymaking more evidence-based.

Goldacre makes an archenemy of "Bad Science" – the title of his weekly Guardian column. Likewise our aim in this paper is to point healthcare professionals toward a practical way of doing "good science" – via methods that are not dependent upon a narrow approach, nor a single tool. He advocates in his book – also called Bad Science (26) – that we each learn how to conduct our own experiments. With this same aim Berwick recommends a tried and trusted approach that requires all NHS staff to "***Learn, master and apply the modern methods of quality control, quality improvement and quality planning***" – at the heart of which is team-based learning conducted not by statisticians nor academics, but by healthcare professionals, managers, and staff. Like Goldacre we too wish to spread a philosophy of science but we see little need for the additional Test, Learn and Adapt (TLA) model he offers because the NHS already has such a model – one which is often more simple to follow – it is called the "Improvement Model"(72) – and via its P-D-S-A mnemonic (Plan-Do-Study-Act) embodies the scientic method. Moreover there is a pre-existing wealth of experience on how best to embed this thinking within organisations – from top-to-bottom *and importantly* from bottom-to-top – experience that was founded in industrial settings and then spread to the service sector – experience that has been accumulating for fully nine decades (59).

Healthcare, as well as the rest of the public sector, deserves a unified approach to the doing of good science. Ben Goldacre's goal of making this possible is both timely and courageous – not least because his choice of collaborator is a group of civil servants who are ostenisbly disconnected from the Dept. of Health (DoH) – and we would love to enlist his energy to the cause of evidence-based scientific thinking – via a small suite of methodologies and tools that are as accessible to healthcare professionals as they are to politicicans and policymakers in general.

Our paper starts by historically positioning the whole idea of the wider application of science, arguing that its time has been coming for five centuries yet is only now fully arriving. We suggest that in a world where many at school have been turned-off science, if rendered accessible and practical, the propsensity to be scientific at work makes a vast difference to outcomes – especially if those who take pride in saying they avoided science at achool, or freely admit they don't do numbers, can get switched on to it.

In making science accessible to all in healthcare, and in work generally, we make a vital distinction between three very different ways of applying science in the workplace – based upon three very different approaches to being evidence-based – labelled "**accountability, improvement, and research**" (65) – as a shorthand way of alluding to the primary purpose of each approach.

We also suggest that Ben Goldacre's nine step T-L-A model is less helpful than it might first appear – and that it can be made more effective by extending it – and at both ends. But that even then the model is less helpful than the "model for improvement" which is already in use in the NHS, and already widely used internationally.

The Model for Improvement has a pedigree originating with Walter Shewhart in the 1920s, then being famously applied by Deming and Juran in Japan in the 1950s. Deming in particular encapsualted the **scientific method** in P-D-C-A (three decades later he revised it to P-D-S-A) as a practical way to enable a learning/improvement method to evolve bottom-up in organisations. After the 1980s Don Berwick then, standing on their shoulders, developed its scientific application in the world of healthcare – initially in his native america. Berwick's approach is to encourage people to ask questions such as .. **what works?..** and **how would we know?** His method, is founded upon a culture of evidence-based learning, providing a local context for systemic improvement efforts. In this way a new organisational culture, one rooted in the science of improvement may, if properly nurtured, may then emerge.

Such a culture may initially jar with the everyday life of a conventional organisation, and the individuals within it. One of many reasons for this is that for hundreds of generations our species has evolved such that ***imagined*** **reality** has been lorded over ***objective*** **reality**. Only relatively recently in our evolution have we witnessed the advance of science levelling up this dichotomy, and we will argue that a method is now needed that enables these two realities to more easily coexist. We suggest that a method that enables data-rich evidence-based storytelling – by those who most know the context and intend growing their collective knowledge – will provide the basis for such a method.

We aim for this method to have appeal for the overly busy healthcare professional, for managerial leaders generally, and for civil servants – indeed for anyone intuitively who feels there must be a better way to combine goals that currently feel disconnected or even in conflict: empowerment and accountability; safety and productivity; assurance and improvement; compliance and change; extrinsic and intrinsic morivation; evidence and action; facts and ideas; logic and values; etc .. indeed for anyone who is searching for ways to unify their actions with the system-based implementation of those actions as interventions.

We will illustrate the method with **case studies**, whilst assessing where we think Ben Goldacre's recent paper contributes. We also propose an approach based upon 'good science' that if adopted would we believe turbo-charge the achievement of his aims – starting in healthcare yet also having a developmental influence right across the public sector. Enroute we will show that by requiring all three paradigms – accountability/ improvement/ research – to be in play, the 2nd of these: the improvement science paradigm, which in conventional organisations is commonly missing, would provide a useful bridge between the remaining two, often estranged bed-fellows.

Allowing all three approaches to play their part is therefore vital, as is having a method that appropriately combines all three – because "good science" requires this. We will also demonstrate how this will enable a focused range of tools to be available – more than just the RCT – and how doing this would make a telling difference to the ability to sustain intended outcomes.

In addition we will show how all this depends upon a consistent **theory of knowledge**. Observed data inevitably has to be filtered by human filters and schemas, and reality can never be wholly complete – and yet data, collected over time, may be used to provide signals that reveal how almost any system in which we are interested is in essence behaving. Generating knowledge about the system therefore requires knowledge about the particular context. Understanding of a system's behaviour – **past**, **present**, *and* **future** – arrives only via thoughtful consideration of data and information, over time, identifying for example whether stability is being gained or lost. A narrative is shaped by those working in and on a system as they collaboratively make meaning by assessing the possible causes of the behavioural patterns they are observing. This narative is all important : the telling of a knowledgefull story – one that becomes a potent interweaving of the reality that is being *imagined*, with the *objective* reality that is being presented as self-evident fact.

In short, the method that comes with improvement science has a historical pedigree that when well adhered to by those who are directly connected to a context can enable the richest of narratives to be created: a truly knowledgeful story. There are many examples of this ocurring in both the public and the private sectors, but because of a lack of strategic context this occurs usually only in isloated pockets. We believe that healthcare in particular is on the cusp of breakthrough, prompted in part by the systemic failures at Mid Staffordshire Hospital, Winterbourne View Care Home and Morecambe Bay Foundation Trust Hospital – which insiders widely acknowledge are the tip of an iceberg.

We will know that we will have succeeded with this paper not just if it is widely read, but if we enlist the likes of Ben Goldacre to the definition of science to which we allude.

We wholeheartedly agree with Ben that for the public sector – not just in healthcare – policy-making needs to become more evidence-based, and in showing how science can be rendered more practical and accessible we aim for **good science** to become the do-able default. In our final section we explore what else this might take?

2. Time for Good Science

The two of us, in our roles as citizen/patients notice that the healthcare "system" often succeeds, but sadly all too often does so despite itself – and largely via the efforts of heroic individuals rather than through a well-constructed system. Healthcare seems too rarely to be **viewed as a system** by those responsible for designing and operating it – many of them blind to its potential or to its failures and/or to the possibility of fully learning from them. In recent years, more and more data seems to have been collected in order to inform political discussion in Westminster, and/or resource allocation in Whitehall – all too often, as if one-club golfers, being achieved only bluntly and arbitrarily via top-down targets, or via performance rankings that merely name-and-shame – meanwhile at the coal face next to no meaningful data is collected. Locked into a carrot & stick ethos, good science seems to be rendered a cultural impossibility, a considerable irony given that the system is driven by highly educated professionals most of whom will have had years of science education.

During the eight decades of our combined working lives, covering both the private and public sectors, if we learned just one thing it was that when people are asked to deliver something of which the system they are leading is currently incapable, there are only three options

(i) Improve the system's capability;
(ii) Distort the system; or
(iii) Distort the data.

The first option requires some understanding of the existing system; some knowledge of how best to change its design; as well as a tried and trusted method for affecting the changes so that they may sustain over time; and a little science so that it can be demonstrated that any observed new outcomes are real. This first option is sadly therefore rarely considered expedient, most people, most of the time opting, with varying degrees of cynicism, for either system or data distortion.

In his recent book "Sapiens – a brief history of humankind"(30) Yuval Noah Harari describes how since a "cognitive revolution" around 70,000 years ago Homo Sapiens has been living in a dual reality: on the one hand, the objective reality of rivers, trees and lions; and on the other hand the imagined reality of gods, nations and corporations. This second reality, amazingly borne of an evolving ability to gossip, separated our species from the other apes and then over time came to dominate the first kind of reality – such that today the very survival of our objective environment, rivers, trees and lions depends on the grace of our values, our gods, our nations and our corporations. To Harari this is both good and bad news for Sapiens.. our ability to invent totemic entities such as national values, brands, human rights, limited liability, the promise to pay the bearer on a bank note, has in effect enabled ever larger numbers of strangers to effectively cooperate – indeed the benefits have been so great that even just in our short lifetimes the planet's population has trippled. The bad news is that the quality of life for all but a small minority is – at least according to Harari – lower than the one that pertained when we were all *merely* foragers.

Harari describes how large-scale human cooperation, and hence human advancement (at least as measured by popoluation size) has been enabled by our collective ability to myth-make, and he suggests that our futures will depend upon how we can take the myths we inherit and continue to remake them i.e. by telling new stories – perhaps this is why our politicians talk so much of the need to "control the narrative"? He also describes how in just the last 500 years, myths have become increasingly shaped by science via ***"a common core of research methods, which are all based on collecting empirical observations***" (30). Harari cites Francis Bacon who by founding the scientific method in his 17th century scientific manifesto "*The New Instrument*" challenged accumulated myth, dogma and superstition – in effect challenging our species to get more scientific. Seventy thousand years after the cognitive revolution, we maybe are finally reaching the point where objective reality reasserts itself – for Sapiens, **a historically important watershed moment?**

In the UK this year we have been experiencing the second General Election in succession dominated by the question: Which party can appear most connected to "reality" – at least the fiscal kind? Perhaps the time has also now arrived for the organs of government to shift their knowledge-creating practices away from the *imagined* reality of policy-based evidence towards the *objective* science of evidence-based policy.. but how straightforward an undertaking would this be?

The trouble is, each of us as individuals operate according to the particular paradigms, sets of scehma that we have either aquired or were scripted to possess, each paradigm by definition an incomplete model of how the world really is. For the advancement of knowledge it is of vital importance therefore that we each start from a position of skepticism and Harari explains how pragmatically knowledge progresses only through a process of **continually discovering our ignorance**. To advance our individual and collective knowledge we require tools – specifically statistical ones.

We two happen to have statistical backgrounds, each now working as volunteers in our local communities. Statisticians are in short supply, and so, like Erasmus's one-eyed man in the kingdom of the blind, we are frequently tempted to pretend to know all the answers! We constantly feel the need to remind ourselves that like science, statistics is itself a field with more than one approach, and a set of research methods that themselves need to continually evolve. As patient/citizens, whenever we hear healthcare commissioners or providers talk about a proposed new intervention, we like to ask them .. **What are the intended outcomes? What might be some of the unintended ones? And as you proceed, just in case you need to change course, how will you know – in real time – what is actually happening?**

The **scientific method** has unarguably already made a massive contribution to how knowledge of the physical world and its phenomena is acquired. To be *scientific* in healthcare a method of **inquiry** is now required one based upon gathering **observable** and **measurable evidence,** that is subject to some set of commonly understood principles of **reasoning** and experimentation – the testing of **hypotheses**. Such a method exists, that has been proven to work well in everyday operational systems. We want to be careful to not propose anything so highfaluting or so esoteric that few will be able to apply it, we nevertheless feel it important – via a coherent philosophical framing – to root our proposals in science and the nature of scientific knowledge. We therefore aim to provide this in sections 6 and 9.

3. Knowledge is powerful if contextual and practical

> **"For *knowledge*, too, is *itself* a *power*" ..** *Sacred Meditations (1597)*
>
> **Francis Bacon** (1561-1626)

In this year's third Reith lecture, US surgeon Atul Gawande(11) talked about the problem of hubris, and the unwillingness of society and medical institutions to recognise the limits of knowledge, as well as the reluctance by professionals to talk about what they can realistically achieve – which itself results in widespread suffering. In addressing for example the End-of-Life process he proposes that when physicians are talking with the patient, they should put to them a couple of vital questions: ***"What is your understanding of where you are with your condition?"*** and ***"What outcome would be acceptable to you in this situation?"***

It may seem remarkable that these kinds of questions would be viewed as so revelatory. In the medical profession the approach has often been merely one of .. just inform the patient of only what is medically pertinent.. job done!

Gawande refers to one RCT study, done at the Massachusetts General Hospital (MGH) by a team who studied patients with Stage IV lung cancer – an incurable lung cancer – the patients subsequently living on average just 11 months. Half of the patients were given the usual oncology care, and the other half got the same care yet also got to see a specialist physician in palliative care – who would discuss with them what their priorities and goals might be for their end of life. This second group wound up with very different outcomes: they were much less likely to go onto the next round of chemotherapy (reducing chemotherapy costs by a third); they stayed in hospital one third fewer days; they were much less likely to die in the hospital or in the intensive care unit; and they started hospice care earlier. They had less suffering at the end of life, and somewhat surprisingly they actually then also lived 25 per cent longer. As Gawande points out.. *"If this were a drug, it would be a multi-billion dollar drug and we wouldn't be asking: oh can we afford it? .. But you know in truth it isn't even a matter of affording. These are basic skills .."*

At the point of the MGH story described by Gawande – with new learning having been achieved largely via standard statistical thinking – it remains a long way from being an embedded improvement that in practice will assure better future outcomes. We do not know what happened next, but it is worth some reflection.. ***what would you do next***?.. and how would you measure the impact of any intervention you might make? The professionals involved at MGH would have to agree that the results are sufficiently interesting to justify change, and then together agree the details of how such patients would be handled in future before attempting to successfully implement this change. For the new model of care to become an implemented improvement in other hospitals in the US, or in the wider world, there is even more to be done. In section six, we explore how this might have been addressed, or might be addressed, as an Improvement Science based project.

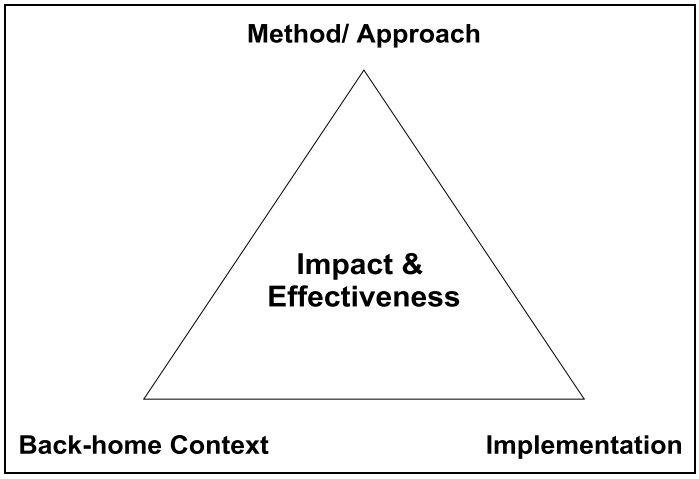
In his book: "Sapiens", Yuval Noah Harari (30) suggests that a distinguising feature of the modern world is that knowledge nowadays means power, and points to the study of statistics becoming institutionally core to undergraduate course design – replacing once pre-eminent subjects such as theology. For Harari however, the real test of knowledge is ***"not whether it reveals truth, but whether it empowers us."*** Whatever the approach or method that gets recommended, be it Ben Goldacre's, Don Berwick's or someone else's, we believe that the level of real empowerment created is a key test.

Berwick makes empowerment central to his conclusions on how to build a safe healthcare system, putting staff and their ability to improve the system (at least their bit of it) at the very heart of his vision of the NHS as a **Learning Organisation**. In his long career he has discovered that making knowledge practically obtainable over time, and in real time, is critical to transforming a system's culture into one that is truly powerful, one that is truly caring.

Having a practical approach to science in the workplace, and through it the accumulation of knowledge, therefore makes a vital difference – Berwick refers to this as "***the democratisation of science***" (6). Whilst Bacon saw Knowledge as power, for us if knowledge is to have meaning it must start with the individual, must be generated collaboratively with other individuals, and must be oriented towards a primary context e.g. the patient and the system that serves them – over time. This means the system as it *has been*, the system as it currently *is*, and the system as it *will be* – one that has been successfully designed to capably deliver better patient outcomes – outcomes that are purposefully defined in terms of the patient as an individual, the patient as a member of a community, and the patient as tax payer.

Efforts to improve safety and the quality of care have in recent years focused upon socio-technical interventions under badged initiatives such as Lean and Business Process Re-engineering, but as those who are actually managing/ leading change quickly find out, what works in one location often does not deliver in another. The nature of the intervention may itself vary more than is realised, and the implementation process too – but the largest source of variation is the environment, and as Deming said: "intervention + context = outcome" (31). Exaggerating to make the point, Paul Bate has even argued that "**context is everything**" (5).

In his (2000-02) study of some Orthopeadic Service Collaboratives, Bate examined three things:



Paul Bate (5)

And in explaining the variation in the outcome(s) of the **"back-home" context**, three aspects specifically stood out:

1. the **Leadership context** – style, method, level of support in programme board, faculty board, region, executive team level, local team leaders.
2. the **Political context** – level of empowerment, locus of decision making, configuration of top-down/bottom-up, and the mix of allies/ adversaries/ opponents/ bed-fellows/ fence-sitters.
3. the **Cultural context** – shared mindsets around quality, risk, participation, etc.

Of particular importance is the way that individuals in the midst of change selectively attend to, interpret, and attach significance or relevance to what they think they're witnessing – and how that feeds the narrative, their personal behaviour and their interractions? Whilst shaping the narrative and utilising storytelling as a tool has certainly become a powerful ally when manging/ leading change, and whilst stories are nowadays central to the change process, suspicion persists that the narrative is often fancilful and at best the source of only partial truth, with only a tenuous connection to what's scientifically real.

"There is surely a role for storytelling to generate narratives that practitioners engage with. However there is also a danger in this approach, as the use of narrative or metaphor risks being dismissed as not properly grounded in science." Paul Bate (5)

Surely there must be ways in which storytelling can be rendered more scienticially robust. But how?

According to the widely used Myers Briggs personality test (MBTI) half the people in the world prefer to value the subjective and what is personal to the individual, rather than what is objectively out there. This half perceives the other half as impersonal and unfeeling, even cold-hearted. Indeed, along with the Sensing/ Intuiting continuum (more of which later) the dichotomy between Thinking (T) and Feeling (F) is central to to MBTI typology – and one reason that it has become so popular is that the tension between objectivity and subjectivity is a perennial one.

Those possessing a strong "T" preference tend to assess what they're seeing according to a set of "scientifically proven" principles, or a given law or policy – typically preferring to analyse things and events according to clear criterion or publically declared standards. T types most respect what to them *is* objectively real.

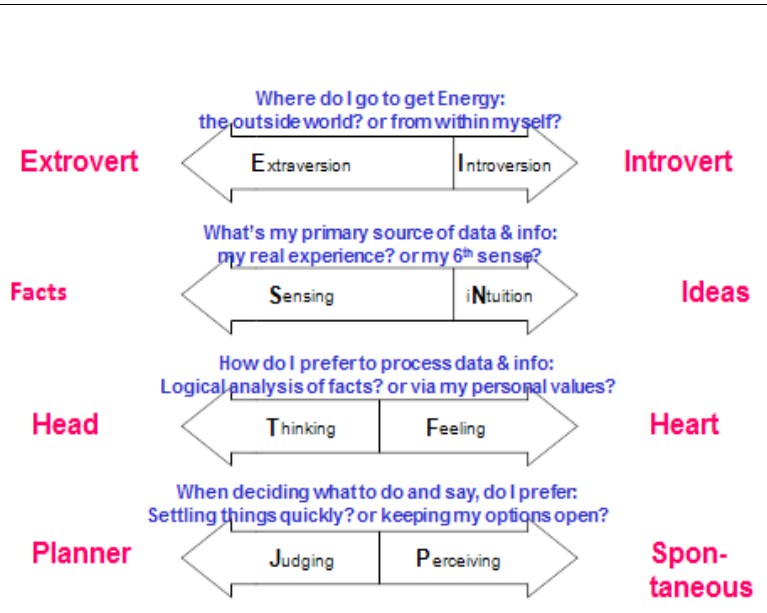
Conversely, people with an F preference *embrace* subjectivity, preferring to judge people, things and events by their personal values, importantly referencing these values as they attach meaning to the data they are observing. What feels right to an F-type must therefore be so – fairness typically having more to do with humanity than metred justice. Rules and laws are there to be bent if this is what it takes to maintain personal integrity.

The point we wish to make is that neither preference is more right than the other, they just *are* – and in the quest for knowledge both are to be prized. Knowing, for example, the price of everything and the value of nothing, is just as debilitating as knowing the value of everything but the price of nothing, and depending on one's preference, the context being observed is bound to appear and be interpreted very differently. One way of dealing with this is to assemble a strong team – one in which both F and T types are amply represented – albeit a robust method of dialogue would surely be needed if the team were to be able to function optimally.

***SIDE BAR*: The Myers Briggs model in outline**

*Here is a summary of how the MBTI model works. There are 4 continuums and we are here focusing on the middle two: **S** ↔ **N** and **T** ↔ **F**. The two central questions are: First, what is my primary source of data/information, is it my real experience or my 6th sense? and Second, how in the main do I process that data/ information, is it via a logical analysis of the facts or is it by referencing my personal values?*

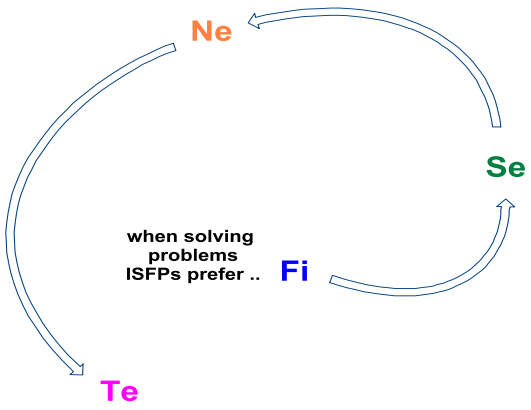
*According to MBTI type theory, the best decisions are those using both ways of **perceiving** (S and N) so that as much pertinent data and information as possible may be gathered, and so that then both ways of **judging** (T and F) can be applied to ensure that all the important factors can be fully weighed .. but few individuals are sufficiently developed to have the required **bandwidth.** Because there are 4 continuums, there are 16 possible types, and each has its own process for solving problems – an **ISFP** for example is Introverted Feeling supported by Extraverted Sensing, and whilst iNtuition and Thinking may come into play, this is usually as an afterthought.*



*Problems would usually be addressed as in this flow diagram below .. starting with introverted Feeling (**Fi**).*

*To the external observer this first function would not be visible – many are therefore mystified by the ISFP's behavior. What the ISFP reports they see via their extraverted sensing function (**Se**) has already been predetermined by what their value system dictates is visible to them.*

To further illustrate the S ↔ N dichotomy, we refer to this particular Myers Briggs type again on page (10).



(10)

The MBTI typology is of course merely a model, and no model perfectly fits reality. Our point though is that as the most widely used way of diagnosing personality type, the model elegantly illustrates how the world becomes too easily polarised between those preferring to deal in the observably tangible – concrete factors, variables, events, objects, causes, effects, the substantively real – and those who are more accepting of subjectivity, such as constructionists who believe that human beings continually generate knowledge and meaning from the interplay of their experiences with their ideas.

Subjectivists, half the world, prefer to prioritise the process of making-sense of what people think they observe, rather than of trying in any absolute sense to pin it down. If this sounds worrisome, the other half are cause for just as much concern in that they are prone to believe what they have pinned-down is all there is. In short, though science may at first blush depend on those who prefer to do the logical analysis, it is the subjectivists who by allowing for all shades of opinion often make knowledge more practical.

To be valid therefore, the study of context has to allow for the likely blindnesses of each person. Even a well balanced team will require a good team process as it tries to uncover a more complete picture about what is really going on. Attempting to understand events that have historically occurred, or are occurring in this moment, or are most likely to occur in the future, will never be straight forwards without the aid of a tried and trusted methodology – otherwise making proper sense of the inherent variation will *within any particular context* be nigh impossible – let alone accommodating the likely difficulties in understanding the inherent variation *between* contexts.

In the interests of enabling outcomes that are truly evidence-based, we therefore need a practical way of embracing a combination of both **heart and head**, so that inferences about our local context can rest openly upon both belief systems and objective standards (54). We believe that storytelling is a vital tool to have in our kitbag, but our experience also tells us that on its own storytelling is dangerous unless informed by the kinds of facts that an MBTI "T" type might through analysis alone deduce.

In healthcare the context is "the healthcare system" but tackling the NHS as a *whole* system is rarely the best starting place. The vast majority of NHS employees think first about their own very local/ very personal bit of the system – and for us that is where the power and the potential for knowledge accumulation lays. The created cultural context needs to be an empowering one, and the improvement science methodology we are recommending needs to allow individuals and their immediate colleagues to self-empower. For us, this is how we should gauge the method itself – **by the degree to which individuals and their colleagues appropriately take power to themselves – and then hold themselves accountable for their own performance outcomes over time.**

The method will be as good as the extent to which it becomes likely that motivated individuals and their teams engage with their immediate context – as a system – authoritatively taking charge of the systemic context. This necessitates a source of data that can tee-up the system's **continual monitoring** – as a platform for continual experimentation – PDSA cycle by PDSA cycle – so that the team may for themselves discover over time what works

and what doesn't. The data and the way it is presented needs also to be able to help the team establish whether or not their system is stable, and hence predicatable – otherwise the system's inherent **safety** must remain in doubt.

It is only the team member, working in the systemic context, under the direction of a clinical or managerial leader who possesses a method for working on the system, who can for all practical purposes maintain its stability – using that stability as a basis for enhancing the outcomes the system produces over time. The power lies in the new kind of accountability that becomes possible in this scenario – whereby **a team of well-led individuals in effect may hold themselves to account**. This trumps the relatively ineffective top-down-only sort of accountability that conventionally reigns down from politicians, lawyers and senior people – meeting their personal needs yes, but only very indirectly meeting the needs of those who they supposedly represent. Of course, this kind of accountability has a place and can't be completely replaced by the bottom-up kind of accountability we are alluding to, so the method we shall outline must include some upward reporting processes within some kind of standardised format – a "**liberating discipline**" (75) that supports the enablement of intrinsically motivated continual improvement at the front-line.

Such a scenario is difficult to conceive of without the pressence of a culture of learning, yet just such a culture *can* and will grow so long as it is nurtured with "**constancy of purpose**" (16). In time a willingness to learn from others, and in other contexts, will assuredly become ingrained as, like a positive virus, the ethos of learning and improvement spreads.

> **".. all *knowledge* and wonder (which is the seed of *knowledge*) is an impression of pleasure in itself " ..**
>
> **Francis Bacon** The Advancement of Learning (1605)

4. Not one, not two, but three measurement paradigms

In an excellent and still pertinent paper published in 1997 three clinicians: Solberg, Mosser and McDonald (65) saw the need to differentiate three faces of performance measurement; **Accountability**, **Improvement**, and **Research**. They had noticed that whenever medical groups started to talk about measurement "there was a great deal of confusion and emotion caused by widely differing understandings of the purposes and means of measurement." Warning strongly of the importance of being aware of, and allowing for, this confusion, they decided it would be helpful to set out their views on the characteristics of measurement for the three primary 'faces' or mindsets or (our preference) paradigms.

Their views are shown in the table below – an amended version of which has been widely quoted and used by such bodies as the NHS Institute and the DoH's National Quality Board – indeed in 2008 it was offered as a foundation stone for the World Class Commissioning initiative – the model neatly illustrating how in the NHS, measurement for improvement gets barely a look in. We prefer here however to return to the purity of the authors' original version. Albeit we have altered the order of the columns.

The **A-I-R measurement model** has the "Improvement" paradigm as the middle column, not only because then improvement appears as a kind of wedge between the A and R paradigms, but because the acronym is so apt – potentially *breathing life* into organisational efficiency and effectiveness. We also like to think of the Accountability paradigm as the natural default in conventional organisations, with the Research paradigm deferred to only as a distant afterthought. The improvement paradigm, more than acting as a wedge between the A and R paradigms may then serve as a bridge – enabling the other two paradigms to function more appropriately.

Allowing all three A-I-R measurement paradigms to properly play their part is vital, not only because "good science" requires it, but because it makes such a telling difference to the ability to sustain intended outcomes.

Because of the way that measurement has recently been used in the NHS, and because of how those who work in the NHS typically view such measurement, we strongly support this 3-way differentiation. We see measurement for improvement (column 2) as a vital underpining for the application of Improvement Science in the NHS.

The A-I-R measurment model

Characteristics of Measurement for Accountability, Improvement, & Research (23)

	Accountability	Improvement	Research
Who?			
Audience (customers)	Purchasers Payers Patients/ members Medical groups	Medical group Quality improvement team Providers and staff Administrators	Science community General public Users (clinicians)
Why?			
Purpose	Comparison Basis for choice Reassurance Spur for change	Understanding of (a) process (b) customers Motivation and focus Baseline Evaluation of changes	New knowledge without regard for its applicability
What?			
Scope	Specific to an individual medical group and process	Specific to an individual medical site and process	Universal though often limited generalizability
Measures	Very few Complex collection Precise and valid	Few Easy to collect Approximate	Many Complex collection Very precise and valid
Time period	Long, past	Short, current	Long, past
Confounders	Describe and try to measure	Consider but rarely measure	Measure or control
How?			
Measurers	External	Internal and at least involved in the selection of measures	External and usually prefer to control both process and collection
Sample size	Large	Small	Large
Collection process	Complex and requires moderate effort and cost	Simple and requires minimal time, cost, and expertise Usually repeated	Extremely complex and expensive May be planned for several repeats
Need for confidentiality	None for objects of comparison -- the goal is exposure	Very High (Organization and people)	High, especially for the individual subjects

Solberg L.I. Mosser G. McDonald S. – 1997 - (with the column order switched)

Two vital differences between columns 2 and 3 are the way that 'experiments' are designed, and the degree to which implementation in a specific context is handled. W. Edwards Deming personally witnessed many attempts at trying to apply the techniques aimed at assuring a count was correct, to the different problem of deciding what method to use in the future. He frequently expressed his frustration when seeing supposedly knowledgeable statisticians involved in applying tests of significance to analytic studies – seemingly entirely ignorant of the inherent contradiction. To Deming, in most studies, certainly the more important ones, we need to be able to predict the future – for this is the source of new knowledge about the cause system – requiring a column 2-type analytic approach. Measures of probability or uncertainty involve judgement that cannot be avoided, and 'p' values and "power calculation" (51) become nigh impossible.

Deming (16) also made an important distinction between two types of scientific study, using the words "**enumerative**" and "**analytic**", and it is important to include this distinction here alongside the A-I-R model. For any statistical study the ultimate aim is to provide a **rational basis for action** – enumerative and analytic studies differ according to where the action is taken. Deming summarized the distinction between the two types of study as follows:

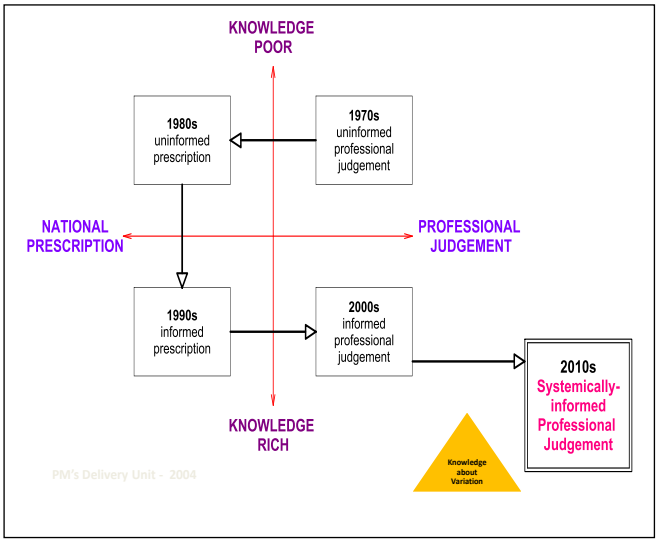
> **ENUMERATIVE STUDY:** a statistical study in which action will be taken on the material in the frame being studied – in a statistical study, the frame is the data set from which the sample statistics are taken.

ANALYTIC STUDY: a statistical study in which action will be taken on the process or **cause-system** that produced the frame being studied – the aim being to improve **outcomes in the future.**

Analytic studies correspond closely to column 2, whilst Enumerative studies are close to column 1. So what about the paradigm being expressed to do with "**Research**" in column 3?

Because the purpose of Research (column 3) is to validate an intervention, thereafter to either continue with the intervention or stop it, a **confidence interval** is a useful summation for the results of an enumerative study, but as Deming stated "a confidence interval has no operational definition for prediction" – this generally invalidates the Randomised Controlled Trial (RCT) whenever the purpose is to understand cause and effect over time. In practice only an analytic (column 2) approach can enable this.

Partly because a column 3 approach is hard to apply in practice, the only game in town for conventional organisations has traditionally been the (column 1) accountability paradigm. Against a historical background of next to no real-time measurement in the public sector prior to the 1980s, this is perhaps easier to explain:



This diagram was used by Prof. Michael Barber when Head of the PM's Delivery Unit in 2004. It neatly describes how during the last half century public sector measurement has been transformed. During the 1970s governmental intervention was in effect run on virtually zero data, but increasingly since then data collection has been imposed by central government, to the point that nowadays masses of data are available and waiting to be mined. The diagram anticipates the pending shift towards "*informed professional judgment*", an almost full-circle shift from the uninformed professional judgement of the 1970s that is now starting to replace the era of "*informed prescription*" and which according to Barber characterised the NHS's approach in the

By 2008 it was indeed more than timely for Lord Darzi to call for yet more informed professional judgment under the heading of "**clinical leadership**" – albeit if Michael Barber had still been in post we are sure he would be disappointed with the lack of progress to that point in history.

The final box is our addition. In the current decade we hope and anticipate that Improvement Science will contribute to the professions becoming able to think more systemically about using *their* data and evidence in a way that leads to decisions that prompt real and sustained systemic enhancement?

In future the main role for central government and intermediaries like the Care Quality Commission should in our view be to create a top down framework within which **bottom-up** monitoring (much of it clinically led) can happen, such that evidence-based professional judgment becomes the norm. This framework needs to be one in which **(self) empowerment, transparency** and team-oriented **accountability** can flourish.

Statistics is sadly viewed by many as something to be avoided, but with the right guidance and support – which if it's on hand and timely, needs only to be minimal – it will become possible for evidence-based systemic intervention and decision making to become the new culture.

Following the Francis and Berwick Reports, these distinctions, in particular the need to adopt column 2 thinking, have become a long overdue imperative – especially given that Simon Stevens' Five-year-view is projecting by the 5th year a funding gap of £30 billon, of which the NHS will itself need to find £22 billions internally – principally via increased efficiencies and system re-design. The need for enhanced quality whilst simultaneously making savings has

therefore never been greater – an achievement which from our perspective can only occur alongside the application of sound research methodologies.

Don Berwick's recommendation that NHS staff *"Learn, master and apply the modern methods of quality control, quality improvement and quality planning"* – is wholly consistent with column 2 thinking, and the new Rose Report (80) into what is preventing a culture change away from what is described as consistent underperformance emphasises the need for *"empowered employees with a willingness to improve.. not held back by a system characterised by poor practices for managing, training and developing people".* To engender the widespread use of measurement and the application of Improvement Science, it is clear that a very significant cultural and attitudinal shift is required by all those involved with the NHS. Don Berwick's report spells out the evidence-based methodology needed to take this forwards.

With his colleagues from the Cabinet Office's "Behavioual Insights Team", Ben Goldacre makes the point that there needs to be more of a focus on meaningful outcomes, as well as prognostically (following an intervention) a focus upon measuring these outcomes to establish whether or not the they are occuring as predicted. However, in suggesting methods that are mostly appropriate for the 'Research' paradigm alone, we suggest that they have missed the practcial vitality of the A-I-R measurment model – *a fuller critiqueof their T-L-A proposal is provided later when assessing the 3 case studies.*

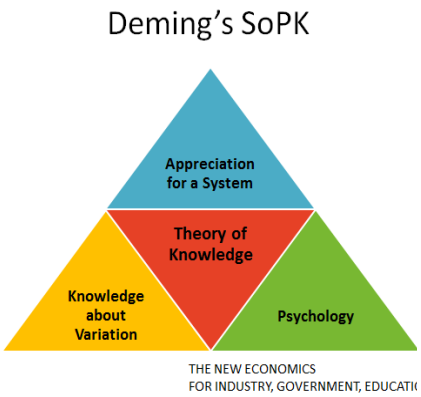
In recent years, increasing numbers of individuals working inside the healthcare system have started to ask just how it can be that an organisation like Boeing can deliver a faultless new aircraft that flies safely the first time of trying without even having been (as a whole vehicle) tested, and how it can be that the same aircraft will then go on to fly for thousands of hours without a hitch? Yet, in stark contrast why it is that when you turn-up at an airport to fly away on holiday the chances of being killed by the process are around 1 in 10 million,whilst in a UK hospital the chances are typically as low as 1 in 300? (21) For us as patients, providing a method for achieving enhanced and sustain outcomes is an imperative.

In addressing the potential for failure in healthcare systems, Atul Gawande (25) emphasises the need for a whole system approach based upon collaborative research methods. Even the UK's healthcare research institutions are coming around to a realisation that this is a primary need – Academic Health Science Networks , for example are currently busy acquiring and devloping the needed skills – albeit finding them currently in short supply.

5. Improvement Science

Improvement Science is a nascent field, emerging along with other fields to answer the call of the Berwick Report for improvement in healthcare quality and safety. The conceptual frame of reference for improvement science allows for a broad sweep of scientific study which in a complex adaptive organisational system like the NHS makes improvement strategies easier to apply.

Langley et al (40) in 1996 were the first to use the phrase "*the science of improvement*" in the first edition of their seminal book: *The Improvement Guide*. The term was used to build upon W. Edward Deming's System of Profound Knowledge (17) comprising the following 4 interrelated parts:



- **Appreciation for a system**: Understanding how the parts of a process or system relate to one another to create an effective systemic whole with an overarching aim.
- **Understanding variation**: Knowing the difference between variation that is an inherent part of a system and that which is a genuine outlier and not typically part of its usual cause system.
- **Psychology**: Understanding how interpersonal and social structures impact the performance of a process or system.
- **Theory of knowledge**: knowledge comes from prediction and theory, and accumulates as the nature of the gap between theory and reality is observed.

Without individuals and teams being practised at cycles of experimentation and standardisation, and being able to articulate their experience as a **theory of knowledge**, they are unlikely to be clear about their mental models or just how assumptions are driving their thinking. Opportunities to either improve or to sustain that improvement get too easily missed – risking the very health of the systems in which we all participate.

Without **knowledge** about the **variation** in our systems we each fail to understand how we spin our own reality. In ignoring or falsely interpreting data we can easily windup self-deluded, even wilfully blind. Better to use the variation that over time is inherent in all systems to gain an enhanced sense of what's actually happened, as well as what's really happening, and, if things stay the same, is most likely to happen. Without being adept at using variation to connect to reality – its past, its present and its future – we spurn the opportunity to design our system for better outcomes.

Without **appreciation for a system** we fail to see how things are connected and in the moment we miss the uniqueness of our own particular context. We also miss the importance of our personal maps, and those of others, for good sense-making – and in missing the sharing of our individual realities we miss the potential to collectively spot what really is causing outcomes – and we forgo the opportunity to act together upon the system as a whole. Interventions then fail or have undesirable consequences.

Without empathy for the **psychology** of others, and being able to build our understanding of what causes human behaviour, our ability to manage change is fundamentally impaired. We ignore the importance of intrinsic motivation and miss opportunities for proactively designing intrinsic motivation into everyday work. In order to reign supreme, bosses who know no better simply default to punishment or reward, or overbearing control – simply destroying the individual's self-motivation, and with it their organisation's agility. Far better that they enable their people to take charge of, and take pride in, their work.

People may think about science in very different ways. We especially want to inquire into how differently Ben Goldacre and Don Berwick are viewing it? A passing look at most any dictionary reveals three main definitions of, or ways of thinking about, science ..

1. **Science is the application of natural laws in order to invent new things;**
2. **Science is a body of knowledge expressed as laws and justifiable logic;**
3. **Science is a process of discovery: testing/ learning cycles built upon observation, measurement and accumulating evidence – via something akin to what Francis Bacon in the 17th century termed the "scientific method";**

The 1st definition is the one that is often uppermost in the public's consciousness – fuelled by daily media reports on whatever the latest scientific or technological breakthrough happens to be. The 2nd is probably closer to how most people remember the science they were taught at school – absorbed well enough to be regurgitated in an end of year examination, or in later life at a pub quiz night. Whilst in most schools only lip service is paid to it, the 3rd definition however feels very different to the second – imbued with movement and change – a process. This third kind of science is sadly too rarely experienced even by post-school undergraduates, and given that only a very few individuals in their working life get to experience in-work scientific research, it is not a definition or mindset that many recognise. In any case, conventional organisational life tends to be more about the achievement of goals and the measurement of individual compliance, than the enhancement of systemic learning (58). Sadly, individuals may reasonably expect nothing else, for adult life simply repeats that experienced as a child (22).

Students in school for example are so used to being measured and formally tested, that when entering a world of work, having one's feet held to the fire by a boss may seem like the most natural of things. Workers fast become institutionalised via appraisal systems that are designed to work wholly or largely top-down in order to make it easier for managers to control behaviours and ensure accountabilty – to them – even though most people suspect , and evidence supports (9) that most of the variation in appraisal comes from the rater rather than the ratee. None of this scenario has therefore anything much to do with what Bacon would have thought of as science.

We are concerned that science is dismally conveyed to children and students, the majority of whom leave formal education without understanding the power of discovery, nor gaining any first-hand experience of the scientific method. If science were to be defined around discovery, and learning cycles, and built upon observation, measurement and the accumulation of evidence – then good science could be viewed as a process rather than merely as an externalisd entity. These things comprise the very essence of what Berwick refers to as Improvement Science – embodied by the Institute of Healthcare Improvement (IHI) and in the NHS's Model for Improvement.

We are suggesting that in a world where many at school have been turned-off science, if rendered accessible the propensity to be scientific in our daily lives – and at work – can make a vast difference to the way people think about outcomes and their achievement. This is especially so if those who take a perverse pride in saying they avoided science at achool, or who freely admit they don't do numbers, can get switched on to it.

Using Deming's SoPK as a springboard, Langley et al stressed two critical ideas that helped define the science of improvement. First was the idea that all improvement comes from developing, testing, and implementing changes – the role of measurement being to set-up feedback (learning) loops so as to gauge the impact of those changes over time as environmental conditions vary. Second is a recognition that the subject matter expert them self plays the lead role in developing changes and establishing the conditions for testing. In launching the "Model for Improvement" – which subsequently has become a touchstone for the NHS – they are in effect favouring the 3rd of the above definitions of science whereby science is viewed as a process.

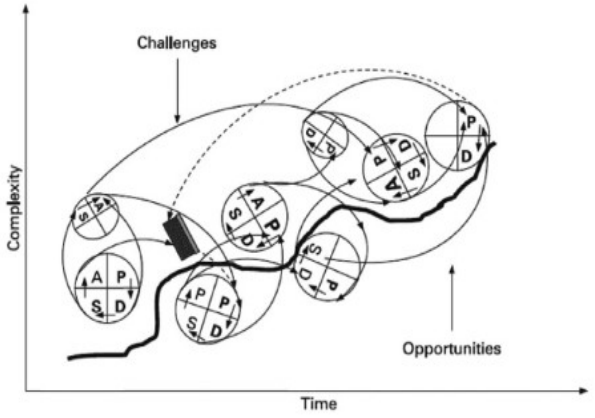
Langely et al are saying that improvement and innovation remain inhibited unless ..

- Good science can not only be permitted, but proactively encouraged.
- Good science can be anchored in real life practice and feasibility via a firm connection to context.

- Good science is constantly supported across the organisation – in the short, medium and long term.
- Staff at all levels are involved (self-empowered) and that any systemic aspect that may have a potential bearing upon healthcare outcomes are legitimised as on-limits.
- Specific local improvements can be made as long as the wider systemic context is properly considered.
- Can be led by local managerial leaders with minimal training, supported via on-going coaching by experienced "Guides", combining to create an environment characterised by personal and team-based discovery.
- The Guides know when complexity or breadth require wider input or advice, and where to get it.
- The fundamentals of improvement science are applied to the issues being considered, however big or small – see the seven propositions below.

Without the support of good managerial leadership to create the right environment for improvement to take place, the application of this Improvement Science approach is difficult, though by no means impossible, provided that as a minimum ..

- Staff can be equipped and consistently encouraged to take part in improvement, and to suggest problems to address or ideas for improvement.
- Honest data, can be collected in real time by the staff themselves, that bad news is not buried, and the messenger never shot.
- Learning from within the team, and learning from elsewhere, is consistently encouraged – this is for example very different from enforced copying from other contexts with insufficient appreciation for the systemic differences.
- It can be accepted that not all improvement works, and that multiple iteration is needed – as illustrated in this diagram illustrating that potential for learning *if permissible* needs to occur cyclically and iteratively. This is the reason why the phrase "continual improvement" is preferred to "continuous improvement", because inquiry is a process that never proceeds at a constant rate.



(51)

The case for such managerial leadership and such a culture is made by Berwick (8) and strongly supported by us.

Incidentally, earlier we referred to Harari's observation that in the modern world the subject of statistics is now core in many academic environments. However, unless the teaching is especially inspiring, in our experience making statistics compulsory can for many students be an eternal turn-off – especially if all that is taught is an approach that feels impractical or disconnected from real life. One of the joys of improvement science is that people can immediately apply it to their own real world context. There are nowadays examples even of patients and (carers) doing just that and transforming the conversations that they have with their physicians (63) – as well as the outcomes achieved.

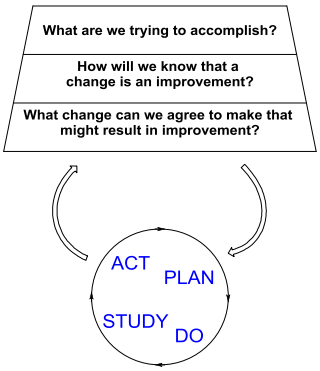
Good science requires that all three paradigms are appropriately in play (even column 1) according to **whatever works** to engender and sustain improvement. In conventional organisation life it is the 2nd column that is missing, which if activated can act as a bridge between the other two usually strange bed-fellows, for the world of work is usually so dominated by column 1 thinking that column 3- type approaches are either denied or avoided, usually becoming merely the post-hoc preserve of academia.

Perla et al (54) describe the following seven propositions of improvement science (IS), which we prefer to think of as foundation stones ..

1. **IS is grounded in testing and learning cycles.**
2. **IS embraces a combination of psychology and logic.**
3. **IS has a philosophical foundation is conceptualistic pragmatism.**
4. **IS employs Shewhart's theory of cause systems.**
5. **IS requires the use of operational definitions.**
6. **IS considers the context of *both* justification *and* discovery.**
7. **IS is informed by Systems theory.**

expanding on each of these propositions in turn:

(1) ***Improvement*** **Science is grounded in testing and learning cycles.**



The Model for Improvement

(API 1996)

Building upon the Shewhart/ Deming concept of continual improvement, the consultancy API (29) 20 years ago launched the "**model for improvement**" – now widely adopted in and across the NHS. The model requires individuals and teams to be clear about what's important and to start by stabilising their system so that it can naturally deliver it – thereafter to use that stability as a platform for improving the system via experimentation.

Each experiment is performed against a baseline measurement, to test clearly stated hypotheses about how each proposed change will lead to things becoming better. If against that baseline no improvement occurs, or if even things get worse, there is still a prize: Learning.

Knowledge about what things, or combination of things, causes improvement then accumulates, and is carefully retained so it may be passed-on to new team members. In this way systemic enhancement can be sustained. Albeit, whenever external pressure is applied e.g. via arbitrary numerical targets, particular care and strong local leadership is needed if the model for improvement and its sequence is not to break down.

Most of us as children are taught that Knowledge is something external to ourselves – and that we each become learn*ed* if we read the right books and successfully reproduce that learning in examinations – see the 3 definitions of science above. Few of us get to pick-up on the very *process* by which that learning was created or discovered prior to it becoming written down and published – or to pick-up on the *principle* that we are each standing on the shoulders of giants aiming to stretch previously gained knowledge to new heights.

It is therefore easy to miss that our predecessors had been at great pains to replicate their experiments – in a variety of contexts – in order to verify any newly emerging knowledge. Individuals and teams in organisations too need to remember that each context is to some extent unique, and that pre-existing science must therefore be justified, and assumptions declared and justified with local evidence.

New knowledge is uncovered by curious individuals who are intrinsically motivated to pursue a process of discovery, and it is continually being verified and rediscovered. It is for these reasons for example that whenever pilot projects seem to have delivered success, science is needed whenever subsequently "rolled-out" to other locations – because each now location is a different system. It is also the reason why the word "continual" is preferred to "continuous" because expecting learning to be anything other than iterative is a key reason why it is so frequently stunted.

(2) **Improvement Science embraces a combination of psychology and logic.**

Carl Jung (25), perhaps anticipating Harari's historical description of our species' evolution, pointed to a vital human psychological continuum: **sensing ↔ intuiting** (*S* ↔ *N*). He presented the "**S**" pole as "that psychological function which transmits a physical stimulus to perception" and the "**N**" pole: as "that psychological function which transmits perceptions in an unconscious way".

This same continuum is core to the Jungian-based Myers Briggs personality type indicator (MBTI) which models those possessing a strong "S" preference as likely to get irritated whenever fiction is predominating over what they would prefer to be witnessing things and events: concrete, fact-based conversation that deals with the here and now. Those with a strong N preference are conversely categorised as more interested in abstract ideas, theories and future possibilities. The tension between Harari's two realities described on page 5 is therefore neatly echoed by these two poles.

There are of course some advantages in being able to use one's intuition. N-types are typically less than 30% of the population yet they often rise to the top of organisational life – many politicians too scoring as N – but as Ben Goldacre suggests there are also serious hindrances ..

> *"Intuitions are valuable for all kinds of things, especially in the social domain: deciding if your girlfriend is cheating on you, perhaps, or whether a business partner is untrustworthy. But for mathematical issues, or assessing causal relationships intuitions are often completely wrong because they rely on shortcuts which have arisen as handy ways to solve complex problems rapidly but at the cost of inaccuracies, misfires and oversensitivity."*
>
> Ben Goldacre *(17)*

Actually everyone is prone to such traps, and if you think this doesn't apply to you, think again. In his book *Thinking, Fast and Slow*, the behavioural economist and Nobel Prize winner Daniel Kahneman (26) describes going through seven years' worth of data on the performance of 25 top wealth advisers at a Wall Street firm. It turned out that there was no such thing as a consistent gift for stock-picking and the advisers might as well have made their choices by tossing a coin. When Kahneman and his colleagues presented their findings to the company's executives, they were roundly ignored. Kahneman further evidences a number of other archetypal ways in which our intuitive responses steer us badly, or literally render us blind – leading to misconception and poor decision-making. "Seeing is believing" turns out to be a highly unreliable heuristic.

There are other classic ways in which we may fool ourselves ..

- **Confirmation Bias** – selectively choosing history or events that match preconceived notions, rather than forming opinions through empirics or reason. Evidence to the contrary is subsequently considered chance.
- What Nassim Nicolas Taleb calls the **"Narrative Fallacy"**: constructing stories around facts, but then later believing the stories such that facts become accommodated into the stories. Sapiens look for history to fit an easily explained narrative, for example by perversely believing that whatever chosen variable is the entirety of the truth, we then each become closed to any further information. In providing an operational definition of a signal – identified via a system behaviour chart – Walter Shewhart offered a practical way out of this conundrum.
- **Smart after the event** – taking past events as predictably determined and then adding-in baseless, opinionated commentary.
- **Externalising evidence** – ascribing the viewpoints of others to evilness rather than hearing the argument being made.

These four aren't the only failures in our processes for forming opinions, but each simplifies the world into a discrete, understandable sequence of events in a way that simultaneously wraps and warps our view of reality. We each have a natural proclivity towards joining the dots – even if this means we have to invent non-existent data in order to make our story flow. We need a **standardised storytelling tool** if only so that our peers may challenge us and via collective inquiry – to help us to recalibrate – see section 9.

The aim is better personal decision-making and maintained motivation. Improvement Science requires that both ends of the *S* ↔ *N* continuum are well-functioning – and that individuals are able to move with agility between its poles – possessing high "***bandwidth***" – for without *both* objectivity *and* imagination in play, reality gets distorted. As Oliver Moody recently notes .. *"One of the best things the Greeks did for us was to challenge* **mythos***, myths and received wisdom, with* **logos***, the conclusions of rational inquiry. Not necessarily the truth, but* ***a better and stronger story****.. but most people's instinct for taking narrative short-cuts through difficult facts is undiminished. We are what the late Sir Terry Pratchett once called Pan Narrans, story-telling chimpanzees."* (34)

Our central proposition in this paper is that what humankind needs now, more than ever before, is data-rich, evidence-based, storytelling – requiring skills in both mythos and logos .. the best of both psychology and logic.

(3) **Improvement Science has a philosophical foundation of conceptualistic pragmatism**.

Deming's thinking was strongly influenced by the philosopher C.I. Lewis who in his own way spent much of his life trying to integrate the *S* ↔ *N* poles, and who in 1929 (30) put forward a position whereby empirical knowledge is dependent upon three things: (i) a sensed 'given'; (ii) the constructive activity of the mind; (iii) a set of a priori concepts which the individual as agent brings to their interpretation of what is observed, the product of the individual observer's social heritage and their particular cognitive interests. He called this philosophy "**conceptualistic pragmatism**".

To square a similar circle Carl Jung by this time had already come up with a second core continuum, this one between **Thinking** and **Feeling** (T ↔ F). He defined thinking as "that psychological function which, in accordance with its own laws, brings given presentations into conceptual connection"; and feeling as "primarily a process that takes place between the ego and a given content, a process, moreover, that imparts to the content a definite value in the sense of acceptance or rejection."

Science is therefore difficult to conduct because individuals make their judgements partly cognitively and partly based upon inherited or evolved values and beliefs – MBTI F-types preferring the latter. Whereas the S↔N continuum concerns the process by which data and information are *gathered*, the T ↔ F continuum concerns how that data and information is then *processed*. The MBTI model predicts that individuals possessing a strong "T" preference absorb what they've taken-in via the *S* ↔ *N* continuum and then process it according to "proven" principles and logic, or laid-down laws and policies – typically preferring to analyse things/ events according to clear criterion or publically declared standards. T-types therefore respect most what to them appears objectively real, whereas people with an "F" preference embrace subjectivity, preferring to judge people, things and events according to their personal values – and importantly attaching their own meaning to the data they are observing. What feels right to an F-type must therefore be so – fairness having more to do with humanity than metred justice – rules and laws existing only to be bent if that is what it takes to maintain personal integrity.

Bandwidth and agility between the T ↔ F poles is again helpful, but this empirically comes only with maturity, with an individual's longitudinal development (44), so the science of improvement warrants a method that enables individuals to achieve what has been called a "power of balance" (52). Fortunately because improvement science is, or ideally ought to be, largely conducted in a team-based context, as long as team members *between them* possess a broad range of psychological preferences, and as long as a productive team process is followed, then sound team-working may naturally achieve the necessary balance. In our experience, having one person in the team who possesses "late-staged" (44) maturity can make a world of difference to a team's ability to achieve a scientific balance, but because such individuals are relatively rare, a sound methodology is vital.

So, what did Lewis mean by "conceptualistic pragmatism? The standard positivist view of empirically acquired data and information has been that observation, experience, and experiment serve as neutral arbiters between competing theories. However, since Thomas Kuhn's impactful book (28) in the 60s, there has been persistent support for the position that these methods are influenced by prior beliefs and experiences. Consequently it cannot be expected that two scientists when observing or experiencing the same event will make the same theory-neutral observations. The role of observation as a theory-neutral arbiter is therefore diminished – even if there were prior

agreement about methods of inference and interpretation. Science, and improvement science, is thus forced to accommodate the normative nature of reality.

A **concept** may also be thought of relatively – if *your* hours are felt to be twice as long as *mine*, and *your* pounds twice as heavy as *mine*, these are not differences that can be tested by assigning physical properties to things, the way the world is experienced is nonetheless real. For Lewis, knowledge begins and ends in experience, and knowledge of some thing or event requires that the verified experience be actually experienced. For the **pragmatist** therefore, verifiability as an operational definition (or test) of the empirical meaning of a statement requires that a speaker knows when (or not) it is valid to use someone else's described experience as a representation of reality – properly allowing for its consequences in all conceivable situations both real and hypothetical. Whereas positivism ultimately reduces the relation between meaning and experience to a matter of logical form, pragmatism allows us to ground meaning in conceived experience.

(4) **Improvement Science employs Shewhart's "theory of cause systems".**

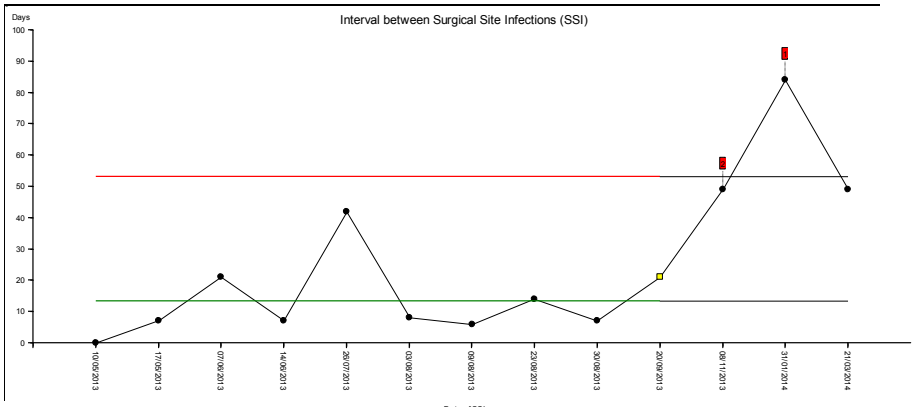
Walter Shewhart (1891-1967), known as the "father of statistical quality control, was a physicist, engineer and statistician. Having joined the Western Electric Company at their Hawthorne Works in 1918, and discovering there that quality was restricted merely to inspecting finished products and removing defective items, he proposed (1924) that quality would be better assured via a tool he had created to display systemic variation over time. He called the tool a **control chart**. The tool and the thinking behind it directly impacted both Deming and Juran who worked with him in the Hawthorne plant.

Generating knowledge about a system requires knowledge about its particular context, and an understanding of the system's behaviour – past, present, *and* future – which if displayed **over time** by those working in and on that system, arises out of the thoughtful interpretation of data being monitored. The metric being monitored may have been carefully selected by them to connect to a particular outcome or to a sub-system that is seen as vital to that outcome; alternatively the metric may be no more than a KPI that has to be reported upon. Either way, converting that data into a control chart – or as we prefer to call it a "***system (or process) behaviour chart"***– almost always transforms the conversation. Usually the first thing to be discussed is whether the system appears to be stable – and given that a system exhibiting signs of instability is inherently unpredictable, such a discussion is vital for any "managerial leader" worth the title, for any system in this state is likely to cause future trouble.

Shewhart wanted those working in and on a system to be able to use its inherent variation to better diagnose cause and effect – and this became the basis of his life's work: exploring the theory of cause systems – work which eventually had a transforming impact in many manufacturing sectors, and eventually in some service sectors too.

Shewhart's great discovery was that *some* variation is predictable, within limits, but not all of it. The purpose of his chart was – in real time – to split the predictable data variation from the unpredictable data variation – dividing the predictable "*common-cause variation"* from the unanticipated, emergent or previously neglected phenomena representing variation outside the system's historical experience base. Shewhart called this second kind of variation "*assignable cause*" variation, later popularised by Deming as "*special-cause" variation,* to indicate that the frequency/ severity of what are usually one-off events is fundamentally different and should be treated as such. If a special cause event is being signaled, some sort of systemic change has probably occurred, offering a timely clue about cause to those working in and on a system. A narrative emerges that is all important: **a knowledgefull story** – a potent interweaving of the reality that is being *imagined* together with the *objective* reality that is being presented as self-evident fact – like warp and weft.

Here for example is a chart of data gathered by a surgical team where the individual values are showing the number of lapsed days between incidents of post-operative infection. The control limits have been calculated using the first 11 data points (incidents) – and at the 12th data point the system changed. How do the team know this change is really improvement?



The team know the system changed because within two more incidents a special cause is flagged – under the Western Electric Rules the 13th point triggering the 12th. Usually when a special cause is signaled it is unexpected, but in this case the team working in the system were expecting it because they had actually decided to make a design change to their system to see if they could reduce incidents of post-operative infection – this was the Plan stage in their P-D-S-A cycle. In this case the chart and its special cause, confirm the effectiveness of the changes they made.

Whatever the change was, their experiment seems to have worked. The team is using a powerful mix of column 2 and column 3 thinking to deliver *and sustain* a significant reduction in post-operative infection rates for their patients.

Shewhart's legacy has been profound – people are now able to distinguish assignable signals from the common cause noise in their system in a way that directly prompts systemic understanding. The system behaviour chart is like a stethoscope for listening-in, tuning-in, to behaviour that otherwise would be missed. It aids diagnosis, *and*, by flagging potential future trouble, also aids real time prognosis. Deming and Shewhart were passionately interested in supporting workers and managerial leaders to be less reactive and more considered in how they intervene in their systems. They wanted them to be aware that their very own actions may inadvertently make things worse – they were keen for example to be able to demonstrate how continual system-adjustment in reaction to non-conformance (counterintuitively) usually increases variation and worsens quality – a syndrome later referred to by Deming as "**Tampering**".

(5) **Improvement Science requires the use of Operational Definitions**.

Earlier we pointed to Deming's notion that "profound knowledge" rests upon having a level of appreciation for a system. He defined a system as "***a network of interdependent components that work together to try to accomplish the aim of the system***" (11) . To better understand our own particular system we might therefore decide to map it, indeed in order to better appreciate how it flows or how the components fit together we may well draw upon *several* maps – exploring for example how it is structured or how it appears to different stakeholders. But how should we monitor its behaviour and performance over time? What metrics should we choose? And as it changes, which metrics might act as the best barometers?

We need to make these choices wisely and with humility because as Deming emphasised, some of the most important aspects of a system cannot be expressed numerically, and even when this *is* possible, "***there is no true value of anything measured or observed***" (11). When it comes to measurement therefore it is vital to understand the measurement process itself, and the "operational definition" it depends upon.

Just what did Deming mean by the phrase "no true value"? Of course, there can be a 'true value' of something that is true by definition, for example 2 plus 2 equals 4 is true by definition. But in real live systems, reality is much harder to pin down. In an economic system for example, stating a single statistic to wholly represent the rate of unemployment at a single point in time is difficult enough, let alone one that may be relied upon over time. All we

can do is agree and share an operationally defined value – and this essentially must include a precise description of the measurement process. Any alteration in definition or in the method of measurement, is likely to yield a different value. Not a wrong one, just a different one. Repeatability and reproducibility over time is the aim.

> *"There have been three scientific numbers that reflect the true speed of light.. I do not know which one is right, all I know is that it is damned fast."*
>
> Deming (10)

An operational definition, to be effective, needs 3 essential elements ..

1. CRITERIA – what is the aim of the system, and what does the customer say is wanted?
2. MEASUREMENT – how will we measure the things relating to customer wants?
3. DECISION RULE – what defines acceptable in terms of these measures?

When you only have the first, all you have is a basis for an argument. It is only when you have all three parts that you have progressed from wishful thinking to an operational definition. When applied to the question of measurement consistency we will need to come up with a statement of what good consistency is, how to test to see if it is present, and how best to interpret the test results.

When all this is shared with the community likely to be working in the system, or dependent upon it, it becomes possible to reduce ambiguity – to as close to zero as possible. That is the goal.

(6) **Improvement Science considers the contexts of *both* justification *and* discovery.**

Earlier we described how science can be defined in more than one way, and how sadly the education system, at least as a process of discovery, diminishes science. The 3rd of the three definitions we offered is the one most pertinent to the science of improvement, built as it is upon the **scientific method** – we therefore need an overall approach that as a minimum aids discovery.

The scientific method is an ongoing and cyclic procedure, which starts with observations about the natural world. As Sapiens we are naturally inquisitive and we love to gossip, so it's hardly surprising that questions arise about things seen or heard, and that then ideas develop (hypotheses) about why things are the way they appear. The best hypotheses lead to predictions that can be tested via experimentation followed by further observations about the system. In general, the strongest tests of hypotheses come from carefully controlled and replicated experiments based upon empirical data. Depending on how well the tests match the predictions, the original hypothesis may require refinement, alteration, expansion or even rejection – this is the A-step in the P-D-S-A mnemonic. If a particular hypothesis becomes very well supported a more general theory may then be developed.

P-D-S-A, as a cycle, is not meant to imply that the first step is necessarily to make a prediction i.e. the "Plan" step. In most circumstances the initial cycle starts with "Study" as in S-A-P-D. Discovery and justification are *temporally distinct processes* – for in the beginning the system needs to be described, description precipitates **discovery**, and each specific discovery then requires **justification**.

Our method, during the process of discovery, needs to be empirically robust – whereas the justification process, the way we do our experiments, needs to possess logical integrity. Adopting the empiricist view, one can claim to have knowledge only when one has a true belief based on empirical evidence. Empirical evidence – experience shared for example via a Flow Chart, or a Gantt chart – may be taken as data or information that justifies a belief in the truth of a claim, but for us, to be worthy of the term "science" more than this is needed: a method or procedure

characterised by inquiry. The map is not the territory, the map itself is a hypothesis requiring corroboration – and not merely at one point in time, but continually over time. Our chosen method therefore has to include systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses. This means that teams need a variety of tools in their kitbag: be it a scattergram, pareto chart, cause & effect (fishbone), run chart, histogram, system behaviour chart, etc, etc.

(7) **Improvement Science is informed by Systems Theory**.

Systems theory is the study of systems – any system: as small as a quark or as large as the universe. It aims to uncover archetypal behaviours and principles that can be applied across all disciplines and all fields of research for example the principle of "self-regulation" and how via feedback, systems are capable of self-correction. Examples of self-regulating systems for example include the physiological systems of our body, local and global ecosystems, planetary climate systems, and systems of human learning – both individually and organisationally. Peter Senge, in his famous book "The 5th discipline" made "systems thinking" pertinent especially to business life, and the late Donnella Meadows too sought to make **thinking in systems** (33) an essential aid in solving problems – both personal and large-scale – practical and relevant in the everyday. There are however several Systems Thinking schools, and Senge and Meadows generally adhere to the **system dynamics** school – which was initiated by Jay Forrester in the 50s with a particular focus on **cause & effect** and **feedback loops**, and the driving of complex systems. This is the kind that are especially relevant to the world of healthcare because of the need for enhanced safety and quality. Forrester's method starts with the system's structure – the many circular, interlocking, sometimes time-lagged relationships and flows between its individual components. Importantly, because there are properties-of-the-whole which cannot be found among the properties-of-the-components, the behaviour of the whole often cannot be explained in terms of the behaviour of the parts.

It has always been important that healthcare provides quality outcomes at reasonable cost, but it is comparatively recently that achieving these things has been considered a discipline all its own. It is becoming clearer that productivity in healthcare – the ratio over time of outputs to inputs – depends heavily upon the way that a particular healthcare system has been designed, and upon how it is subsequently maintained and improved – and this doesn't mean only the buildings and the kit, but also the way that activity (and funding) flow, and the way that the people who work in the system continually solve problems to optimise each system.

It is also becoming clearer that productivity is an outcome that depends on how safety and quality are being achieved – aspects of a system that often are not necessarily seen as key drivers. To many the notion that quality and low cost go hand-in-hand remains a counterintuitive one.

Improvement Science (IS) helps ensure health services operate in a way that makes it possible to offer the best possible clinical care and thus improve people's health, and IS methodologies are currently being developed by many healthcare organisations around the world. Four years ago the charitable Health Foundation, a leading champion of improvement science in the UK, identified 67 centres claiming a 'special interest' in improvement science – 36 of them based in N. America, 21 in the UK, 8 in mainland Europe, and 2 in Australia. Today there are more.

In the UK, the DoH have recently declared that the organisation of health improvement requires rationalisation and everyone in the field now awaits their deliberations. Meanwhile the landscape of NHS improvement and innovation support remains fragmented and cluttered, and the theories and methods for achieving rigorous evidence-based practice continues to be embryonic with healthcare professionals still relatively unskilled in how they are applying the new thinking. Education and training programs are only just beginning to systematically include IS in the education of future healthcare professionals. In writing this paper we hope to provide a philosophical touchstone to all these efforts so that a coherent methodology can properly evolve.

In their T-L-A proposals Ben and his colleagues at the cabinet Office are offering their own methodology, one also intended to support good science, and it is exciting that they want to include all the public sector – not just the parts directly responsible for health. It is our hope that our paper – especially the **A-I-R measurement model** and the **seven IS foundation stones** (54) – can contribute to a fleshing-out of their thinking in a way that makes sustained systemic enhancement practically accessible to all those who work in public sector systems?

6. A more detailed critique of "Test, Learn and Adapt" (TLA)

> ***"The RCT is a powerful, perhaps unequaled, research design, to explore the efficiacy of conceptually neat components of clinical practice: test, drugs and procedure. For other crucially important learning purposes however it serves less well. Fans of traditional research methods view RCTs as the gold standard, but RCTs do not work well in many healthcare contexts".***
>
> **Don Berwick** (7)

Ben Goldacre (et al's) paper (27) suggests an overall approach of Test, Learn, Adapt (T-L-A), adding 9 more specific steps to operationalise it..

Test

1. Identify two or more policy interventions to compare (e.g. old vs new policy; different variations of a policy).
2. Determine the outcome that the policy is intended to influence and how it will be measured in the trial.
3. Decide on the randomisation unit: whether to randomise to intervention and control groups at the level of individuals, institutions (e.g. schools), or geographical areas (e.g. local authorities).
4. Determine how many units (people, institutions, or areas) are required for robust results.
5. Assign each unit to one of the policy interventions, using a robust randomisation method.
6. Introduce the policy interventions to the assigned groups.

Learn

7. Measure the results and determine the impact of the policy interventions.

Adapt

8. Adapt your policy intervention to reflect your findings.
9. Return to Step 1 to continually improve your understanding of what works

Randomised Controlled Trials (RCTs) in particular are proposed as at the heart of the method; "*the best way of determining whether a policy is working"* (29). We very strongly support the aims behind this proposal, but for reasons we will now explain we think our proposals for 'good science' via the 'Improvement Science' approach offer a better, more complete approach. RCTs have a place in this broader approach but are not the default.

Our key criticisms about the T-L-A approach are ..

1. It does not enable, nor encourage, involvement by improvement of staff at all levels – or in all aspects of Healthcare
2. It is incomplete, in that it omits the important steps b*efore* and *after* the three T-L-A steps, respectively, the prompting of understanding and ideas about the actual system currently being run by a team of healthcare professionals and, in a wide range of different situations, the implementation and monitoring of systemic improvement.
3. It is too easy to regard any 3-step process in a linear way (64). Some people for example replace P-D-S-A with Plan-Do-Review thereby losing the inherent potential for learning – something which *if permissible* must to be effective occur cyclically as iterative inquiry.

Moreover, RCTs, as the T-L-A authors freely admit, require academic support. This immediately limits the type and spread of activity and risks disenfranchising most healthcare staff from participation. This in itself will significantly reduce the amount of improvement that can and will take place. There are also additional risks to widespread parallel improvement activity and we will address these later, however, the loss of potentially huge improvement energy is of particular concern to us.

RCTs are quite a complicated tool, especially if there are several options to explore, or several metrics of interest – which is commonly the situation. They are cost effective when used well but do imply quite formally conducted improvement, often organised centrally. Designing the randomisation usually requires considerable thought and skill. They also require detailed calculations on sampling, sample size and probability error types. This is why expert support is necessary.

RCTs can lead to the testing being done not on the process as is, but on the theoretical process. This is especially likely when the people involved in everyday work are not fully involved or engaged.

Once an RCT has shown an intervention to be 'of value', it still needs to be implemented successfully, and then continued long term. Anyone who has attempted to implement a good idea knows what dangers there are at this stage. The T-L-A approach includes no simple methodology for monitoring the results at different locations, nor over time (including the long term). In fact, during each T-L-A iteration, RCTs are suggested as The Tool to be used.

Furthermore, at the start of the improvement process, we note significant omissions. It is our very clear experience, that in Healthcare and in a wide range of situations, it is only those who work in the system who can know how it currently works. Those who are more distant inevitably have an incorrect or incomplete understanding. Apart from in practice often failing to deliver improvement, interventions suggested from afar run an increased risk of unwanted side-affects. The costs and disruption often lead to failed improvement, unnecessary chaos, or worse. It's little wonder that staff get disenchanted by initiatives.

In summary, to be considered science therefore, it is to us essential that ..

- A representative set of people who work directly in the system are fully involved
- The current approach to the system is studied to establish how it currently operates, and what causes problems, delay and variation?
- Ideas of causes and of solutions are discussed, and appropriately explored and prioritised by those involved in the system (in some improvement work RCTs may have a role in this)
- Only after the above may selected interventions then be tested
- With a small amount of training, much of all this can be led by line managers.

Depending entirely on Column 3 for the organisation's learning is tantamount to outsourcing it .. *"Measurement for research is typically too slow, too expensive, too elaborate to be useful for improving processes in single clinics or hospitals" (68).*

P values too should be interpreted with considerable care and skill, a level of care and skill that is rare in managerial and clinical leaders. *"A p Value measures whether an observed result can be attributed to chance, But it cannot answer a researcher's real question: what are the odds that a hypothesis is correct? Those odds depend on how strong the result was and, most importantly, on how plausible the hypothesis is in the first place."* (54). There are several ways for the unscrupulous to cheat with p values, for example *"if you measure a large number of things about a small number of people, you are almost guaranteed to get a statistically significant result"* (9).

The Department of Health has recently suggested a reinterpretation of the AIRmm model (51). This attempts to makes it even clearer that things have to change, and that column 2 "Improvement Science" is needed as an alternative to the overbearing (column 1) compliance approach that whilst holding people to account, motivationally appeals largely to the extrinsic, and frequently feels overly judgemental – directly prompting reactivity in staff as well as an entrenched moribund culture.

SIDE BAR

Conventional organisational life is dominated by the column 1 (accountability) paradigm – the primary aim being to motivate individuals and sub-groups to meet some arbitrary performance target – often with some kind of extrinsic incentive. As an undergraduate in the early 70s Julian learned about column 3 (research) statistics and, blissfully unaware of its limitations, was forever keen during vacation jobs to try out some of the techniques – e.g. Critical Path Analysis. His bosses always initially thought of him as some kind of oddity – at least that is until it became obvious that performance had been greatly enhanced, at which point the typical organisational response was to find ways to prevent the ***"talisman with magical powers"*** *from moving on to another job. There was rarely any interest in learning about how the apparently magical results were being achieved. Painfully aware that as soon as the day came for him to leave, he knew that everything would return to how it was before his arrival. Nowadays he wonders whether some of Ben Goldacre's frustration is borne of that same painful feeling – that if only people could learn to apply a few statistical basics then the world would rapidly change for the better – and even be sustained in that better place.*

Julian spent the next 15 years of his working life getting used to a dual-mode of working: one to fit with how his bosses viewed the world (mainly column 1) and the second to fit with the statistically-based view that he had been taught in academia (mainly column 3). He learned to keep the two modes in two very separate boxes.

Aged 36, Julian's then boss decided he might prevent him from leaving by offering to fund a part-time MBA, so somewhat reluctantly Julian trundled off back to university. It was with a leading institution for MBAs (Bath), so maybe they would have some integrative answer to his dual-mode purdha? It was not however until he had completed the 2 year programme that a neighbour (a chief engineer with a local bus company) stopped him in the street and said .. "Julian, I'm being sent on a course about something called Total Quality and have been told to read up on two chaps called Deming and Juran. Since you're the guy with an MBA I thought I'd ask you about them?" Somewhat embarrased he was forced to admit that he had never heard of either of them, but agreed he would look them up.

Purchasing a copy of Deming's "Out of the Crisis" (16) Julian immediately experienced epiphany. Finally .. the answer he had been seeking! Deming not only seemed to offer a wholly new way of thinking about statistics, but had an explaination as to why western businesses were in decline – how our organisations were actually set-up to treat our customers disrespectfully – how the only way to reverse the decline would be to target quality improvement, which then like a chain reaction would lead to reduced costs and better customer outcomes – simultaneously! Deming wrote about how it is possible to achieve cultural transformation in any organisation prepared to recognise they are in crisis, and provide ***"constancy of purpose"*** *for their employees. He explained just how it could be that Japanese-owned companies were so easily able to raid western markets – and how 40 years earlier he had personally taught them how to do it. Most importantly, Deming spoke about how nurturing intrinsic motivation in any organisational system – not destroying it with well-meaning yet highly counterproductive HR policies – is the key to sustained improvement.*

Twenty five years later, and now retired, Julian finds himself working as a volunteer in his local NHS and has discovered that healthcare too is finally waking-up to the reality that whilst not being driven out of business, there truly is change (maybe even transformation) in the air – and not just the usual deckchair-style restructuring that comes along after every change of government. Culturally, it feels to him a little like being transported back to the paradigm that dominated western engineering/ manufacturing in the 80s, yet because the NHS is everyday confronted with the life and death issues, just as much as in those days whole organisations were in existential crisis, there seems to be a great deal of intrinsic motivation just waiting to be awakened. In particular, the NHS is having to face up to the fact that despite scores of performance targets and huge financial commitment from tax payers, as well as masses of investment into health research, patients all too often are still on the receiving end of poor quality outcomes. The Francis inquiry into the Mid-Staffordshire deaths with its 290 recommendations, together with the Berwick Report which specifically examines how to address patient safety in healthcare organisations, are heralding a whole new beginning.

In column 2, magic becomes possible because suddenly the other two columns (paradigms) – which used to be disconnected – now make practical sense. Allow us to demonstrate by developing Atul Gawande's RCT study concerning patients diagnosed with Stage IV Lung Cancer referred to earlier on page 7.

Massachusetts General Hospital – a column 2 approach

We are not privy to how the Massachusetts General Hospital proceeded after receiving the results of the RCT study, but let us suggest how the whole exercise might look from an Improvement Science perspective. We wish to explore this from two aspects.. first, how does it support our critique of the T-L-A approach and the suggestion that RCTs are the 'best method'? And second, how might the 'improvers' at Massachusetts General Hospital (MGH) have approached their task of enhancing patient outcomes using improvement science methods?

One of our key criticisms of T-L-A is that the 9 steps omit any discussion of implementation of the findings in 'normal practice'. It must be obvious to most that the palliative care approach is not yet normal practice and that there are many obstacles to it becoming so in MGH – let alone in other hospitals. The improvement cycle is therefore far from complete.

It is very likely that the consultants in MGH would have wished to consider these results very carefully before moving forward to an implementation phase. Firstly they would likely have checked the benefits and costs suggested by the results. They would have needed to confirm that the overall affect was worth the effort of change. They would need to consider the impact on resources of more palliative care contact (how much, how often and for how long), less chemotherapy, less bed time, … they would need to make sure that the resources can be reallocated as needed, and to plan to do so. They would also need to agree how to continue monitoring the new treatment – measures of some of length of life, suffering, time in hospital, amount of chemotherapy treatment – they should do this to confirm that the benefit continues. In the event that it does not, they would need to review why, and consider whether this had been due to implementation issues, or doubt over the original conclusions. Measurement in real time and simple graphs (run charts, or better: XmR system behaviour charts) would allow this to be considered in an ongoing way – in near real time.

They would need to convince all the professionals involved that this change was well-founded and significantly beneficial, *and* to explain the new approach in full with careful explanations for all aspects of the proposed new approach. Without both these steps; the change would most likely end up either poorly implemented or not implemented – leading to poor outcomes.

Whilst some aspects of these implementation steps are touched on in steps 7 and 8 of T-L-A, we suggest it is being considerably underplayed.

In other hospitals, implementation is even more uncertain. How would they hear of the results? How well would they understand the trial and its implications? Would they believe it applies to them? Would they be able to implement it well – adjusting the approach well enough for their particular circumstances? Even if they were to hear of it, will this change be a priority compared to other issues that are considered important?

There will be contextual differences between hospitals and between consultants that could affect the findings. It is to be expected and appropriate for other hospitals and consultants to be initially sceptical. The level of scepticism should depend on how well the original trial was conducted and explained, the reputation of those conducting the trials, and how many relevant differences between hospitals are being recognised. However, as experience of applied Improvement Science is gained, practitioners should be more and more willing to believe that 'such results are likely to be relevant to us'. This should not result in simply trying to copy the MGH. They should consider the MGH results and implementation approach, adapt it to their situation and then plan a simple but good scientific test of the change. They could use the simplest Improvement Science approach.

This involves –

- assembling their current relevant and carefully selected data,
- plotting this data over time (run or XmR chart) to establish the current baseline performance
- making the change, in the same area as for the collected data
- plotting the new results in real time
- using the simple rules underpinning the XmR chart to indicate whether or not this has led to improvement
- comparing the results with those of MGH, if not similar (better or worse), exploring possible reasons for this and acting accordingly.
- sharing results, warts and all, with others including MGH

MGH no doubt ended up doing some or all of what we are suggesting, but if following an IS-type of approach we would expect to see evidence of the following:

1. **Selecting a clear aim, from all those possible**
   All care can be improved; there are always disappointments and improvement ideas around. Within MGH cancer care department, some discussion at ideally one of the regular prioritisation meetings, would have led to agreement that the care of Stage IV lung cancer was appropriate for investigation at this time. MGH would have had some disappointments to reduce or already some improvement ideas being put forward. An improvement team would have been set up which included representatives of key roles within the current care process. An improvement aim would also have been agreed at this time; possibilities include better care, better patient satisfaction, better outcomes, longer life, lower costs – and one or more of these could have been explicitly included.

2. **Studying the current system as broadly and deeply as needed. Selecting one or a few causes of variation, delay or problems to address**
   The depth required here depends on whether they are being problem or improvement idea driven. If the latter, they can move more quickly into step 3. However, it is always worth ensuring the whole current process as practised is well understood. If there are many different processes in operation, we may need first to use the first cycle of 9 steps to explore which current practice is best. Subsequently here, we assume there is a reasonably consistent current process. Flowcharts are a great way to do this. They also help check that current data collection is good for the purpose of this trial. The current data should now be studied with simple charts and analyses using 'tool' that are easy to use and understand; run charts, XmR charts, Pareto charts, separation of data into groups of interest. These analyses will enable improvement ideas to emerge or for the existing improvement idea to be tested for feasibility and relevance.

3. **Selecting one or a few improvements to try**
   In the MGH, example, they may have decided that patient understanding and control over their care was the key change of interest. They may have decided to include a palliative care consultation be included to support this. (There are other possible scenarios.) This could have been based on aspects of their analysis in step 2.

4. **Testing these in the real situation or one very close to it**
   They still needed to agree the details of how the new process will operate; where things will be the same and where they will be different. They needed to agree when the change will take place. They needed to agree what data they would collect live and how it would be collected. They will need to explain very carefully to all those involved what they need to do differently when. The analysis of this data in real time needed to be put in place. These analysis methods will determine whether a change of outcome has resulted; the bigger the change the quicker it will be confirmed. Of course, no improvement (or worsening) may result, so some formal review dates should also be agreed. They will be comparing the process and data before and after the change.

***SIDE BAR*: Don Berwick and the Science of Improvement**

In 2008, whilst President and CEO of the Institute for Healthcare Improvement, Don Berwick published a seminal paper in the Journal of American Medicine entitled "The Science of Improvement" (7). Berwick is Clinical Professor of Paediatrics and Health Care Policy at Harvard, and specialises in the management of health care systems, with an emphasis on using scientific methods, evidence-based medicine and comparative effectiveness research – aiming at improving the trade-off between quality, safety and costs. In the paper he makes a number of profoundly important points about science and the research methodologies and philosophies that underpin it. "Where is the RCT?" is for many purposes the right question, and often can be, but for many others it's the wrong one, a myopic one. A better one is broader .. ***"What is everyone learning?"***

Healthcare's main goals of relieving illness and pain require research methods of several different kinds – basic, clincial, systems, epidimiological – which needs to be determined by the kinds of knowledge being sought. It is important therefore to adjust the chosen research methods to fit research questions in order that the vital job of redesigning healthcare system can then proceed – and for crucially important learning purposes RCTs really don't work well – especially in many healthcare contexts. Berwick names 4 good science prerequisites:

(1) Embrace a wider range of scientific methodologies – "Many assessment techniques developed in engineering and used in the quality improvement – statistical process control, time series analysis, simulations, and factorial experiments – have more power to inform about mechanisms and contexts than do RCTs .."
(2) Reconsider thresholds for action on evidence. The traditional rules of inference (p> 0.05) embed an aversion to accpeting a null hypthesis even when it it true.

(3) "To study a linear, mechanical or natural, tightly-coupled causual relationship .. (for example to determining benefits of Beta blockers for heartfailure) an O-X-O design like an RCT may be exactly correct. O-X-O stands for: observe a system (O), introuduce a perterbation (X) to some participants but not to others, and then (O) observe again. Properly measured, the changes in outcome are, with a calculable degree of certainty, attributable to the perterbation". But with social/technical system improvement (e.g. via a new social programme) "multi-component interventions, some of which are interpersonal, all of which are non-linear, in complex social systems.. ***other richer, equally disciplined, ways to learn are needed****". A better way is to equip the workforce to study the effects of their efforts, actively and objectively, as part of daily work is via C-M-O (Context + Mechanism = Outcome), often a better alternative to an O-X-O design, in particular for interventions into social systems requiring a method that ensures the particular social and cultural context drives the particular (local) intervention mechanisms chosen. In aiming to eliminate possible bias the O-X-O approach intentionally seeks to remove local knowledge of context and mechanisms. This is wasteful for almost always the individuls who are making changes in care systems know more about the mechanisms and context than 3rd party evaluators can learn with RCTs" .. and "Insensitive suspicion about biases, no matter how well intended can feel like attacks on sincerity."*

(4) When providing external support to practitioners be especially careful about mood. ***Academics and frontline caregivers best serve patients and communities when they collaborate on mutually respectful terms.***

There is an "apparent gap between science and experience (which) lies deep in epistemology." When evaluating an intervention it is dangerous to ignore any epistemological incongruence in the ways we seek to get at "truth". The process of social change (over time) is multi-component and complex .. and "in such a terrain .. an RCT is an impoverished way to learn .. (people) who use it as a truth standard in this context are incorrect. Berwick quotes Pawson and Tilley (53) who coming from a realist position, argue that RCTs exemplify the overused O-X-O paradigm, leading all too often to "heroic failure, promising so much and yet ending up in ironic anticlimax.. The underlying logic .. seems meticulous, clear-headed and militarily precise, and yet findings seem to emerge in a typically non-cumulative, low-impact, prone-to-equivocation, sort of way". We find out whether an intervention works, but gain all too little knowledge as to how and why – and in what contexts.

In short, Don Berwick is telling us that an exclusively Column 3 approach (including their centre-piece the RCT) has significant drawbacks, and thinking more widely about research methodologies in whatever particular context we are dealing with is likley to lead to better designs, better interventions and better data. Pragmatically speaking, RCTs are like reaching for a spanner and finding a lump hammer in your hand.

Note, the before-and-after comparison is much simpler to set up and analyse than for an RCT. Care is needed around the data definition and collection but the same is required for an RCT.

People may have a concern about biased judgement; this is not a worry if the data is objective rather than subjective for example “length of stay in hospital”. It is also helped by the commitment to ongoing monitoring.

In the next section we illustrate how being conscious of the A-I-R measurement model makes for a more powerfully integrated approach to measurement and that RCTs have a useful situationally specific role to play when deployed alongside improvement science and a nuanced approach to accountability – all combining in a way that adds up to better science.

7. The right tools make a telling difference: 3 case studies

With the aim of bridging the 3 paradigms in the Solberg model, in 2010 a statistical tool called BaseLine® (24) was created to instantly convert data that has already been collected for column 1 purposes into data that can be used to identify systemic change over time – as if collected for column 2 purposes. The tool is based upon the Individuals & Moving Range Chart which prompts an immediate inquiry into cause and effect over time. It is a tool designed to be easy to use even for a person who thinks of themselves as someone who "doesn't do numbers". The user simply opens their spreadsheet, highlights the date/time column together with the column containing the metric of interest to them, and clicks copy. Next they simply open the BaseLine® software and click Paste .. *et voilà*, the chart appears. In that instant, the variation over time is displayed, and the conversation – be that with yourself or with others – is often instantly transformed. In many cases merely using such this tool initiates the very kind of paradigm shift being required by Berwick-style transformation – one characterised as being essential to a systems thinking "Learning Organisation"(22) and one where cause and effect over time becomes a continual discussion, prompting both diagnosis and prognosis. Those drawn into making such a leap then quickly learn that more pertinent data could have been collected to further aid their inquiry into what's actually happening (in real time), and what is most likely to happen (in the immediate future).

Collecting the data that you yourself – even better if also agreed with your team – think is currently most useful to create knowledge of how the system is running, is in itself an act of self-empowerment – the most valuable kind of all (18) – and incidentally the most efficacious basis for self-management, not just by healthcare teams but by individual patients too.

In short, the BaseLine® tool makes it practically easy to operate within the Column 2 paradigm. It also aims to make the second step/question in the Improvement Model (on page 19) – ***how will we know a change or intervention we make equates to improvement?*** – a practical one for people to address. Culturally, it may still be a challenge for people and organisations wedded only to a Column 1 mind set, but the tool may also be directly used to convert data collected for Column-1 purposes into a chart that identifies systemic shifts over time – and this render these cultural barriers instantaneously surmountable.

Simply having access to data in near to real time – data that indicates the impact one's intervention is having – prompts a dramatic shift in the type of conversations that people working in and on the system may have. Pretty soon, remarkable outcomes can get initiated. Such is the powerful potential inherent within Improvement Science. Here are 3 examples of IS in action that collectively illustrate how data may be used to evidence *both* sustained systemic enhancement, *and* to generate engagement by the people most directly connected to what is systemically happening in real time.

1. A surgical team using existing knowledge established by column 3-type research as a platform for column 2-type analytic study – to radically reduce post-operative Surgical Site Infection.

2. 25 GP practices are required to collect data via the Friends & Family Test (FFT) and decide to experiment with being more than merely compliant. In 2 practices they collectively pilot a system run by their PPG (patient participation group) to study the FFT score – patient by patient – as they arrive each day. They use IS principles to separate signal from noise in a way that prompts the most useful response to the feedback. Separately they summarise all the comments every 2 months and feed their analysis into the next PPG meeting. The aim is to address respectively both "special cause" feedback and "common cause" feedback in a way that can prompt sensible improvement activity in what most feel is an over-loaded system.

3. A patient is diagnosed with NAFLD and receives advice from their doctor to get more exercise e.g. by walking more. The patient uses the principles of IS to monitor what happens – using the data not just to show how they are complying with their doctor's wishes, but to understand what drives their own

mind/body system. The patient hopes that this knowledge can lead to better personal decision-making and maintained motivation.

Case Study 1

Elective abdominal wall hernia repair is one of the commonest planned surgery procedures with over 60,000 operations performed annually in the UK – mostly as day cases. Post-operative wound infections – thought to originate with the patient's own skin flora – occur typically in about 6% of groin hernia surgery and usually become manifest in the first two weeks – leading to pain, poor healing rates, and delayed return to work, sometimes requiring further treatment with antibiotics, even hospital admission or more surgery.

One hospital Department of General Surgery (24) receives an email in May 2013 announcing the new availability of an antimicrobial-coated suture – together with a proposal that they should adopt a standard skin preparation using alcoholic chlorhexidine which has been evidenced by a RCT conducted between 2004 and 2008. The results of the RCT showed a statistically significant benefit of using chlorhexidine-alcohol skin preparation compared with povidone-iodine. No evidence to support the efficacy of the new kind of suture was offered.

Several reactions to this proposal ensued: one surgeon questioned the scientific sense of introducing two interventions at the same time; another surgeon questioned the evidence for the skin-prep having found separately published data that showed the exact opposite (iodine proved better than chlorhexidine); another surgeon questioned that the suture could have such a significant impact and queried the evidence; the managers too had their say, asking whose budget the money would come from to pay for the extra cost of the more expensive sutures?

One of the surgeons then discovers that the Department of Health, using the same RCT evidence, had in 2010 proposed a "care bundle" approach – making several interventions at once, as a package. He wonders about using a before-and-after study on such an approach: measuring the incidence of surgical site infection (SSI) as a baseline, then introducing the new protocol, and measuring the post-intervention SSI? The evidence on offer seemed to point to the efficacy of a care bundle approach when specifically designed for day case hernia surgery?

On further study, he discovers that over the years a variety of related RCT studies had been conducted, yielding only marginally conclusive results. Even a meta-analysis had revealed no clear consensus. He discovers however that "care bundle" interventions in other forms of surgery have often been shown to be effective. So he assembles an anti-infection bundle of care – comprising the seven simultaneous interventions (see below) that seemed supported by at least some positive evidence. Incidentally, the theatre process itself had also recently been redesigned and now delivered an increased patient throughput of 30% with no additional resource.

<u>The anti-infection Care Bundle</u>

1. Hair removal by clipping
2. Single-dose of prophylactic antibiotics on induction
3. Intra-operative warming to achieve normothermia
4. Antiseptic skin prep with aqueous providone iodine
5. Subcuticular wound closure
6. Wound sealing using skin glue
7. Moisture resistant dressing (lasts> 48 hours)

Pragmatically, this list was the maximum number of interventions considered allowable within the team's existing operational and financial framework. Perfect compliance by all those involved in the process of implementation

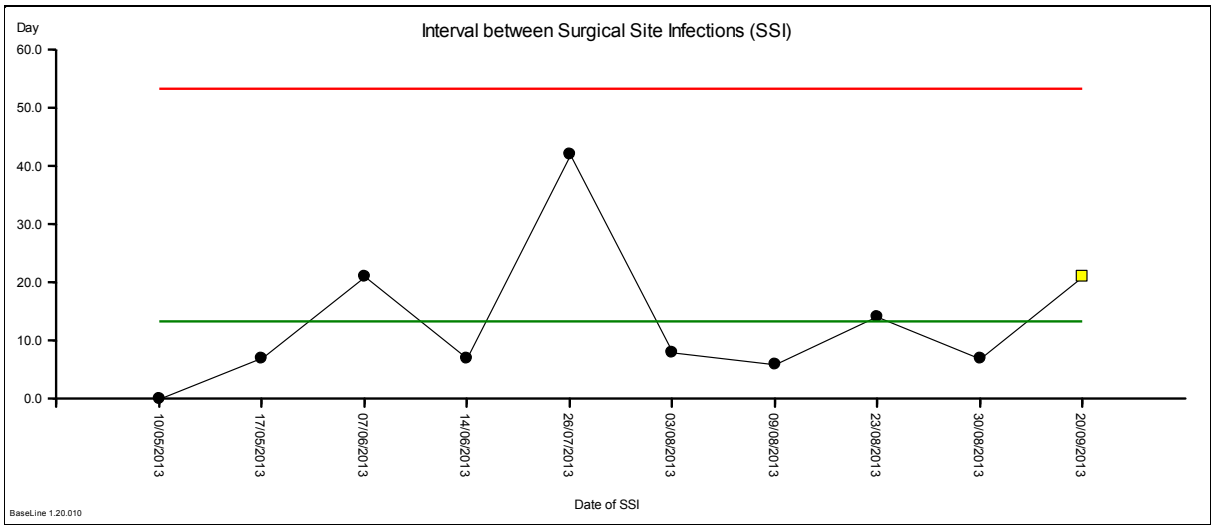
would clearly be vital, so a "Safer Surgery" checklist – provided as a standard by the World Health Organisation – was deployed.

The baseline SSI data was collected from April 2013-August 2013 (168 days) the timing influenced coincidentally by the move to a newly opened day surgery theatre. With the baseline data in the can, the care bundle was then introduced from the beginning of September 2013.

No formal analysis of the outcome data was attempted for a further period of 168 days in an attempt to minimise "expectation bias".

The outcome?

The chart below is an Individuals & Moving Range Chart (XmR) showing the time (in days) that lapses *between* SSI

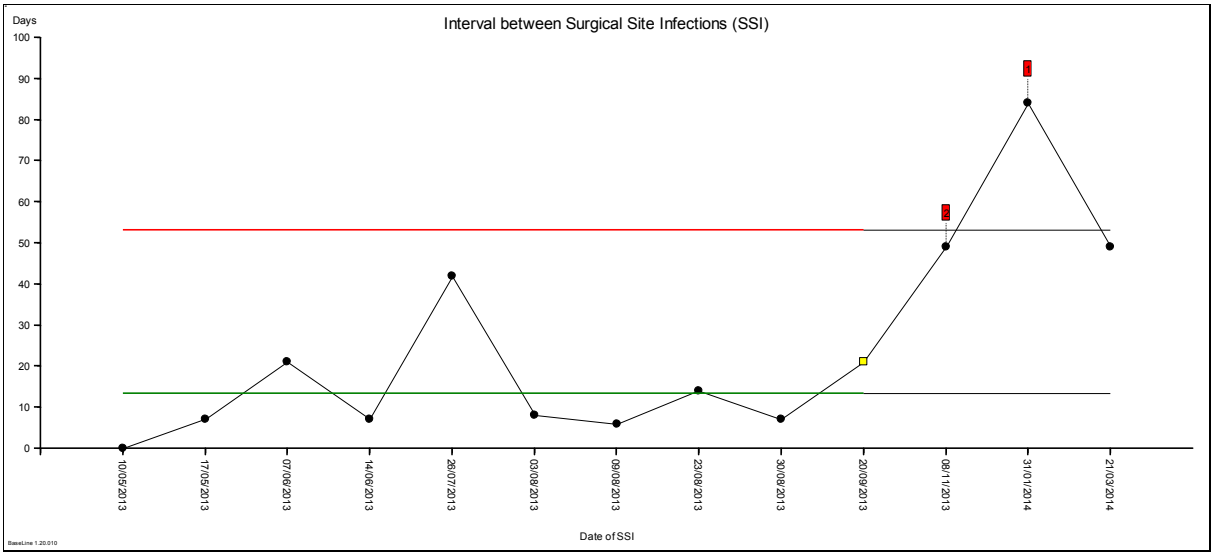


Segment	Start	Finish	Mean	Sigma	Count	LCL	UCL	Stable
1	10/05/2013	20/09/2013	13.30	13.30	10	-26.6	53.19	Yes

Note: To keep things as simple as we can we are not using the Moving Range chart for additional support.

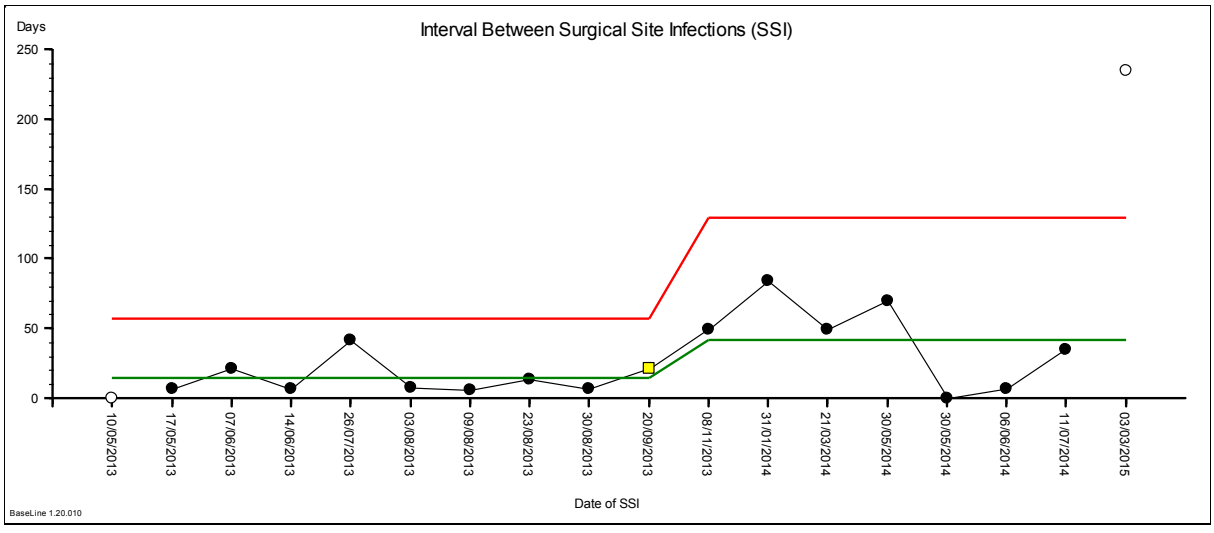
Before the care bundle intervention it is not unusual to have to wait only a few days before a fresh infection event. In the pre-intervention period (up until the point marked □), out of a total of 118 operations, 8 resulted in an SSI event being recorded – a not exceptional rate of 6.8% for this type of surgery. Before the intervention the average period between SSI incidents is 13.3 days.

At the point of the intervention the chart's limits are locked so that the impact of the intervention can be seen.



The XmR Chart offers an "**upper control limit**"(UCL) of 53.19 days and the tool, BaseLine®(), uses what is known as the four Western Electric Rules to identify a flag : for example 2 indicates that Rule 2 has been triggered – indicating that 2 out of 3 successive points are more than 2 sigma away from the mean. In this case two flags are triggered because the chart clearly evidences that the system has been operating differently since the care bundle intervention commenced.

How large is the change? After the next seven SSI incidents the following chart is revealed ..



Segment	Start	Finish	Mean	Sigma	Count	LCL	UCL	Stable
1	10/05/2013	20/09/2013	14.78	14.18	9	-27.8	57.33	Yes
2	08/11/2013	03/03/2015	42.00	28.96	7	-44.9	128.9	Yes

But that of course isn't the end of the story. Because the team are plotting time-in-between incidents they are always awaiting the next one. At the time of writing (30/03/2015) it has been no fewer than 235 days since the last incidence, so the next point on the chart will be at least as high as the one indicated by the circle ○. A Rule 1 flag will be triggered, indicating a possible second shift – albeit care will be needed because time-in-between data can sometimes be exponentially spread. Only the team themselves will sensibly be able to ask themselves what (if anything) has changed since the last incident (11/07/2014)? It may also be a transient cause. If it turns out to be a

sustained improvement, over and above that achieved via the care bundle intervention, this can't be known until several more incidents occur. Indeed, it now appears that the team may start even to think about selecting a new way of measuring their system's improvement? Because the energy for improvement comes from within the team, more bottom-up than top-down, selecting new metrics whenever the circumstances require it becomes a possibility – a capability that is a vital part of sustaining the improvement ethos.

To reiterate, the interval between infections clearly rises and the change is sufficient for this to be a signal on the team's chart, and after only two data points. Such a signal is very reassuring to all involved and encourages them to maintain the new regime.

All in all, there can be little doubt that patient outcomes have been improved – that a dramatic reduction in unintended adverse outcomes has been achieved. Furthermore, because in measuring the initial outcome the team decided to followed up all patients around 10 days after their operation – previously patients were simply referred back to their GP for any further follow-up – it becomes immediately obvious that the data collection process itself is prompting enhanced patient outcomes since it is now apparent that all but a few patients are religiously adhering to the standard advice of "*don't return to work for 6 weeks*" – as if some kind of sell-by date. In 90% of cases the opportunity is now taken to tell them that they can get back to work much more quickly. Repeat prescriptions, hospital readmission rates, and incidences of further surgery – all of these are now evidenced as falling sharply.

Case Study 2

On December 2014 all GP Practices in England have been required to collect feedback from their patients using The Friends & Family Test (FFT) – which centres upon a single standard question – that following the Francis Inquiry has been in use in hospitals since 2012:

How likely are you to recommend Sunnyside Surgery to friends and family if they needed similar care or treatment?

Extremely Likely	Likely	Neither Likely nor Unlikely	Unlikely	Extremely Unlikely	Don't Know

Please tell us why you selected the above response

Patients are prompted to respond to this questionnaire after each visit – and promised anonymity.

Coming at a time when there seems to be no end to the additional pressures placed upon GP practices, three North Somerset PPGs (patient participation groups) decided it would be useful to be able to handle this new expectation in a way that minimises resources whilst also generating some genuinely useful feedback for the practice – in the hope that this might lead to better outcomes for patients?

In the months prior to the start of the scheme, most of the remaining 22 practices in the county opted to focus merely upon avoiding being inadvertently non-compliant, preferring to pay £40 per month (plus £0.15 per returned questionnaire) to a private company to process the data on their behalf. This would buy them a standard kit: questionnaires, envelopes, ballot box, and an A3 poster for display in the waiting room. Some of these Practice Managers were openly indicating that, because no Response Rate target was being set by their paymasters NHS England, they would aim therefore at being only minimally compliant whilst provoking as little feedback as possible. This may appear to be a cynical response, but when under pressure to deliver performance whilst also required to pursue a relentless stream of new initiatives, it might equally be seen as a pragmatic response.

In a spirit of self-empowerment however, the three PPGs decided to offer to provide this same service for free to their practices – as well as to process the feedback in real time, whilst applying improvement science principles to

make the most of the feedback. In effect they are acting as a pilot for the rest of the county, and a review meeting with all the county's Practice Managers is being set-up by the lead Practice Manager, to take place by the end of 2015. The pilot is also fully supported by NHS England.

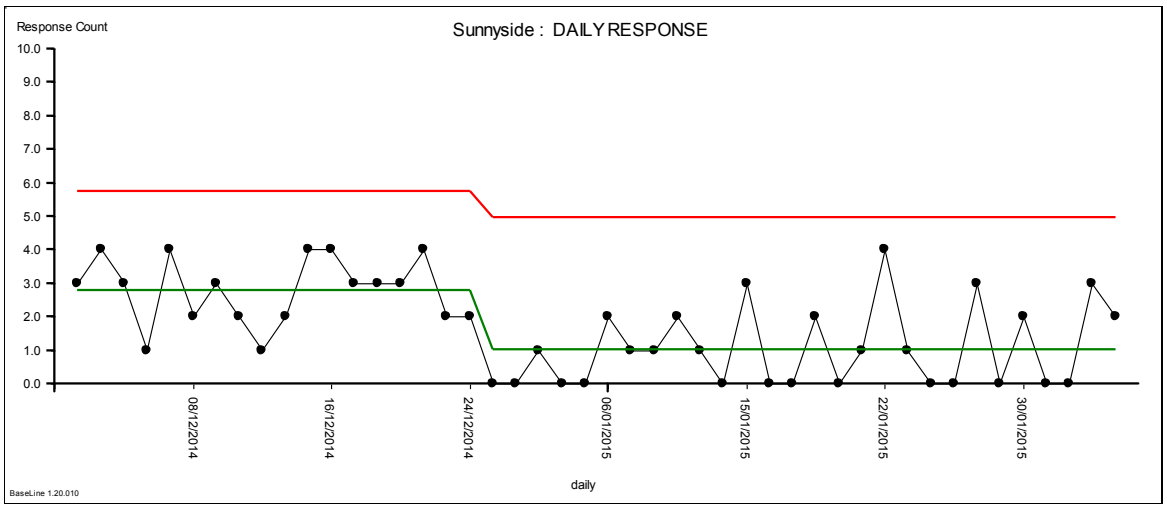
The outcomes thus far

As stated in the Berwick Report .. *"Patient feedback is instrumental to measurement, maintenance and monitoring of safety; Feedback should be collected as far as possible in real time and responded to as quickly as possible".* In this spirit, the aim of the pilot is to obtain rich feedback that prompts useful and timely action, but not over-reaction. Considerable energy was put into the design of the spreadsheet so that data can be entered easily – as far as possible in real time – and so that stratification and analysis by a PPG member can be straight forward. The intention is to be able to have a standard spreadsheet (latest one below) for use by all of the county's practices – which will make benchmarking possible.

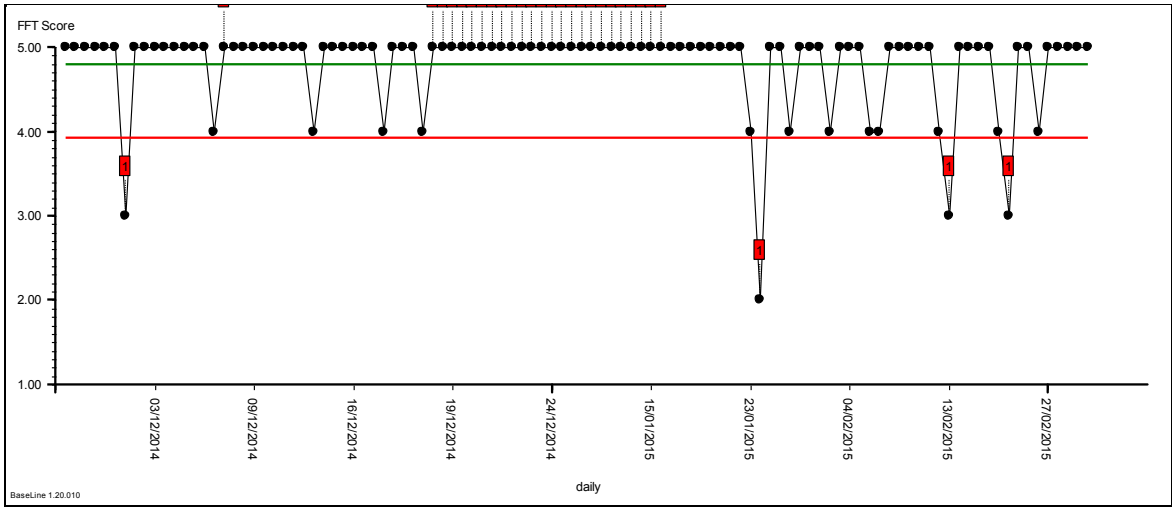
Note that a separate study has shown that older patients give higher FFT scores, so when there is sufficient data the spreadsheet makes stratification by age easy to do and analyse.

Patient No	DATE FORM RECEIVED	age group	FFT Score	Collection method (practice / website)	Comments Why selected above response	Comments One thing to change
65	22/01/2015	E	5	p	The doctors, nurses and reception have been excellent.	None.
66	22/01/2015	F	5	p	Efficient and friendly.	
67	22/01/2015	G	5		Having been at this surgery since it opened, I have been treated exceptionally well having had my fair amount of problems over the years also I find it has a friendly atmosphere.	None.
68	22/01/2015	E	5	p	Very well run, friendly practice.	More appointments outside working hours.
69	21/01/2015	E	5	o	It was fantastic to have the doctor ring me with my test results rather than make me come up to the surgery. He also gave me lifestyle advice which immediately prompted me to get more daily exercise.	Online access to test results
70	23/01/2015	G	4	p		Would like to get through on the telephone - always engaged then all appts have gone.
71	28/01/2015	U	2	p	Appointments are always running late!	To not be late on every appointment!
72	28/01/2015	E	5	p	Because the Drs and Nurses are kind and find time for one.	
73	28/01/2015	D	5	p	Dr Horner is excellent.	
74	30/01/2015	C	4	p	It's closer than Riverside to our home.	Make general appointments slightly longer to avoid delays.
75	30/01/2015	F	5	p	Always had the best of treatment!	
76	03/02/2015	F	5	p	Excellent appointment system. Good caring GPs and Nurses.	Some gentle soft-piped music in the waiting room - to allow confidential conversation and distract from other noise eg coughing.

The pilot has been running for 4 months and despite various attempts to increase it, the Response Rate (RR) has stubbornly remained low – less than 0.5% of patients.



Note that the RR was higher in the first month, and then dropped to a level of around 1 patient per day after the Christmas break. In contrast, hospitals, who have been required to use the F&F Test for 3 years, are typically getting a 20% RR – perhaps because there are targets set? The low RR has meant that stratification is not as yet possible, so all of the data is being lumped together in the chart below ..



This somewhat simplistic approach to differentiating signal from noise – patient to patient – is being justified as the *liberal* application of IS principles in a way that pragmatically handles the need to identify outliers given a context that is required to be patient-focused and puts safety first, and where data arrives (so far at least) only infrequently.

Co-production

The updated spreadsheet is sent daily to the PPG member (a patient) handling the processing. Many of the FFT scores are the maximum "extremely likely" to which we are awarding a score of 5. Thus far, all the scores of ≤ 3 are showing up as outliers. In each case a discussion ensues – between the Practice Manager and the PPG member – as to whether any immediate action is appropriate?

NHS commissioners are nowadays required to involve patients in decision-making, not just in a token way, but in a way that means decisions are "co-produced" with patients making significant contributions to the delivery of a valid and meaningful outcome. The aim is to be able to demonstrate a high degree of co-production in the pilot, and subsequently to be able to replicate this in other practices. The kinds of discussions now taking place between the practice staff and the PPG's patients is prima facie evidence of co-production, albeit the quality of this co-production will require continual review.

Having separated signal from noise, all the verbal comments is then analyzed and presented in a way that can be discussed at each bi-monthly PPG meeting. Here is how the data is being presented ..

please tell us why you selected that particular F&FT response (88% awarded the top score of "extremely likely")

Friendly/ welcoming/ polite/ helpful	Kind/ responsve/ caring/ attentive/ patient/ considerate/ supportive/ committed/ respectful	can get appointments easily	excellent professionals	excellent reception	consistent/ efficient / reliable	informative / preventative culture	went the extra mile	convenient	clean	good waiting area
33	40	14	40	24	35	3	3	1	1	2

if I could change one thing it would be ..

nothing (fine as it is) / changes appreciated	more/ better seating	need gentle piped music to distract from eves-dropping conversations	*Not being addressed by my christian name.*	better reading material/ more reading for men	early morning telephone appointments system is poor	appointments not running to time / more explanation needed	need more appointments 2 or 3 days in advance	need more evening/ weekend appointments	faster appointment access to doctor of choice	need better appointment system for on-going treatment	make appointments longer to avoid delays	add online apointment booking for nurses	text reminder needs time and date	need night visits by doctors	more assistance with lifestyle changes	style of professional	online access to test results	friendlier front of house staff	parking
36	3	2	2	3	8	10	4	3	9	2	4	1	2	1	2	2	4	4	2

Outlier (low score) Comments:

Stick to appointment times, improve booking system.	*Treatment and GP ok. (Just time waiting).*
	To not be late on every appointment!
	Two consecutive visits where the appointment has been more than 45 minutes late.
	Appointments are always running late!
	Waiting time to see GP
	Because I don't like waiting.

At the meeting it was felt that action should ensue around Appointments – the nature of this is yet to be decided upon. It is expected that a PPG sub-group will study the existing system, and report back on possible changes.

It was felt that even with the initially low RR, we are learning a lot about what our patients are currently saying, even if the data is as yet barely statistically representative. To increase the RR, three PPG members volunteered to get patients to complete a FFT questionnaire on say one day each month – and to take the opportunity also to raise the profile of the PPG. Two of them were able to attend on the chosen day and, in just an hour, 30 patients were happy to fill-in the FFT questionnaire. This exercise is now to be repeated more regularly until a tipping-point of patients is achieved.

One recent study assessing interventions in primary care is that there is currently insufficient good evidence to support many of the health improvement interventions being undertaken in general practice and more widely in primary care. (84) CCGs and Trusts are starting to re-think their assurance systems, aiming to balance the need for providing assurance and that targets are met via local action to improve things. Staff engagement in patient experience work is therefore more critical than ever. Patients are becoming more involved in patient experience work, from gathering data to providing insights and working with staff on data-led improvement work. The ways in which patient experience work is carried out is changing the nature of patient and public involvement work and the potential roles for patients.

There are more and more examples of patient experience teams working alongside healthcare professionals to improve services – in a spirit of co-production. But as MES report (85) increasing operational pressures are making it difficult to embed the work. These are some of the things they suggest can help:

•Persuading staff that patient experience is as valuable as – and can contribute to – clinical outcomes and safety

•Engaging staff by having them carry out patient experience work (gathering and using data)

•Involving patients and carers in dialogue about what data means and what can be done about it

Local leadership – from clinicians and/or senior managers – is essential, but as this case study demonstrates, leadership can come from patients too. The very notion of co-production is being defined by how feedback mechanisms are enabled and processed in ways that can be considered to be scientific.

Case Study 3

A patient is diagnosed with Non Alcoholic Fatty Liver Disease (NAFLD) and receives advice from their doctor to get more exercise e.g. by walking more. The patient uses the principles of IS to monitor what happens – using the data not just to comply with their doctor's wishes, but to help them understand what drives their mind/body system. The patient hopes that this knowledge can lead to better decision-making and maintained motivation.

The patient is not over-weight and so the diagnosis comes as a surprise. The symptoms are sporadic severe pain under the rib cage and understandably his immediate goal is relief from this. His GP advises that there are as yet no drugs that can be prescribed, so more exercise and better diet are the only ways forward.

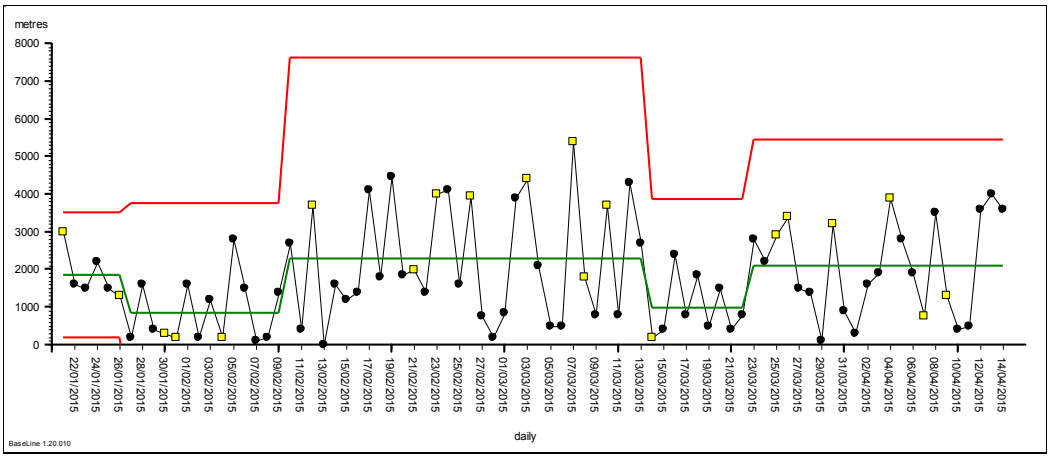
Because of the pain, there is no time to collect baseline data – which would have enabled the impact of any in intervention (before and after) to be demonstrated. Nevertheless, some baseline data is available because the patient already routinely collects daily data on BP, Heart Rate and Weight.

To keep the job of identifying cause & effect easier, the patient might wish to make one intervention – and even that can be tricky given that there is usually some kind of delay before an effect is observable. In this case the patient's GP has recommended more exercise (specifically more walking) as well as better diet (specifically consume less saturated fat) because the annual blood test result has indicated an abnormally high level (7.4 compared with 4.9

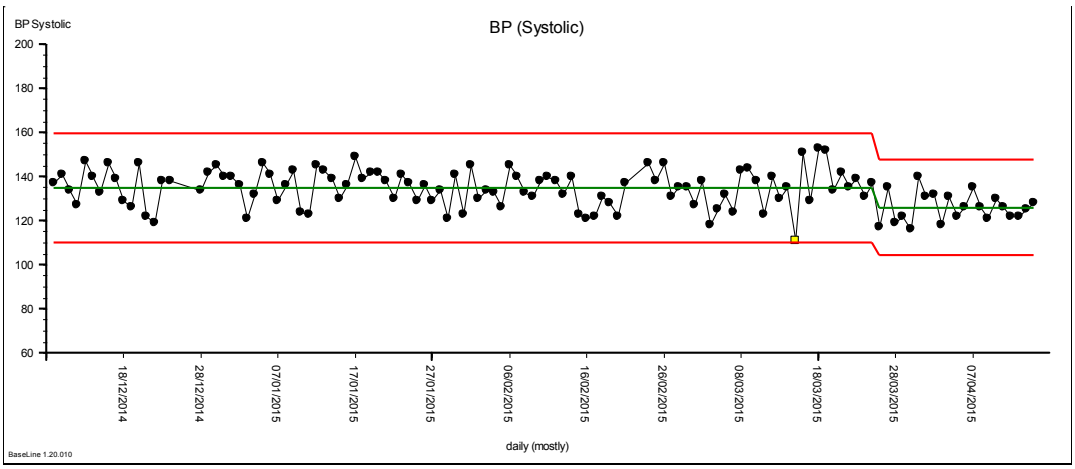
the previous year). The patient therefore decides to adopt both behavioural changes at the same time – much as the surgical team in the first case study did.

The patient also knows that he has been eating a lot of sugar during recent months and 2 weeks after the conversation with his GP the BBC happens to broadcast a programme called “The truth about sugar” featuring a young man also diagnosed with NAFLD. Inspired by this, our patient immediately decides to cut out all sweets and deserts.

He keeps a chart on the metres walked. It tells a story.. after an initial burst of activity for a few days – averaging almost 1850m a day, this was followed by two weeks at an average of only 850m per day. On studying the chart, the patient realises that something more radical would be needed, so he decides that for all local trips he should default to walking rather than driving. This leads to a sustained period (32 days) at an average of 2280m. The dip to just 983m for a week is because the patient undergoes a biopsy under general anaesthetic and takes fully a week to recover. After this he returns to a sustained average of 2km + per day. Whilst there was no baseline information prior to the phone call from the GP, the patient estimates that he had been walking around just 2Km per week.



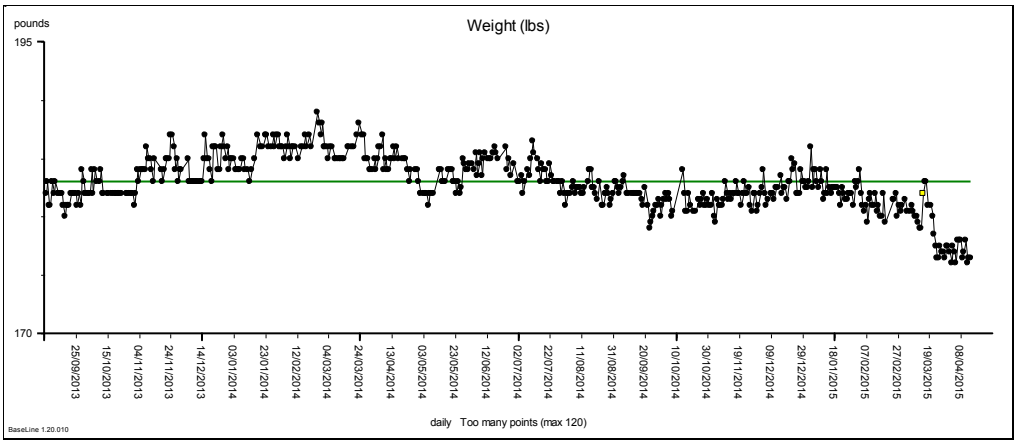
Segment	Start	Finish	Mean	Sigma	Count	LCL	UCL	Stable
1	21/01/2015	26/01/2015	1850	549.6	6	201.1	3499	Yes
2	27/01/2015	09/02/2015	850.0	968.4	14	-2055	3755	Yes
3	10/02/2015	13/03/2015	2280	1782	32	-3065	7625	Yes
4	14/03/2015	22/03/2015	983.3	964.1	9	-1909	3876	Yes
4	23/03/2015	14/04/2015	2107	1116	23	-1242	5455	Yes



Segment	Start	Finish	Mean	Sigma	Count	LCL	UCL	Stable

1	09/12/2014	25/03/2015	135.0	8.251	102	110.2	159.7	Yes
2	26/03/2015	15/04/2015	125.9	7.225	21	104.2	147.6	Yes
	15/03/2015	Biopsy under GA						

It is early days, but the new lifestyle seems to be having significant benefits. Blood Pressure, as evidenced by the systolic chart above, seems to have fallen back from a level that has been mildly hypertensive for the past 3 years. Weight to, as evidenced by the Run chart below, shows a consistent fall following the conversation with his GP in late January – apart from the week following the hospitalisation in mid-March. The patient's weight is now running at 6 lbs below the median. More importantly, the NAFLD symptoms are much reduced, and because the daily connection to evidence serves as a continual reminder there is now every chance of the new behaviour being maintained.



How are these 3 cases demonstrating good science?

The first case illustrates that a single surgeon and a single surgery team of around 10 individuals, in a locale that is well understood by them, can find a practical way to achieve more with just a single trial than the total combined efforts of a host of RCTs – and in doing so can build their knowledge of cause & effect in their system. The RCTs were initially vital in informing the surgical team ow best to assemble their care bundle. Of course, an RCT might have been deployed to test the care-bundle, but this would probably have been felt to be tricky to set-up without expert statistical support, and even if such support had been on hand the results would have taken much longer to materialise than the improvement science approach – and even then may well have been disputed.

Like many aspects of healthcare, surgical site infection is multi-factorial, and even if the surgical team were to assume that causal factors act independently of each other, any attempt to dissect out the essential causal mechanisms using a RCT to test, one at a time, each single factor would feel like it could take forever. And even then, the synergistic effects of multiple factors might easily remain undetected. Life is literally too short. This is an inherent limitation of purely column 3 approach, for when the effect of each factor is small the 'signal' can get lost in the 'noise' of the uncontrolled variation. The traditional approach to this problem is to 'power the study' by increasing the size of the study groups. Improvement science approaches this problem of signal-and-noise in a different way – it avoids comparing one 'system' with another at a single moment in time, and instead over time compares the same system with itself before and after the intervention is made. The problem of variation potentially swamping the outcome of interest is managed in a different way – in this case by adopting a policy of 100% Compliance with the Care Bundle and focussing only on *clean* hernia surgery.

The IS approach also has a number of other practical advantages not least that Time Series Analysis does not generate 'p-values' because there is no theoretical probability distribution assumed – indeed, for the vast majority of non-normal data, XmR Charts turn out to be remarkably robust (79). For healthcare practitioners this feature makes the column 2 approach much less intimidating than the column 3 approach.

Moreover, understanding for its own sake the mechanisms driving the system is only a *secondary* goal. The primary goal of the care bundle is a sustained, clinically significant and objective improvement in patient outcome. We believe that the approach of the surgical team is consistent with what Pawson and Tilley call (and supported by Berwick) the CMO approach: "Context + Mechanism → Outcome (CMO)" research design(20), in preference to an Observe → intervention X → Observe (O-X-O) design which ignores context. An intervention is being made, but designed by those involved and with full knowledge and allowance for the context.

The original case study is written-up in full as a narrative (24) because this is symbolic of the way that this type of approach works – it is highly context-dependent, and that context is as much to do with individuals involved, and the organisational culture in which they live, as it is to do with the methodologies considered and used, or the outcome as over time it unfolds. In the column 2 world view: narrative and evidence, imagined reality and objective reality, need to work as one whole. It is also written up as one person's story – as an insight into the difference that one clinician prepared to take the lead can make. This individual didn't wait to be engaged, they decided to self-empower and then build ownership with those holding key positions in the flow of work. To us Ben Goldacre too is an example of a clinician who is not slow to adopt a leadership role. Nowadays there are growing numbers of clinicians who if offered the chance are ready to grab the reigns, and locally-led Clinical Commissioning Groups are but one example.

It is the outcome that is all important – and by studying the variation over time building knowledge of what causes which outcomes. A key aspect to the philosophy of improvement science is that practitioners should.. "worry more about what works than being able to explain exactly why it works". The purpose of research is to draw valid conclusions from experimental observations. As it turns out, it is not difficult to equip clinicians, nor managers, to engage in this kind of science. Moreover, we don't have to wait until large numbers of individuals have been trained-up, for with timely coaching, even if only available virtually, much can be achieved by anyone prepared to be proactive and go for it.

Column 1 thinking makes many people nervous about small sample sizes, but in this particular case a single clinician, with their team, was able to make a massive difference to the outcomes of most importance to them. As long as the data collected can be viewed over time, Column 2 thinking means that sample size is often much less of an issue than it is under the Column 1 or Column 3 paradigm.

To summarise, the SSI case study illustrates a *before-and-after* study i.e. being explicit about one's hypothesis, choosing a metric that relates directly to that hypothesis, so it becomes possible to collect baseline data on the system as-is. Then, when the intervention (X) that is designed to test the hypothesis is made, its impact can be clearly viewed – over time – and talked about.

An exclusively Column 3 approach, over several years, had yielded largely inconclusive results – yet empowered by the practical possibilities inherent with the Column 2 approach, the two approaches have effectively been productively combined. If the RCT had been the only tool available, what new knowledge might eventually have been generated might well have been by comparison greatly inferior and would have been generated so slowly it would to the team have felt disheartening – and hence disempowering. The Column 2 approach has nevertheless been richly enhanced by the Column 3 studies that had already been conducted.

How are the FFT and Patient NAFLD case studies demonstrating good science?

The SSI case demonstrates the importance of enabling individuals and teams to take charge of their locale in its systemic context. Creating a genuinely empowering culture takes time, but is now an imperative for an NHS that has to make the most of every penny the tax payer can afford. The kind of empowerment we mean here is really "self-empowerment" – and by intrinsically motivated individuals and teams. All three case studies amply illustrate what this can mean – both for evidence-based enhancement of a system, and for patient outcomes.

How then do the three cases rate against the the other main recommendations in the Berwick Report that will lead to the NHS becoming a "Learning Organisation"?

1. ***Placing the quality of patient care, especially patient safety, above all other aims.***
2. ***Engaging, empowering, and hearing patients and carers, throughout the entire system, at all times.***
3. ***Fostering whole-heartedly the growth and development of all staff – including their ability to support and improve the processes in which they work.***
4. ***Embracing transparency unequivocally and everywhere, in the service of accountability, trust, and the growth of knowledge.***

The Berwick Report (2013)

We have taken Berwick's four main recommendations as headings and used them as a framework for how we see a culture of Improvement Science being nurtured. We have used this framework as a way of highlighting, albeit in simplistic binary way, how our 3 case studies serve to illustrate the key elements of such a culture in play ..

	Surgical Team reduces post-operative infection rate	Patient Groups enhance FFT feedback in GP Practices	A patient self-monitors to manage their LTC
1. Placing the quality of patient care, especially patient safety, above all other aims.			
(a) "Constancy of Purpose" flows consistently from the top, such that bottom-up frontline empowerment is likely			
(b) Responsibility for safety is clear and simple, cooperatively supported by each and all of the agencies involved	✓	✓	✓
(c) The story that evolves over time is locally context driven, data rich and evidence-based	✓	✓	✓
2. Engaging, empowering, and hearing patients & carers, throughout the entire system, at all times.			
(a) The wisdom of patients and carers is heard – not tokenistically, but as co-producers		✓	✓
(b) Patients leaders are developed with as much dedication as for professional staff			
3. Fostering whole-heartedly the growth and development of all staff – to support and improve the processes in which they work.			
(a) NHS staff get career-long help to learn, master and apply modern methods of Improvement Science.			
(b) Pride and joy is nurtured, not fear	✓	✓	
(c) Blame as a tool is abandoned – errors are handled differently from misconduct	✓	✓	
(d) The Model for Improvement, and a scientific method (PDSA) is used – to continually justify assumptions and to continually discover what works	✓	✓	✓
(e) "3 column" Improvement Science – with tools chosen pragmatically to provide evidence over time	✓	✓	✓
(f) A philosophy of "conceptualistic pragmatism" prevails, such that both logic and psychology may contribute to the accumulation of knowledge	✓	✓	✓
4. Embracing transparency unequivocally and everywhere, in the service of accountability, trust, and the growth of knowledge.			
(a) Transparency is seen as essential – for professionals and patients	✓	✓	
(b) Measurement connects people to real time reality – using Shewhart's theory of cause systems – to build Knowledge of how systems are changing over time	✓		✓
(c) Metrics are locally determined and operationally defined within a nationally set framework	✓	✓	✓
(d) Targets are applied with caution - and are never an end in themselves	✓		✓

All three case studies are examples of bottom-up systemic enhancement – and of the kind that Atul Gawande applauds in his Reith lecture. However, these kinds of initiative more often than not wither on the vine unless top-down support is eventually forthcoming. None of the case studies therefore earn a tick for "**constancy of purpose**"

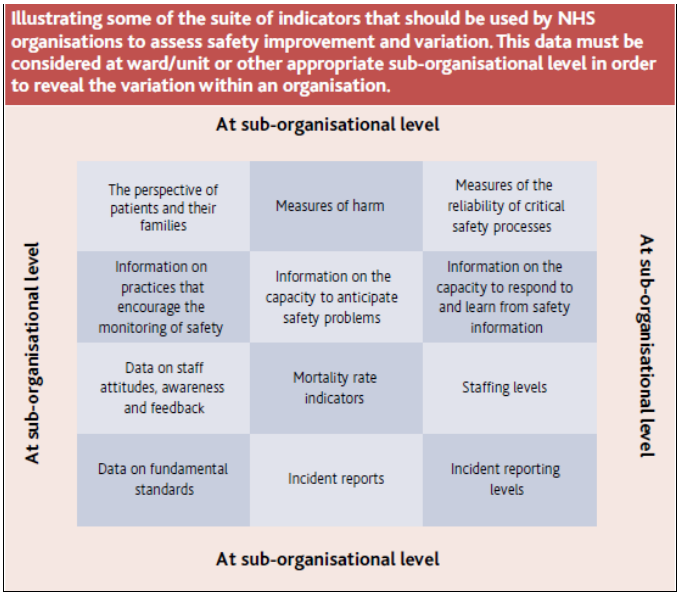
and this can only come from the top. In its absence, even a groundswell of goodwill can lead only to burnout and cynicism.

.. just because you have a roadmap does not mean anyone is going to follow it. There are barriers to over come .. In one healthcentre, staff may not wash their hands because they don't know it's important; in another, because they don't have sinks or running water in the delivery rooms; and in another they simply have not made it their habit and no one cares.

That last phrase I think is the critical one: if no one cares when someone takes the trouble to do things right, nothing changes. And the overwhelming message to the people who work at the frontline of care around the world is that no one notices excellence and no one cares. That is the biggest source of burnout and discouragement for healthcare workers everywhere.

Atul Gawande (11) 4th Reith Lecture (2014)

Measurement is an essential element in enabling safe and evidence-based healthcare, but metrics have to operate dynamically in their immediate context, and Berwick knows this. He also knows that people crave examples and in his report, even at the risk of being crass, he offers some ..



The Berwick Report (8)

But he knows only too well that although it will be the last thing to change, it's culture that trumps everything ..

"in the end, culture will trump rules, standards and control strategies every single time, and achieving a vastly safer NHS will depend far more on major cultural change than on a new regulatory regime".

Don Berwick

8. Good science requires a theory of knowledge

We will in this section define science, and show how IS stands upon the shoulders of several giants of science and philosophy. Here are two definitions that appear in the Wikipedia entry (1st April 2015) ..

> **Science** (from **Latin** *sciatica*, meaning "knowledge"[1]) is a systematic enterprise that builds and organizes **knowledge** *in the form of testable explanations and predictions about nature and the* ***universe***.[nb 1] ..
>
> *In an older and closely related meaning, "science" also refers to a* ***body of knowledge*** *itself, of the type that can be rationally explained and reliably applied.*

These two definitions are very different from each other. The way we have been using the term "science" so far in this paper best fits the first one, whereas the second definition might better suit a professional teacher or examiner working inside the education system. A school science teacher for example typically feels held to account for taking a relatively fixed body of knowledge and imparting it to their students – both as a series of presented facts in formal lessons, and via replicated experiments when space is available in the school labs. Such a professional might well feel more comfortable with the older second meaning?

The concept of "Knowledge" features throughout this paper, and the Latin root of the word science is knowledge, but not knowledge as a relatively fixed set of facts. Interestingly in the Wikipedia definition above, the word "knowledge" has a note against it ..

And that note invokes a third definition ..

> *"... **modern science is a discovery as well as an invention**. It was a discovery that nature generally acts regularly enough to be described by laws and even by **mathematics**; and required invention to devise the techniques, abstractions, apparatus, and organization for exhibiting the regularities and securing their law-like descriptions." — J. L. Heilbron (Heilbron & 2003 p.vii)*
>
> *"science". Merriam-Webster Online Dictionary. Merriam-Webster, Inc. Retrieved 2011-10-16. 3 a*

The fictional Dr Frankenstein might have preferred this definition, implying as it does that whilst there are natural laws to be discovered, with a little added imagination it will be possible to invent not only ways of showing how those laws work, but of bending to them to invent new entities, ones that nature could have invented for itself, but as yet hasn't. Many people today express fears about such science – GM foods, three-person babies, self-recreating computers, artificial intelligence, etc.

But the same note also takes us back full circle – happily for us as it's the one we are exalting – to the first definition ..

> *.. knowledge or* ***a system of knowledge covering general truths****, or the operation of general laws, especially as obtained and tested through* ***scientific method ..*** *such knowledge or such a system of knowledge concerned with the physical world and its phenomena.*
>
> *"science". Merriam-Webster Online Dictionary. Merriam-Webster, Inc. Retrieved 2011-10-16. 3 a:*

The "**scientific method"** is a body of techniques for investigating phenomena, acquiring new knowledge, or correcting and integrating previous knowledge. To be termed scientific, a method of inquiry is commonly based on empirical or measurable evidence subject to specific principles of reasoning. The *Oxford English Dictionary* defines the scientific method as *"a method or procedure that has characterized natural science since the 17th century,*

consisting in systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses."

In 1733 Voltaire invited Francis Bacon to France introducing him as the "father of the scientific method". Known also as the father of empiricism and experimental science, Bacon's works popularised inductive methodologies for scientific inquiry – planned procedures for investigating all things natural. In so doing he established a framework for science, much of which influences methodology today. Interestingly Bacon's researches included health and the prolongation of life itself – perhaps the first epidemiologist since the Greeks?

Since Bacon favored an inductive method of thought and research, one that proceeds carefully from evidence to conclusion, it made sense for him to observe the world by noting both sides of any controversial issue – the key being to examine the evidence before reaching his own conclusion. Bacon wrote both for the seeker of power and the man about town, and wanted to train that man to look beyond the obvious, to coolly and deliberately assess his position before advancing.

Harari (30) reminds us however that .. *"Mere observations are not knowledge. In order to understand the universe, we need to connect observations into comprehensive theories. Earlier traditions formulated their theories in terms of stories. Modern science uses mathematics"* .. To Harari, the new era of modern science is built upon our willingness to admit ignorance. However, whilst hugely expanding Sapiens capacity to understand how the world works, it presents us with a serious problem with which most of our ancestors didn't need to contend – the new assumption that we do not know everything, that current knowledge is tentative, that our shared myths whilst enabling millions of strangers to get along with each other, are not real, and that knowledge is not fixed, nor immutable. It is the goal of Improvement Science to enable those closest to the context to be able to uncover the way their world works so that armed with this knowledge they may enhance that world – or at the very least make it a safe one.

The phrase "the science of improvement" was the first used in 1996 in *The Improvement Guide* (Langely et al). It was used as a way of building upon Deming's System of Profound Knowledge (refer back to page x) and to stress two critical ideas: First, the idea that all improvement comes from developing, testing, and implementing changes – the role of measurement being to create feedback (learning loops) to gauge the impact of these changes over time as conditions vary in the environment, and .. Second, the idea that the subject matter expert plays the lead role in developing changes and establishing the conditions for testing – that together increase the degree of belief that the changes will lead to improvement.

As we referred earlier, Perla et al (54) have more recently offered seven more specific propositions that underpin the science of improvement, which ..

1. Is grounded in testing and learning cycles.
2. Embraces a combination of psychology and logic.
3. Has a philosophical foundation is conceptualistic pragmatism.
4. Employs Shewhart's theory of cause systems.
5. Requires the use of operational definitions.
6. Considers the context of *both* justification *and* discovery.
7. Is informed by Systems theory.

As a set of tools that is underpinned by new thinking, we believe the application in the NHS of Improvement Science to be overdue, for it has been used for decades in other sectors – many contending with just as complex a set of systems. In recent years the NHS has seen many "quality-first" initiatives – but for the most part these were destined to wither on the vine, starved of consistent top-down support. The Modernisation Agency and the NHS Institute for example were set up, then later disbanded – somewhat ironically being too slow to learn that their impact was insufficient. The kind of skills development now recommended by Berwick therefore has been sporadic and barely connected to either central or local strategic thinking. One reason for this is the lack of band-width on the Myers-

Briggs *S* ↔ *N* continuum, leading organisationally to a binary world view whereby individuals possessing an N preference get disproportionately promoted into senior positions – disconnected from those with an S preference who remain close to the coal face. The well documented top-down nature of its moribund siloed bureaucracy, making it nigh impossible for the S and N preferences to prevail in a sufficiently integrated way.

Interestingly, *in 1910,* around the time that Carl Jung was incubating his thoughts on psychological type, research which eventually evolved into the two perceiving functions: **Sensation and Intuition** – which as the *S* ↔ *N* continuum later became core to the Myers-Briggs type indicator – Clarence Irving Lewis, a major influence upon Deming, was writing his undergraduate thesis "*The Place of Intuition in Knowledge"*– in which he aimed to uncover an integration between realism and idealism.

At Harvard, Lewis had studied metaphysics with Josiah Royce, and simultaneously studied Kant and epistemology with R.B. Perry, and he felt compelled to forge a neo-Kantian middle road. Lewis argued that the justification of knowledge requires both the mind's legislative or constructive activity, and "givenness".. his word for intuition.

Eventually, in 1929 in "*Mind and the World Order",* Lewis's seminal work, he advanced a position he called "**conceptualistic pragmatism**" according to which empirical knowledge is dependent upon three things: a sensed 'given'; a constructive activity of the mind; and a set of *a priori* concepts which the individual (agent) brings to their interpretation of what's observed – the product of the individual's social heritage and of particular cognitive interests – their knowledge therefore being "pragmatically *a priori"* as *opposed to a priori* in the sense of it having been given absolutely. The terms ***a priori*** ("from the earlier") and *a posteriori* ("from the later") are used in **philosophy** and specifically in **epistemology** to distinguish two types of knowledge justification: *A priori* knowledge is independent of **experience**; and *A posteriori* knowledge which is dependent on **experience** or **empirical evidence**.

An unduly narrow *S* ↔ *N* "**bandwidth**" in managerial leaders is further compounded when those individuals preferring to analyze situations using their trusted logic (MBTI T-types) think there is but one way of processing what they are observing – and as Lewis shows there are actually many alternative systems of logic, each in its own way apparently self-evident. Equally, the other half of the world (Myers-Briggs F-types) are likely to be constrained by naturally defaulting to the processing of their observations according to their values – such that what is seen and interpreted as "right" is determined more by personal scripting than by logic. Bacon's concept of empiricism – that knowledge comes primarily from sensory experience, from evidence, from testable and falsifiable hypotheses – more than from innate ideas, tradition or revelation – pragmatically therefore becomes difficult. Lewis's notion of "conceptual pragmatism" handles this difficulty by assuming empirical knowledge is dependent upon *(1)* a sensed given (2) the constructive activity of a mind, and (3) a set of *a priori* concepts which the agent brings to their perception and interpretation of events.

Avoiding "Tampering"

In his own attempt to apply this thinking to how a managerial leader best approaches their work, Deming (16) created a concept he called "Tampering" which is meant to convey how inappropriate actions often adversely affect organisational processes or systems. Deming wanted people working in and on systems to understand that they are prone to over intervening under an assumption that applying their "common sense" is the best guide as to what to do.

Tampering in other words is managerial leadership by reaction, intervening without sufficient understanding of the natural variance that occurs in a system or process. It is a common fault of inexperienced and untrained managerial leaders to react to common cause variance – often in a heavy handed fashion. Their actions, albeit with good intent, result in tampering with the system leading to a consequent amplification of the variation, and making "special cause" problem more likely – oblivious that the problem has been self-induced. Tampering in short is the inadvertent amplification of "common cause" variation, causing the de-stabilization an otherwise controlled system.

Avoiding tampering depends upon the ability to be able to maintain a system in its optimal state by accurately sensing it in real time (S) together with the intuitive sense (N) to know when one's own "common sense" behaviour

is detrimental to it. Knowing and internalising the concept of "tampering" in this Deming sense means possessing a heightened awareness of when yourself or others are falling into a lazy "common sense" way of thinking. Here are a few everyday phrases and thoughts that may indicate tampering, indeed bad science, might be in play ..

- ***Trust me, I know this will work;***
- ***I am in charge, we will do it this way;***
- ***It's really just common sense!***
- ***It's my sense, and my sense is right, everyone knows it's for the best;***
- ***Do as I say. I'm the boss, I know best. If you can't show me that things have improved by the time we next meet, we'll all know you didn't do as I told you;***
- ***We don't need to investigate, the solution is obvious;***
- ***We need action not study;***
- ***It works in X company so it will work here;***
- ***I'm the most experienced person around here, when everyone else is as experienced, better decisions will be the norm;***
- ***For me, it simply comes naturally .. I'm really just brilliantly intuitive;***
- ***It's only what most everyone else would do?***
- ***In fact it's what most everyone actually does .. isn't it?***
- ***I'm the most normal person I know. The rest of the world is really nonsensical. Don't expect me to have their kind of sense;***
- ***I realise it may not be common practice, but wouldn't the world be a great place if it were;***

Deming taught that Knowledge is built upon explicit theory, on making the kinds of thinking illustrated in the list above, as transparent as possible. To him "theory is a window into the world – it leads to prediction – and without prediction, experience teach nothing. To copy an example of success, without understanding it with the aid of theory, all too easily leads to disaster. In his words "any rational plan, however simple, is prediction concerning conditions, behaviour, and performance of people, procedures, equipment, or materials" (17). He gave us, as a map – his "*system of profound knowledge*" – to help us address the task of building a system that can be sustainably maintained, continually improved – and if required redesigned.

Theory therefore has to be pragmatically grounded so that awareness of cause and effect may become sufficient for wise systemic choice-making to become possible. Deming's recommended methodology is to identify and operationally define a system's vital signs so they become monitorable over time – which allows previously invisible systemic change to be spotted – and wisely responded to – in near to real time. The Plan in his simplification of the scientific method (P-D-S-A) is meant to embody this grounded pragmatic theory such that it may be iteratively tested in a way that generates learning. This is the essence of improvement science, and of good science.

SIDE BAR **: Our own working definition of science:**

A process by which Sapiens try to increase their awareness/ knowledge of what's real.

It uses history to formulate theories about how the future is likely to be .. so that the future can then be experienced as it occurs in order to assess the gap between my/our prediction and what is observed as actual – all the time recognising and accepting that what is observed is a matter of perception and that therefore reality can never be more than merely partially observed.

In trying to get objectively closer to what's real I/we may aim to reduce the subjective elements in what I/we perceive, but because both the observer and the context are each unique – and measurements are variable over time – the process of observing and measuring any particular context is inevitably a function largely of the context itself, and of the psychology of the individuals interacting with it. Moreover, because the nature of what's predicted depends upon the quality of the question(s) posed, and hence upon the level of awareness thus far achieved, it is in practice virtually impossible to inquire into what isn't yet conceivable i.e. what one doesn't yet know one doesn't yet know.

The boundaries of Knowledge therefore expand only like a slow dance, with the occasional lurch. Such is the nature of continual improvement, but over time if efforts are effective in storing it knowledge accumulates.

We started this section by looking at three definitions of science. For many people, science is merely a body of facts that they struggled to get their head around when at school, or worse science is something to be wary of – an intimidating concept that may a hold dark prospectus for Sapiens.

For us science is about seeking an ever greater understanding of the way life works, so that we may make better informed choices – a way of thinking that when properly conducted can in every sense be life enhancing. As Atul Gawande says, through knowledge we find hope.

"Creating the systems required for the fundamentally transformed needs of human existence is .. a tall order. We have no magic solutions, but I do think we have a path – and that is to invest in a science of exploration and discovery of how our systems succeed and fail, just as we have invested in a science of how our bodies' systems do. Because when we pull back the curtain, we find not only knowledge, we also find hope"

Atul Gawande (2014 Reith Lecture)

9. Telling a knowledge-full story

> ***"Narrative is radical, creating us at the very moment it is being created."***
>
> **Toni Morrison – novelist and winner of the *Nobel Prize* in Literature (1993)**

Storytelling is one of the oldest and most pervasive of human activities and according to Harari (30) was borne of Sapiens' ability to gossip, and to create trust in social relationships. Stories help us make sense of experiences our own and those of others. It is through stories that we learn to be part of a family, a group, a society, an organisation or a culture. Stories also convey tacit knowledge – the things we know but don't necessarily know we know.

For many the telling of our personal stories is a primary way of dwelling upon and making sense of our lives, be that through anecdote or by the capturing of memorable events in a diary, or perhaps talking through events with a friend as a way of coping with a personal crisis. Such a process is a way to make sense of our lives – and an essential part of being human – it creates meaning, for as Socrates famously advised: "the unexamined life is not worth living".

Story telling may be seen therefore as a kind of therapy, as a way of restoring psychological well-being, for many people however it is not yet viewed as a tool with the potential to help with describing a work process or a healthcare system that is failing to deliver what is needed by the population it serves. What's more, as Donald Schön cautions (55) .. *"Stories are products of reflection, but we don't usually hold on to them long enough to make them objects of reflection in their own right."* – in other words we need some commonly understood way of capturing our stories, as well as recording and sharing them.

Nowadays one might opt for writing poetry or songs, or taking photos, or making quilts, or gathering as a bee to chat and knit scarves. A thousand generations ago people passed on their stories by drawing pictures on the walls of the caves in which they dwelt. Before that people just exchanged gossip – maybe building a narrative to fit their survival needs – probably embellishing any aspect not immediately verifiable and feeling free to ignore any facts that might cause the story to be ruined. Indeed sacred texts are mostly collections of stories – think of them as "his-stories".

For the next phase of Sapiens evolution to come about however we believe that what's needed is a data-informed view of history so that a more objective diagnostic process may then ensue. The future too demands a level of objectification because, powerful though stories are for painting pictures of what might yet happen, by definition no observable data is yet available, and as everyone knows some aspects of the future are more predictable than others. In short we will need better ways of collaboratively engaging in both diagnosis and prognosis.

According to Harari once Sapiens discovered how to gossip, how to create and convey an imagined reality, it became possible to cohere in ever larger groups. This capability was so potent that Sapiens were able see off all other animal species including the apes, as well as all other human species. This happened because our ancestors discovered how to tell stories people could believe – learning how to make compelling myths. That was 70 millennia ago, but merely half a millennium ago another revolution occurred as modern science started to challenge myth-making and superstition.

We believe the time is ripe for a fuller integration of these two ways of making sense – the objective reality of science, informing the subjective reality of storytelling. We need a way to embrace our natural need for creating narrative whilst simultaneously embracing what scientific observation tells us is "*actually*" occurring. As Harari says we now have to learn how to tell better stories – stories that are enriched by knowledge – for Sapiens this is now of vital importance not only because threats like climate change now challenge their very existence, but because increasingly people expect the narrative to be fact-based. Politicians in particular are nowadays felt to be less and

less believable, and their statistics get especially scrutinised. Websites like “Ask for Evidence” and “Sense about Science” are proving popular, along with shows like Radio 4’s “More or Less” which seeks to delve into the numbers behind topical news items. The Royal Statistical Society has even felt the need to launch an online way for UK citizens to email their local General Election candidate to ask their local candidates to sign up to take some basic statistical training (which the Society will provide in Parliament) should they get elected on 7 May. And as we write this, the last King of England to die in battle is being reinterred in Leicester Cathedral.. some 530 years after the event we are now being confronted with the “facts” that the hitherto popular account of his life and reign was largely a story that was spun by officialdom – subsequently with powerfully lyrical support from William Shakespeare.

As we have seen, central to the Jungian-based Myers Briggs personality test (MBTI) is the distinction between sensing (S) and intuiting (N). Those having a strong “S” preference getting irritated whenever fiction predominates over what they would see as concrete, fact-based conversation that deals with the here and now. Those with a strong N preference are conversely interested more in abstract ideas, theories and future possibilities. The tension between Harari’s two realities is vitally echoed by these two poles. Individuals who can engage equally with, and integrate, these two ways of seeing reality are in short supply, yet in the face of societal and organisational challenges requiring potent leadership they can possess a kind of super-charged capability – see the side bar Bob Dylan/ Joan Baez example.

SIDE BAR **: The power of "Bandwidth"**

Taking a famous historical example, in 1962 when first meeting Bob Dylan the singer Joan Baez was as enchanted by his poetic lyrics as she was dumfounded by his apparent disinterest in his own genius for bringing deep meaning to her and to many of her contemporaries. He was an enigma to her and the rest of her generation. David Keirsey explains (38) that because of his "S" preference, he was actually very connected to what was really going on whilst uncannily connected to those with an "N" preference, many of whom, driven by the new ideas of the age, wanted to change the world.

Baez, and many in the media, tried to project their own agendas onto Dylan, but in being strong enough to resist, the more that words were put in to his mouth the more he refused to adopt the mantle that others would assign him. To him their naivety and hypocrisy, their pretentiousness, the dearth of authenticity – were all too transparent, and as an artist his simple goal was to be able to capture and mirror it .. "The answer my friend is blowin' in the wind" .. and "You don't need a weather man to know which way the wind blows".

MBTI theory of course is more complicated than this simple S--N dichotomy – see David Keirsey's books and his website (37) – it is nevertheless clear that for someone so young, Bob Dylan in the 60s somehow possessed a wide ***bandwidth*** *on the* ***S*** ↔ ***N*** *continuum, and this gave him huge charismatic power. Keirsey suggests that Dylan (as an ISFP – "perhaps the most misunderstood of the 16 MBTI types"), whilst appearing to shrug his shoulders and walk away from what others were seeing as his responsibility to lead, was actually craving social impact – to be a strong enough presence to be able to affect the course of events.*

In living his life along the full length of the S--N continuum Dylan possessed unusual power. And many, inspired by him, went on to become societal leaders of one kind or another – frequently relying on Dylan's lyrics as a kind of touchstone. In Julian's experience as executive coach he sees in some individuals a super-charged potential to bring forth historically needed change (38). Seemingly, the ***S*** ↔ ***N*** *"****bandwidth****" expands naturally over time, and then more quickly after the personal shift to "post-conventionality" (62). In time sufficient numbers of individuals evolve this capability – collectively perhaps making historically significant societal shifts more likely. Julian's observation is that just such a bifurcation point maybe fast approaching, and one manifestation of this is an awakening to the need for good science conducted by empowered people – those who can collectively agree about what has been happening, what is happening in this moment, and what is most likely to happen if (a) nothing changes, or if (b) together we can instigate and lead the warranted change.*

*Note that our concept of "****bandwidth****" can be inappropriate for conventional individuals – typically those in pre-midlife – for as Jung pointed out (62) development is best facilitated by focussing on one's natural preferences, by accentuating the poles. However, what works well in the first half of life does not necessarily serve in the second half. Such "post-conventional development" (35) typically comes – unless arrested in some way – via a process of what Jung called self-realisation – a new found quest for completeness. Suddenly, being able to draw upon a wider range of parts enables a new level of flexibility – the individual can then choose from a wider bandwidth to express themselves according to the situation in which they find themselves.*

So what strategies can be adopted by individuals in their first half of life? The main strategy for conventional beings – around 15% of people in organisations (62) – is to learn how to appreciate the differences in others, and how to collaborate – for others have the parts you personally are unaware of. In short, teamwork and cooperation works.

It is important not to see the S & N poles as default positions to be traded-off against each other. A few highly evolved people (see side bar) are naturally able to integrate these two poles – they might be said to possess high bandwidth – but the rest of us need assistance. Any narrative that emerges will be either woven, without good data, and by people who are disconnected from what's really happening, or will merely remain hidden – waiting for analysis by people who may have assembled the facts (data and information) but are not able to build a narrative that coherently ties the facts together. So, how can the effects of polarisation be mitigated?

The ***T*** ↔ ***F*** poles too require integration – some way of locating wisdom, for as car Jung said ***"where wisdom reigns, there is no conflict between thinking and feeling"***.

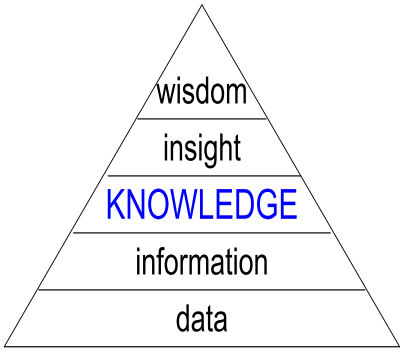
An approach and methodology is needed that can liberate politicians, managers, healthcare professionals, indeed everyone working in the public sector, from this binary world view – and in a way that generates knowledge.

Harari points out that information isn't the same as knowledge, and that it usually comes in the form of statements, typically presented as "facts" – but not ones that are necessarily easily tested. He reminds us that .. "mere observation however is not knowledge. In order to understand the universe we need to connect observations into comprehensive theories" (30). Knowledge, insight, even wisdom can be made accessible when we make our theories explicit enough to be testable.

> **"Wisdom is not a product of schooling, but of the lifelong attempt to acquire it" .. Albert Einstein**

Good knowledge and understanding about a system requires much more than collecting only information about it, it may therefore be helpful to think of knowledge creation as part of a hierarchy as in this diagram (1).

This was Deming's thinking too when he developed P-D-S-A thinking to enable data and information to be converted into knowledge via an approach that is simultaneously pragmatic and conceptual.



Since publishing the SoPK, "**knowledge management**" (KM) has separately almost become an industry of its own. Here's seven principles that might provide some guidance to those seeking to integrate Deming's ideas with KM:

1. Knowledge resides within processes (a kind of muscle-memory) as well as within individuals, and at least half of it is "tacit", rarely being made explicit.
2. People acquire knowledge from established organizational routines, the entirety of which is usually impossible for any one person to know, hence the importance of effective teamwork, and partner collaboration.
3. Knowledge is only ever volunteered it can never be conscripted. It can however be made more explicit if the right processes and good methods are attended to.
4. Individuals (and Teams) always know more than they can tell, and can always tell more than they can write down.
5. Individuals (and Teams) only know what they need to know, when they need to know it. If there's an organizational need for codification, then that's the moment to ask them to make it explicit.
6. Knowledge is created from data which has been converted into information. It can be created by individuals in conversation with themselves, or with others. It requires clear theory (a good question), one that enables the individual (or Team) to make a prediction. This is central to the philosophy behind "continual improvement" (PDSA)

7. It's good dialogue that leads to profound insight and understanding. As hidden assumptions and mental maps get exposed, they can be changed or removed. To some this is the source of wisdom.

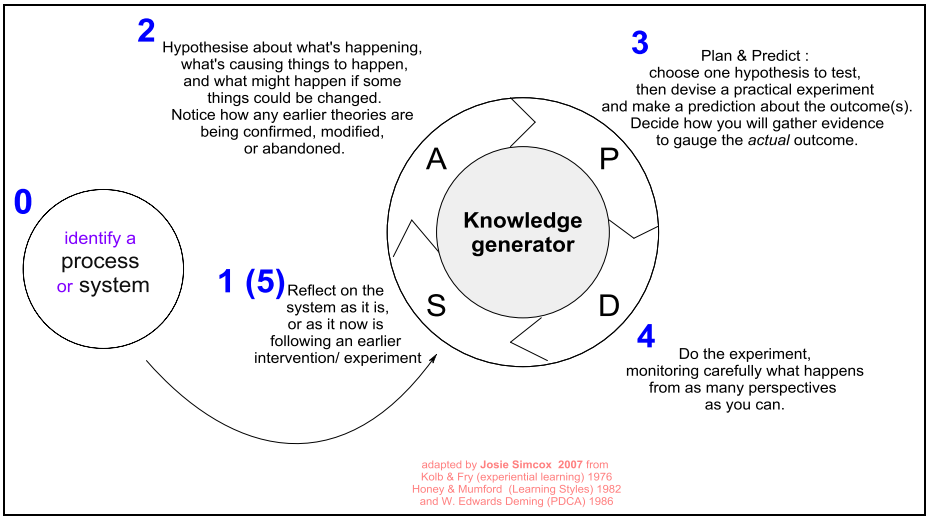
In the post-modern world we more than ever need tools that effectively integrate constructivist and modernist philosophies. "Constructivists" argue that the universe is continually being created by each of us as actors, and that each of us with our unique set of cultural filters can never see that universe exactly as it is. However we come from the perspective that the universe can be known (or at least closely approximated) and continually discovered via scientific experimentation. The trouble with holding only the constructivist position is that "**confirmation bias**" tends to be accepted as a given and not as something that can be readily challenged whereas the two of us choose to believe that Constructivism and Scientific Modernism not only *can* coexist, but when combined can give us a more complete view of the world – see our comments on conceptualistic pragmatism.

Information Departments in NHS Trusts have for years been generating ever more quantities of information, yet only a small proportion of it ever gets converted into knowledge because **commissioner-led systemic inquiry** is at best sporadic. Instead, the focus is all too often on reacting to the latest performance ranking handed down from on high – or finding information that supports the case a politician has already decided they need to make.

What commissioners really need is an answer to the question WHAT INTERVENTION SHOULD WE MAKE TO ACHIEVE IMPROVED PERFORMANCE – not as dictated by political imperative, but according to patient/ citizen outcomes as perceived by those closest to them. League Table rankings provide little more than a snap shot of the system at just one point in time. As Deming warned: ***"One may learn a lot about ice, yet know little about water."*** Performance can only really be understood when viewed by those closest to the context, and if seen as something that is varying naturally and observable over time.

Deming died in 1993 but not before he bequeathed us his System of Profound Knowledge (SoPK) in order to frame a methodology for connecting people to the reality of the system in which they work, and as a platform for understanding its behaviour and securing its sustained enhancement over time. Those who work in the system can bring about such enhancement by collaborating with others closest to the context as it flows – with the support of a managerial leader who in knowing how to apply the methodology, can assess the system's current capability and choose whether to monitor it, maintain it, continually improve it, innovate or redesign it.

At the heart of this methodology is the notion that knowledge builds out of continually asking the question "**how will I know?**" – which both embodies prediction and calls forth a scientific method such as P-D-S-A. The following diagram summarises such this method in action.



As Berwick recognises, "*stories are necessary to gain knowledge*" .. so we would like to make this more explicit in the methodology. What then might this method look like if in order to represent both ***S*** ↔ ***N*** poles we additionally incorporate the need for an informed narrative?

Creating a data informed story

In the Case Studies (section 6) we see how being able to view data pictorially as it changes over time and within its natural limits – whilst simultaneously being able to relate systemic change with a particular context – provides opportunities for powerful insight. But observations need to be noted, and by each collaborator, preferably in real time – otherwise insight is lost and memory fades, leaving the mind prone merely to make up its own un-evidenced story.

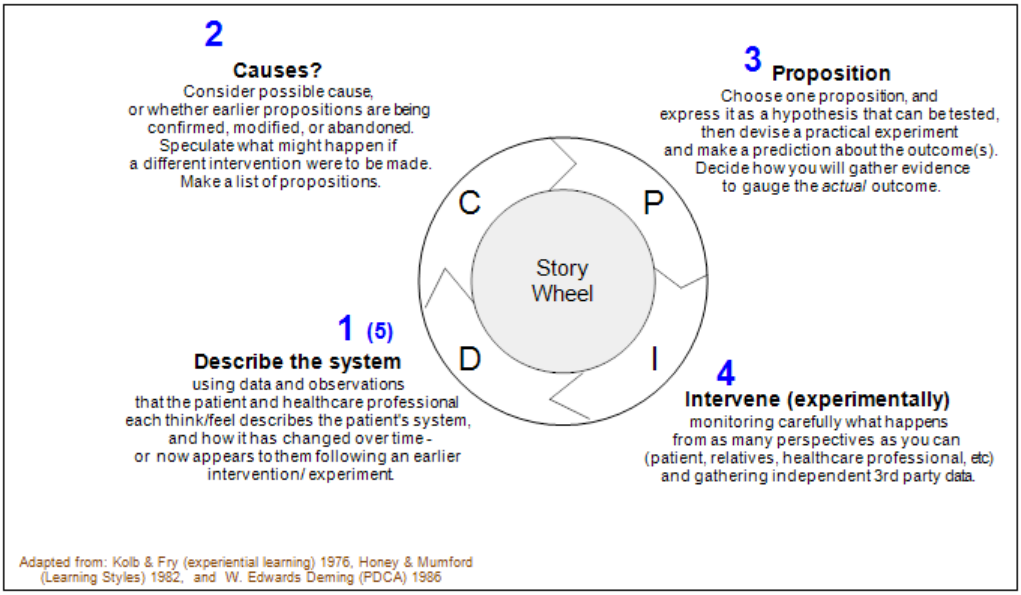
How then might noted observations be best structured – not just so that collaborators can make good sense of them but also to bring some science to what might be causing outcomes to occur, and so that with carefully designed experimentation those outcomes may later be intent-fully enhanced?

Stories don't always flow easily, not just because people fear the consequences of admitting mistakes, but also because in the flush of success people tend to forget what they learned along the way. The knowledge-full story cannot therefore be forced, it has to be teased out, and a little structure helps.

Here is a list of things that in our experience need to be alluded to when note-taking, whether by the patient or the physician – or both ..

1. **Context** – pertinent details about the immediate environment – especially about how it flows – illustrated with a map with input and ownership from every collaborator.
2. **Presenting symptoms** – linked to the steps and stages of the map, current issues and historical systemics.
3. **Agreed Aim** – ideal outcomes defined for each specific stakeholder.
4. **Baseline indicator(s) of events** – if possible numerical and preferably collectable over time – answering the question ***"how will I know the effects of any intervention we make?"***
5. **Annotation** – by each collaborator in real time.
6. **Interventions** – each described with timings – each expressed as a proposition and a prediction.
7. **Reference information** e.g. clinical guidelines.

Most importantly, the notes should enable the emerging story to be told as a series of cycles – cycles of co-inquiry. For this reason we are calling our model "**The Story Wheel**" which is designed to echo the P-D-S-A mnemonic. The wheel metaphor is important because like its sister PDSA there is no end to the number of cycles – it never stops rolling.



Stage 1 (and stage 5) : Describe the system

Using data and observations that collaborators think/feel best describes the patient's system, and how it has changed over time – or now appears to them to be changing following an earlier intervention/ experiment.

Stage 2 : Causes?

Consider possible cause(s). Or whether earlier propositions are being confirmed, modified, or abandoned. Speculate what might happen if a different intervention were to be made.

Make a list of propositions.

Stage 3 : Proposition

Choose one proposition, and express it as a hypothesis that can be tested – then devise a practical experiment and make a prediction about the outcome(s). Decide how you will gather the evidence to gauge the actual outcome.

Stage 4 : Intervene (experimentally)

.. whilst carefully monitoring what happens from as many stakeholder perspectives as you can, and gathering any independent 3rd party data. The BaseLine® tool helps keep track of any numerical data that is collected over time.

After stage 4, it is back to stage 1 again (stage 5) – to describe the system as it now appears following the earlier intervention/ experiment.
In this way, all collaborators may work together, sharing and making sense of the story as it unfolds – exploring causes before setting-up and monitoring each agreed intervention. All these aspects are picked-up as the Story Wheel rolls along.

'How can I help people if I don't know the right stories to tell them?'

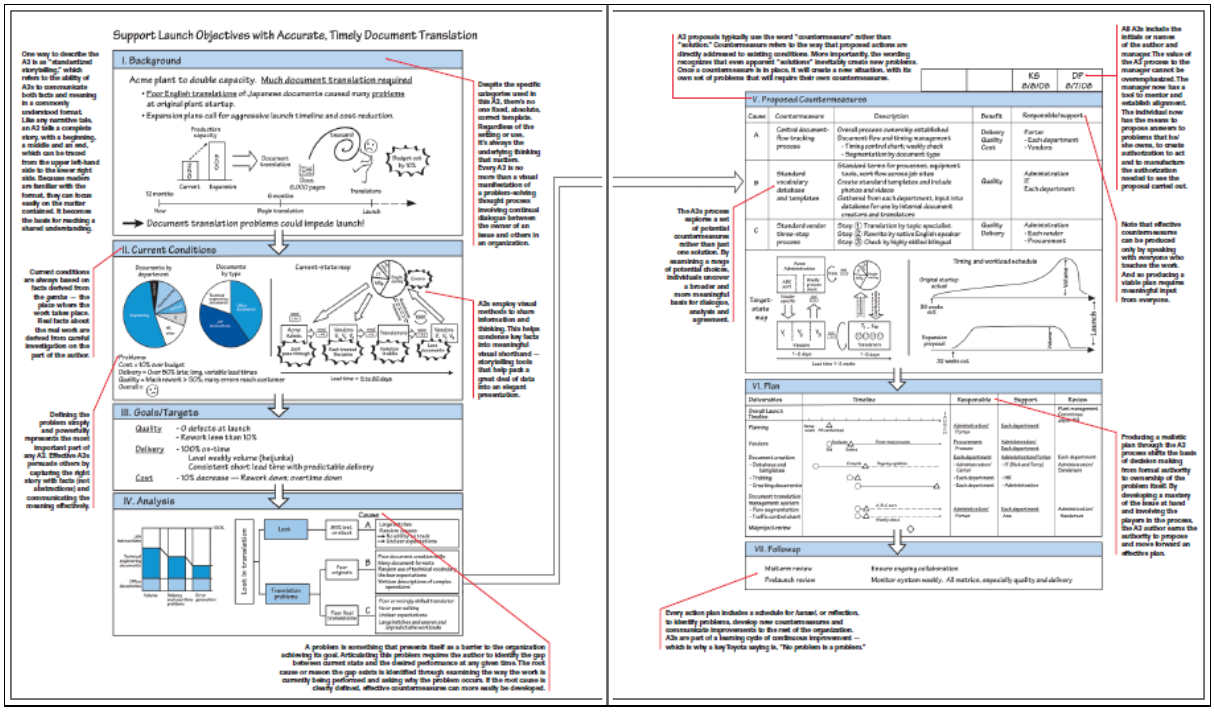
Johnny Cash in "Walk the Line" .. the 2005 film

Standardised story-telling

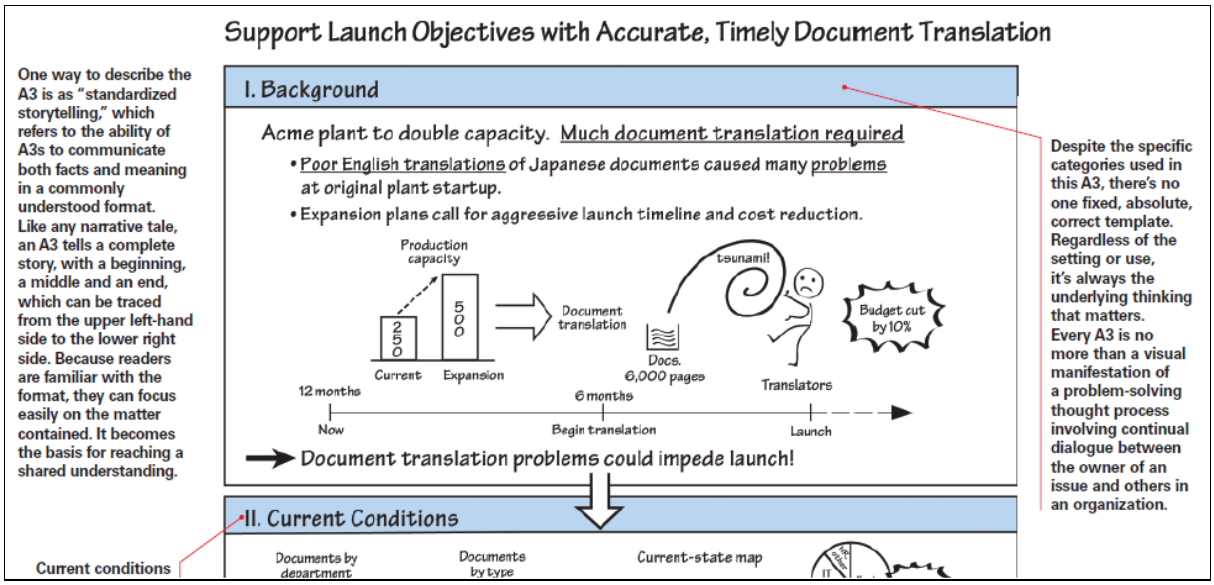
Organisationally, and individually, it makes a great difference if a small number of tools can become widely accepted and used as standard, and in our experience we know that storytelling requires such a standard, for narrative knowledge is what we each use to make meaning via cognitive, symbolic, and affective means – knowledge which provides a rich, resonant comprehension of a singular person's situation as it unfolds in time, whether captured as a

novel, a newspaper story, a movie, a scripture, a courtroom transcription, a battlefield account, or a documented illness.

The Toyota Production System (TPS) for example is famous for both its effectiveness and efficiency – achieved and archived over several decades through its culture of continual improvement. Their key tool for communicating and sharing learning is the "A3 Report" – its name is derived simply from its size which at twice A4 was a good practical fit for a fax machine. Here is one example ..



(64)



(64)

As the notes in margin suggest "because readers are familiar with the format they can focus easily on the matter contained". In the NHS as a system we need our own equivalent standard – one that institutionalises the use of PDSA in narrative form, and one that embeds the notion that understanding the aim of the system and the causes of the variation being experienced in the achievement of that aim can come only from genuine inquiry.

Deming was fond of reminding us that "***Life is variation*** and there will always be variation: between people, in outcome, in service, and in product. The key question is ***what is the variation trying to tell us about the system, and about the people that work in it?"*** This is the kind of question that initiates real knowledge-generating inquiry.

After choosing a system to work upon, the first step is to study it. Mapping it in a variety of ways is always helpful, but unless you have a way of monitoring its outcomes over time, science is rendered impotent.

There is variation inherent within every system and Deming's recommendation was that this be used to listen in to the system in real time – much like a doctor listens in to their patient via a stethoscope, but in this case continually over time.

Over the years the world of improvement has experienced several schisms, e.g. pure process improvement vs business process reengineering, and one of the elegant things about the 6M Design model is that in working from right-to-left as well as left-to-right it successfully unifies both improvement and innovation. The six steps are also an excellent structure within which to capture a narrative: Map, Measure, Model, Modify, Monitor, and Maintain.

It is all too easy to fool ourselves. A standardised tool will help everyone to protect themselves from the made-up story – the story that fails to achieve the full bandwidth needed between the ***S*** ↔ ***N*** poles and drifts too far towards the N pole. Individuals make their way through life by adopting their own rules-of-thumb or heuristics, and these become the bedrock of what we each refer to as "common sense" – the trouble is that we are not naturally aware that these heuristics are sometimes more personalised than common. And even those that are apparently common, universally true, may amount merely to collective delusion (see Deming's concept of "Tampering" page 53).

10. An invitation to Ben Goldacre, and other activists

We of course are sending Ben Goldacre a copy of this paper and we hope he can respond positively to the feedback and to the proposals. We want explicitly to recognise the significant contribution that his T-L-A paper makes, as far as it goes. Ben has worked hard for more than a decade to earn himself a position of influence .. including over 0.4 million followers on Twitter (see below) .. so we are sure it is being widely read – and we hope ours will sit alongside it as a sister paper? In the media there are too few people standing up for science in health, and Ben's writing continually informs the public and sometimes even prompts changes in policy. Unashamedly, we aim to ride on the back of that influence and the goodwill he garners for more science and for more evidence-based thinking, so that improvement science may reach the wider audience it now deserves.

In a world that gets clouded by conspiracy theories we personally love the way that, especially in his books, Goldacre dispassionately accumulates evidence. Like us he wants people to awaken to the real world not only an imagined one, and this is our own quest too. In a field where 99% of people feel intimidated, his indignant passion to make a radical difference comes through in spades. More recently via the AllTrials campaign – advocating open science practices in clinical research – he has been especially impactful. He started it in 2013 after he experienced people in industry too easily brushing aside the evidence being presented. The campaign calls for all trials to be registered, together with the full reporting of both methods and results – and this means ALL trials, past present and future, as well as on all those concerning currently prescribed treatments – no exceptions, no loopholes.

Last year the World Health Organization, the European Commission and the US, UK and Canadian Governments have each announced they are taking further action. Literally hundreds of organisations have signed-up including the medical establishment – and dozens of patient groups too.

In a fast changing world, Ben sees it as increasingly important that the layman has a sense about what good science is, and recognizes that good science is not necessarily common sense. By focussing on pop science he is aiming at anyone who is even remotely motivated so they can get themselves clued-up about what is really happening, and equipped to ask good questions.

In *Bad Pharma,* his second book, he diligently illustrates how common place it is that new drugs are tested only by the people who manufacture them, in poorly designed trials, on hopelessly small numbers of unrepresentative patients, and analyzed using techniques which are flawed by design – all in order to exaggerate the benefits of treatments. Worse still, when independent trials throw up results that the makers don't like, they then feel entitled to hide them away, so the public only ever see a distorted picture of any new drug's real patient outcomes. Clinicians get to hear about what works mainly via ad hoc oral traditions, or from colleagues, sales reps, patient groups, and journals – and the drug companies know just how to distort this process to maximize sales. Ben understands the importance of challenging the grand edifices of knowledge and methodology – and he possesses the courage required to do this. We need him to continue, preferably on a wider front – one that empowers clinicians to drive improvement for themselves.

More than this, Ben debunks the currently practiced medical epistemology, ably demonstrating how new medical knowledge is frequently juxtapositioned against false or misleading comparators, rather than against best current treatments; as well as how some trials are simply stopped too soon – new treatments being assessed over unrealistically short time scales – subsequent complications being omitted from the analysis. Negative results are of course just as important to medical practice as positive ones, yet routinely negative studies are ignored, filed away, unpublished, maybe never even being written up. The potential for harm by ignorance is huge – be it by missed benefits or unrecognised side effects. Bodies like NICE are then left with no choice other than to do their cost benefit analyses based upon incomplete information.

Moreover he is not frightened to condemn his own professions – medicine and academia – saying that they lack ambition, seemingly content to expose themselves and their patients open to exploitation.

> *".. (We) have failed to implement the basic principles of evidence-based medicine properly..*
> *We could be running large randomized trials cheaply, as a matter of routine, in everyday clinical care, to find out which treatments work best — but we've failed to do so.*
>
> *We could be aggressively auditing the extent of withheld trial data on each treatment, to shout about what's missing — but we've failed to do so.*
>
> *We could invest in disseminating and implementing the evidence we have on what works best to decision makers, doctors and patients so that deceitful pharma marketing would be a peripheral irrelevance — but again, we've failed to do this adequately."*
>
> Ben Goldacre (29)

There is precious little to disagree with in Ben's list, though we would want to add that disseminating and implementing evidence depends on individual and organisational motivation and that if we are to make the practice of evidence-based medicine a norm, empowerment (and self-empowerment) of those who work in (and on) the system is a vital ingredient. Ben's work is of huge importance and we are sure that he will continue to point out deficiencies in clinical custom and practice. More than this he makes an apparently impenetrable field accessible to the uninitiated. A key purpose of our paper is to recruit him to the science of improvement – however *poppy* he needs to make it – because such a combination will be more appetising to *both* to the healthcare professionals *and* to the strategists in Richmond House – as well as to the local commissioning leads that have to make the ideas from Whitehall work.

Our wish is for Ben and others, for example the Royal Statistical Society, to look at Improvement Science and see just how much it offers – and how its widespread use would enable the achievement of their goals too. If Improvement Science were to be widely applied and supported, RCTs for example would then get used more appropriately – and as one of several tools positioned alongside others like the XmR Chart.

We hope that Ben and his Cabinet Office co-authors will see our critique and challenges as coming from people with almost identical passions and interests to their own – offering additional benefits to their proposals, and building upon them?

We further hope they will get behind the methods proposed here by us (and the many others we have referenced) – in particular the AIR measurement model – in its entirety – and in so doing increase substantially their uptake in the NHS, the whole healthcare sector, and the rest of the public sector – thus truly helping make a bigger impact.

We would happily meet with them or others at any time in any way that might help this increased take up.

11. What else will ensure policy-making becomes more evidence-based?

> *"There is nothing a Government hates more than to be well-informed; for it makes the process of arriving at decisions complicated and difficult"*
>
> **Lord John Maynard keynes**

Ben Goldacre is as interested in policy-making becoming more evidence-based as we are. He will have in his sights politicians, as well as senior civil servants and their advisors who are contending with a rapidly changing strategic context. In recent years for example the pendulum has swung inexorably towards de-centralisation, and with it devolution too. Healthcare has not been immune from this trend and each of the four UK provinces are now persusing their own approach to policy and how it translates into delivery. As policy becomes more locally driven Commissioners and Local Authorities too are being forced to work together, albeit still strongly steered by the centre.

Meanwhile, individuals making a career in the civil service continue to be required to make a career path choice between *either* policy (making) *or* process (delivery) for Whitehall has historicaly preferred to keep these two activities separate – which is one reason why during the naughties the NHS was split into commissioners and providers. There was an economic rationale for doing this, but it also reflected this historic partition – apparently there are only two kinds of people, those who are bright enough to invent policy, and those who are not – therefore being better suited to mere implementation. When taking the Myers Briggs Test the former frequently score as intuitives (N). In 2007-08, healthcare commissioners (Primary Care Trusts) were required to become "World Class" (56) and much of the scoring for this (yes there was actually scoring, and ranking) was based upon how good the policy statements looked – there being little or no evidence required of actual outcomes.

There are several ways of configuring models for commissioning – especially if policy (theories about beneficial interventions) and process (implementation) can from the outset be integrated (61) – whereby representatives from the system as a whole, work together from the start to study the current data, collect and display appropriate (usually new) baseline data over time, and then co-design interventions that can be implemented initially on a small scale – whilst observing what happens to the baseline. This is improvement science (P-D-S-A) in action and is very different from an approach that can end-up with £ billions spent on what merely *seems* to be a good idea – such as the 2002 initiative to "reduce health inequalities in the UK population" – a policy which by the time (7 years later) it got reviewed by parliament, it was nye impossible to tell whether anything at all useful had been achieved – these words appearing on the final page of their 50 page report:

> **"Our most damning criticism of health inequalities policy is that it is designed & introduced in ways which make meaningful evaluation impossible."**
>
> **Parliamentary Health Committee Report on Inequalities** – March 2009

P-D-S-A is a reminder that people should apply the Scientific Method in order to maximize the learning that is collectively achieved whenever decisions get made and action ensues. It replicates the 17th century work of Francis Bacon: PLAN = agree a "**hypothesis**" → DO = do an "**experiment**" to test the hypothesis → STUDY = collectively

"**evaluate**" what has been learned → and ACT = use that learning to modify the hypothesis and modify the system – ready to start at PLAN once more. This is a never ending cycle of learning enhancement and knowledge accumulation – occurring via cycles of conversation referred to in the diagram below as a **Learning Loop**.

The commissioning cycle we are proposing is P-D-S-A in action. The way commissioning is currently done follows a P-D-R cycle: Plan-Do-Review. Which means simply that someone on high creates a de facto **Plan**, the plan then gets

delivered or **Done**, and finally it gets **Reviewed** or audited to ensure the outcomes were as originally planned – *regardless* of whether during the doing any new learning emerged. Oddly, the rest of the world seems to copy the UK .. This is how Gary Banks, Chair of the Australian Productivity Commission, describes it ..

> *A major failing of governments in Australia, and probably world-wide, has been in not generating the data needed to evaluate their own programs, in particular, there has been a lack of effort to develop the baseline data essential for before-and-after comparisons. As an aside, I should note that quite often even the objectives of a policy or program are not clear to the hapless reviewer; indeed, one of the good things about having reviews is that they can force some clarification as to what the objectives of the policy should have been in the first place (3).*
>
> *In situations where government action seems warranted, a single option, no matter how carefully analysed, rarely provides sufficient evidence for a well-informed policy decision. The reality, however, is that much public policy and regulation are made that way, with evidence confined to supporting one, already preferred way forward. Hence the subversive expression "policy-based evidence"! (3)*

As illustrated earlier, for several years the NHS has actually been recommending the widespread use of the "model for improvement", yet though such a recommendation might appear simply common sense, it is as yet far from being common practice. Why is this?

A key reason is that the thinking underpinning the model for improvement feels so counter-cultural. Throughout the 90s we each worked with several organisations in the automobile supply chain to facilitate their adoption of what in those days was referred to as Japanese methods. Many of our client organisations were being driven out of business by Japanese transplants (Nissan, Honda, Toyota) or were being expected to dance to their tune, and having belatedly realised that these new competitors were not simply loss-leading to gain market share, our clients were desperately trying in double-quick time to catch-up. For many it turned out to be a last throw of the dice, and it quickly seemed easier to be driven out of business than to learn fundamentally new behaviour. In the words of one M.D. "*it is too late for me, I am too old to learn to be Japanese*".

It is now widely recognised that it was "the system" that was Japanese, not the people – indeed the system being used by the transplants turned out to be merely an adaption of some thinking that they had themselves learned post-war, ironically largely from the West. After the initial training had been completed, it turned out that the new Japanese-owned plants such as Nissan in Sunderland (15) were run almost entirely by indigenous Brits. In recognising the significance of purposefully designed training, Don Berwick is urging all NHS staff to "*Learn, master and apply*" the "*modern methods of quality control, quality improvement and quality planning.*"

For all these reasons, those specialising purely in policy are unlikely to be skilled in the use of evidence-based thinking, nor the scientific method. Typically they might think of a Focus Group as their primary tool, informing the marketing process with little attention being paid to being able to analyse the outcomes. Worse, policy specialists are often especially attracted to studies that have emanated from overseas – often without the full context being understood – rather than those closer to home where evidence is easier to grasp and hence challenge. Transparency too is generally to be avoided – especially regarding data, assumptions and methodologies – such that analysis might be replicated, or understood by those likely to be affected by the interventions that follow on from the policy.

The Berwick Review counters this by setting out a vision for the NHS as "a system devoted to continual learning and improvement of patient care, top to bottom, and end to end" (8) and this includes the Research arena – witness the current appointing of "improvement experts" capable of working simultaneously across columns 1, 2 and 3 in the 3-paradigm model that is core to our paper. Patient Safety Collaboratives too are being set-up by NHS England with a

mandate to build improvement capability that is firmly rooted in evidence-based methods. It seems likely that we might at last be creating a "nationally consistent" measurement model being advocated here by Mike Durkin ..

> *"We need to create a nationally consistent measurement model that supports everyone to understand what's working, to measure progress and to identify which interventions are the most effective. Yet whilst doing so we mustn't override the fundamental importance of local ownership and leadership."*
>
> **Dr Mike Durkin: NHS England National Director of Patient Safety** (21)

The two of us are feeling optimistic about all this because the newly emerging structures seem to be designed to work primarily bottom-up, and to depend mostly upon local leadership within nationally established guidelines. It is early days, but the watchful eye of high profile activists like Ben Goldacre can help ensure that this time the culture as a whole really does shift ..

> *"The ecosystem of evidence based medicine is a hopeless patchwork of poorly coordinated players with no real clear design, or plan, or sense of efficiency. That has resulted in tremendous loss of life, in ways that haven't been adequately recognized."*
>
> **Ben Goldacre** (29)

So, to answer our own question: just what will it take for policy-making to become more evidence-based? The short answer is simply: a little appliance of science, but we pick out two enablers which we see as of critical importance:

1. **Stop inadvertently disempowering frontline teams** – for disempowerment leads to disengagement which according to one recent survey () by Gallop is a particular problem in the UK, recording 17% of employees as engaged, or emotionally invested in and focused on creating value for their organizations every day – whereas 57% were not engaged, and 26% actively disengaged or emotionally disconnected from their workplaces.
2. **Start to standardise those things that enable frontline teams** to take charge in a way that is evidence-based and connected to the outcomes that their patients would recognise as good ones. This requires what Deming referred to as "constancy of purpose" and the application of IS-based principles and standards.

In their paper Perla, Provost and Parry (54) conclude that to ensure the continued development and relevance of the Science of Improvement, its epistemological foundations and reasoning methods need to be understood. They say that.. *"If improvement efforts and projects in health care are to be characterized under the canon of science, then health care professionals engaged in quality improvement work would benefit from a standard set of core principles, a standard lexicon, and an understanding of the evolution of the science of improvement".* Here is our *own* illustrative list of principles that are formulated to stop the disempowerment of frontline teams, together with an **illustrative list** of standards that are needed to enable them to deliver continually improved systems – note that these have been uncovered during many years of practise, and we are certain there are more for us to glean, for learning about the science of improvement has to some extent to be a personal journey.

PRINCIPLES : a few examples

- Everyone who work in each system are well led by someone who works on that system and knows how to hold themselves collectively to account using pertinent evidence-based data

- Every managerial (or clinical) leader is required to "Learn, master and apply" the "modern methods of quality control, quality improvement and quality planning" – these methods becoming a core element in the annual appraisal round.

- Appraisal methods are designed with care, firstly to recognise team working, secondly to be based upon sustained performance in systemic capability, and only thirdly to be based on personal contribution.

- Every managerial/clinical leader is supported as they learn system design and improvement methodologies in order to improve systemic capability

- Within each organisation, measurement and accountabilities are systemically-based and principally operate from the bottom-up, yet within a top-down framework – i.e. metrics, both for outcomes and activity, are locally chosen in alignment with national guidelines

- Between organisations, the customer requires suppliers (e.g. commissioners requiring providers) to operate in an scientifically evidence-based way, adopting all of these principles listed here, and continually demonstrating that learning and capability is being enhanced.

- Standardisation is viewed primarily as a bottom-up concept – something to be agreed locally after each P-D-S-A cycle with the aim of reducing variation. Standardisation is an essential to the evolution of ever safer systems.

- Systemic changes are dynamically monitored in as close to real time as practicable. Local teams continually report performance as and when "special causes" (16) occur, or when SPC chart control limits are altered. Explanations of the interplay between their interventions and changes in their system's performance (mean and variation) are upwardly reported.

- SPC charts – to separate signal from noise over time and enable the achievement of stability – are used as a baseline for prognostic experimentation (P-D-S-A).

- Every managerial/ clinical leader knows how to keep the "measurement process" itself stable and consistent.

- Qualitative and quantitative data collected over time needs to be simultaneously analysed so that a data-rich story can be constructed by those who best understand the context.

- All "special causes" are treated as possible early warnings – and hence are "analytically" (16) reported upon.

- Pertinent outcome data (unless specifically protected) is published online, warts and all, and in real time. What is pertinent is agreed with patient representatives.

- Systemic learning is reported in a way that it can be shared across the locale, the region, the country.

- Numerical performance targets may be set, but merely for guidance purposes – and are always referenced against statistically valid external benchmarks. Whether maintaining the local system for stability, or continually improving it, local managerial/ clinical leaders should respond only to actual data – in effect ignoring any arbitrary targets.

- All of the above are conducted in a culture of whole system enhancement – requiring for example a radically new approach to how financial control is achieved. **Productivity** is seen as vital but only as an outcome of **Safety, Flow,** and **Quality** .. in that order.

STANDARDS : a few examples

- A standard set of tools for improvement and innovation – framed by the three paradigms: improvement/accountability/research. In particular a single SPC Tool is recommended to all, so that no one need question the validity of the tool, only the data being charted, and what it is revealing about the system – in a way that renders systemic change identifiable in real time, and signals early warnings. This tool needs to feel empowering to the non-statistical user.

- A standard methodology for improvement and innovation constructed of course around a scientific method, but alongside recommended ways of proceeding such as the API model for improvement Outlined on page 19, and the SFQP® model : safety-flow-quality-productivity (86)

- System performance is continually (upwardly) reported using a standardised format – something akin to the Toyota "A3" tool

- Operational Definitions are required to be unambiguous within the national guidance provided, and are thoroughly tested (bottom-up) before being set in stone (top-down)

- CCG and Provider websites containing real time outcome data, presented in SPC chart format, are accessible to all – with links published on local HealthWatch websites

- Data liquidity and interoperability () to ensure that data can be used across the NHS's many agencies and departments – given proper safeguards, of making the anonymised, pseudonimised, aggregated data available to patients, citizens and researchers – thereby empowering citizen/ patients and care providers, patients and researchers to make better informed decisions, spur new innovations, and highlight prospective efficiency gains

- Learning is always shared, and leads to evidence-based guidance – for example – on what to do about low Friends & Family Test response rates without having to resort to the setting of arbitrary numerical targets, or how to define an episode of care, or how to choose metrics for services with small/ intermittent demand?

- A glossary of succinct terms and definitions is centrally provided, and continually updated according to collective learning about what works best

Note that we are making a distinction between "standardisation" that comes upwards from the frontline, and "standards" that are needed to enable the system as a whole to evolve healthily and which can only emanate from those who have responsibility for the whole system – usually senior leaders.

The point of all these principles and standards is that they collectively make good science do-able. As an epidemiologist Ben Goldacre is like us especially passionate about public health and enabling good science to take place is the only way to enhance the public's health over time. Since the Lansley reforms in 2012 Public Health as a body has been under the political control of local councils, and it has been keeping a pretty low profile. Maybe something spectacular is incubating, but a clear steer has now been given by the new NHS CEO, Simon Stevens, who refers to Berwick's concept of "**the triple aim**" – the idea that we should judge our success relative to population health improvement, the quality delivered for individual patients, and a wise stewardship of resources that ensures high value care – all to be achieved via what Stevens calls four game changers:

(1) **Personalisation** – moving away from care geared for the "median" patient and towards a much greater tailoring of the individual treatments on offer;

(2) **Anticipatory care** – moving away from healthcare systems that rely principally on people pitching up to see a health professional when they get sick – towards healthcare systems that are much better able at stratifying risk, identifying upstream care opportunities, and targeting interventions accordingly; and

(3) **Standardisation** – that originates bottom-up from teams that exercise empowerment (*our words*);

(4) **Co-production** – recognising that it is often the lay "experts by experience" who bring the assets, insights and commitment that will reshape the way care is provided.

In making public health the central aim, it is important to be reminded that historically it has adhered to an epidemiologically driven risk-factor approach to health, targeting for example obesity, alcohol, smoking, and so on, but what's now needed is more of a focus on the whole lives of people, and in the context of their local communities. This should bring a broader and deeper meaning to the concept of health, taking account not only of conventional determinants, but also: educational, social and participatory, and environmental aspects such as food, nutrition, violence, poverty, loneliness, employment.

Traditional epidemiology is a powerful scientific tool for exploring the causal determinants of disease. But when translated to the arena of public health, epidemiology is restrictive and reductive in its vision.

Public Health needs a renaissance. As the world moves into an era of sustainable development, UK public health should not miss this opportunity to rewrite the contract between health and society.

Richard Horton, Editor-in-Chief The Lancet - Dec 2014

12. Summary & conclusions

The two of us, in our roles as citizen/patients notice that the healthcare "system" often succeeds, but sadly all too often does so despite itself – and largely via the efforts of heroic individuals rather than through a well-constructed system. Healthcare seems too rarely to be **viewed as a system** by those responsible for designing and operating it – many of them blind to its potential or to its failures and/or to the possibility of fully learning from them. In recent years, more and more data seems to have been collected in order to inform political discussion in Westminster, and/or resource allocation in Whitehall – all too often, as if by one-club golfers, being achieved only bluntly and arbitrarily via top-down targets, or via performance rankings that merely name-and-shame – meanwhile at the coal face next to no meaningful data is collected. Locked into a carrot & stick ethos, good science seems to be rendered a cultural impossibility, a considerable irony given that the system is driven by highly educated professionals most of whom will have had years of science education.

During the eight decades of our combined working lives, covering both the private and public sectors, if we learned just one thing it was that when people are asked to deliver something of which the system they are leading is currently incapable, there are only three options ..

(iv) Improve the system's capability;
(v) Distort the system; or
(vi) Distort the data.

The first option requires some understanding of the existing system; some knowledge of how best to change its design; a tried and trusted method for affecting the changes so that they can sustain; and a little science so that it can be demonstrated that any observed new outcomes are real. This first option is the good science option, but sadly is rarely considered expedient – most people, most of the time opting, with varying degrees of cynicism, for either system or data distortion.

Ben Goldacre continually admonishes bad science and like him we too worry about it, so in this paper our aim has been to uncover the nature of "good science" and to point the healthcare sector and its professionals toward practical ways of doing more of the good kind of science, less of the bad. Ben too wants to make good science easier to do, but we question the need for the new Test, Learn and Adapt (TLA) model he offers because the NHS already possesses such a model – one which in our experience is more complete and often simpler to follow – it is called the "Improvement Model" – and via its Plan-Do-Study-Act mnemonic embodies the scientic method in a way that the inventor of that term Francis Bacon would have recognised. Moreover there is a pre-existing wealth of experience on how best to embed this thinking within organisations – from top-to-bottom *and importantly* from bottom-to-top – experience that has been accumulating for fully nine decades – and though founded in industrial settings has long since spread to services.

Healthcare, as well as the public sector as a whole, deserves a unified approach to the avoiding of bad science. Ben's goal of making this possible is both timely and courageous – not least because his choice of collaborator is a group of civil servants who are ostenisbly disconnected from the Dept. of Health. We would love to enlist his boundless energies to the cause of evidence-based scientific thinking – via methodologies and tools that are as accessible to healthcare professionals as they are to politicians, and policymakers in general.

In offering a 3-way definition of science this the longer version of our paper rues the dismal way in which science is conveyed to children and students, the majority of whom leave formal education without understanding the power of discovery or gaining any first hand experience of the scientific method. If science were to be accessibly defined around discovery, and learning cycles, and built practically upon observation, measurement and the accumulation of evidence – then good science could be viewed as a process rather than merely as an externalisd entity. These things comprise the very essence of what Berwick refers to as Improvement Science – embodied by the Institute of Healthcare Improvement (IHI) and in the NHS's Model for Improvement.

We also aimed to bring an evolutionary perspective to the whole idea of science, arguing that its time has been coming for five centuries, yet is only now more fully arriving. We suggest that in a world where many at school have been turned-off science, if rendered accessible and practical, the propsensity to be scientific in our daily lives – and at work – makes a vast difference to the way people think about outcomes and their achievement. This is especially so if those who take a perverse pride in saying they avoided science at achool, or who freely admit they don't do numbers, can get switched on to it.

The NHS Model for Improvement has a pedigree originating with Walter Shewhart in the 1920s, then being famously applied by Deming and Juran after WW2. Deming in particular encapsualtes the scientific method in his P-D-C-A model (three decades later he revised it to P-D-S-A) – his pragmatic way of enabling a learning/ improvement to evolve bottom-up in organisations. After the 1980s Don Berwick then, standing on these shoulders, applied the same thinking to the world of healthcare – initially in his native america. Berwick's approach is to encourage people to ask questions such as .. what works?.. and how would we know? His method, is founded upon a culture of evidence-based learning, providing a ***local context*** for systemic improvement efforts. A new organisational culture, one rooted in the science of improvement, if properly nurtured, can in this way emerge.

However, such a culture may initially jar with the everyday life of a conventional organisation, and the individuals within it. One of several reasons, according to Yuval Harari, is that for hundreds of generations our species has evolved such that *imagined* reality has been lorded over *objective* reality. Only relatively recently in our evolution has the advance of science been levelling up this imbalance, and we argue that a method is now needed that enables these two realities to more easily coexist. We suggest that a method enabling data-rich and evidence-based storytelling – by those who most know about the context and intend growing their collective knowledge – will provide the basis for an approach whereby the two realities may do just that.

A vital facilitator of this new organisational culture is the 3-paradigm "**Accountability/ Improvement / Research**" model of measurement (**AIRmm**) reflecting as it does the three archetypal ways in which people observe and measure things. It is a tool originally created by healthcare professionals to help their colleagues, and policy-makers, to unravel the confusion that commonly pervades our everyday working lives, and to help people make better sense of the different approaches they may adopt when needing to evidence what they're doing. An amended version of this model is already widely quoted inside the NHS, though this is not to imply that it is yet as *widely* understood or applied as it needs to be.

The 3-paradigm **AIR** measurement model underpins the way that science can be applied by, and has practical appeal for, the stretched healthcare professional, managerial leader, civil servant – indeed for anyone intuitively who feels there must be a better way to combine goals that currently feel disconnected or even in conflict: *empowerment and accountability; safety and productivity; assurance and improvement; compliance and change; extrinsic and intrinsic motivation; evidence and action; facts and ideas; logic and values;* etc .. indeed for anyone who is searching for ways to unify their actions with the system-based implementation of those actions as interventions. In particular the model makes it immediately plain that an alternative to the apparent **Accountability/ Research dichotomy** is readily available which risks transfering all the responsibility for "learning" to academia.

In many organisations currently, the Column 1 (Accountability) paradigm is the only game in town. Column 3 (Research) may seem attractive as a way-out, but it also feels inaccessible except via a graduate in statistics if one happens to be available? Moreover, the Randomised Controlled Trial feels altogether overblown and lacking in immediacy. As Berwick says *"Fans of traditional research methods view RCTs as the gold standard, but RCTs do not work well in many healthcare contexts"* (7). An RCT can feel much like reaching for a spanner and finding a lump hammer in your hand.

Like us, Ben is frustrated by the ways that healthcare organisations conduct themselves – not just the drug companies that commercialise science and publish only the studies likely to enhance sales, but governments too who implement politically expedient policies and subsequently invent evidence to support them .. policy-based

evidence rather than evidence-based policy. His recommended Column 3 T-L-A approach is however more likely to make day-to-day sense to people and teams on the ground if complemented by a Column 2 approach – and bridging the chasm that normally exists between the coulmn 1 and column 3 paradigms.

One reason why Improvement Science can sometimes fail to dent established cultures is that it gets overly-simplified by organisational "experts" – many of whom then use what little knowledge they have gleaned only to try to make themselves indispensible, not realising the extent to which everyone else as a consequence gets disempowered. In our papers we therefore wanted to take the opportunity to outline the philosophical underpinnings, and to do this we have borrowed the 7-point framework from a recent paper by Perla et al (54) who suggest that **Improvement Science**

1. Is grounded in testing and learning cycles.
2. Embraces a combination of psychology and logic.
3. Has a philosophical foundation is conceptualistic pragmatism.
4. Employs Shewhart's theory of cause systems.
5. Requires the use of operational definitions.
6. Considers the context of *both* justification *and* discovery.
7. Is informed by Systems theory.

To complete this philosophical framing we use the 4 elements in Deming's System of Profound Knowledge because we especially want to underscore that **contextualised knowledge** is the overall aim, and that this is all the more powerful if continually generated in context-specific experimental cycles. Deming showed that good science requires a **theory of knowledge** based upon ever-better questions and hypotheses. We two are now aiming to bring further pragmatism to each context by developing new ways of building knowledge-full narratives.

There are two further elements that are central to what we wanted our paper to convey. First we have tried to demonstrate that **context** is of the very highest importance to improvement and innovation, but also that an understanding of context can never be complete – and that only through our attempts to describe its multi-dimensional-ness can we connect to it powerfully enough to change it. Those however who talk about the importance of context often tend to be those who are anti-science and we hope our paper prompts more personal integration for organisational leaders (at every level) – especially clinical leaders who are nowadays doing (or refusing to do) most of the managing. Because MBTI is the most widely used diagnostic of personality we have used it to illustrate how reality is a combination of what people perceive and how they process their perception. We wanted to show that *both* mythos *and* logos are needed if people are to be able *to both* make sense of THEIR world, *and* make sense of THE world.

Secondly, that **storytelling** is a core human skill, one that we've been honing for at least 70 millennia – but a skill that can be honed and much enhanced by more connectedness to "scientific reality", by more consciousness about learning cycles, by recognising the different aspects of ourselves, and with fuller empathy for fellow collaborators. Very many people who extol the virtues of better storytelling (now commonplace in the NHS) appear simultaneously in denial about the importance of connecting in any scientific way to their context as it ***has been***, as it ***is now*** in this moment, and as it predictably (or not) ***will be***. If the reason for enhancing our collective storytelling ability is only to build a believable story – any story, however disconnected from science – then leadership is dangerously diminished. A proper integration between mythos and logos becomes impossible.

Yes, MBTI is the most widely used tool, at least in the western world, for enabling individuals to first understand themselves and then understand others – and as such is a powerful lens through which individuals may understand their reality – at least their personal preference for the way they view it. MBTI is also widely used by NHS leaders, but those who most refer to it are also the least likely to want to be scientific about their work in the way that an engineer might do. We hope our paper confronts that notion.

Why is this so important? Well, the requirements of managerial leaders in the NHS and the wider healthcare system are fast changing – as recently acknowledged here ..

> *Current NHS architecture depends on alignment and consensus rather than the use of crude levers. As we move forwards, leaders are needed (who are) grounded in **common values** with a broad outlook that is patient-centred, population focused, and cost aware. Leaders with experience of **innovation, improvement,** and **implementation** at pace, (who are) empowered rather than hindered by the system.*
>
> Mahiben Maruthappu & Bruce Keogh
> "The future of health system leadership" – Lancet – June 2015

At the start of this paper we referred to the 2013 Francis Report into why upto 1200 patients unexpectedly died in the Mid-Stafforshire hospital between 2005 and 08. Actually, these are the official dates, but if even the most basic improvement science had been in place the problems would have been identifiable at least 3 years earlier than that. To the two of us this is what evidence-based knowledge needs to mean – being able to test and respond in near to real time. At that time in the West Midlands, and across Britain, there was plenty of column 1 thinking in place but it was column 2 thinking that was needed, which had it been in place would very likely have saved a considerable number of lives.

None of that is possible without the cultural transformation heralded by the Berwick Report. We wanted to show how a combination of "**bandwidth**" in individual leaders (62) *and* a well thought-through team-oriented method – one that can become standardised across the NHS, provides the necessary human dimension to the task of enabling a Berwick-style culture – one characterised by continual systemic improvement, and innovation. That in a sentence is our ambition with the paper. We will know that we will have succeeded if it becomes widely read, but more than this we aim to enlist activists like Ben to the definition of science embodied by Improvement Science.

Don Berwick's proposals for the building (designing and monitoring) of safe healthcare systems *are* being taken up but only slowly, yet with Ben Goldacre's encourgement, of intervention that is evidence-based, this can happen more quickly and more assuredly. However, in recommending RCTs alone, Ben Goldacre and his colleagues at the Cabinet Office risk making us all look like one-club golfers. Sometimes things are a bit more complicated than that – as Einstein warned *"everything needs to be made as simple as possible, but no simpler".* Actually, we believe the risk to be low, but only because RCTs are so hard to apply in practice they are destined to remain the preserve of professional statisticians and/or academic institutions.

Nowadays, young healthcare professionals are complusorily inducted into research methodology, but it is often badly taught, and worse, the fact that it is made compulsory sometimes destroys what intrinsic worth it might have had – just as most people are forced to study Mathematics when at school, but leave to join the world of work determined not to do numbers ever again. The crucial thing is that everyone who works in the healthcare system needs to be able to the live the appliance of science. Most people at school were taught about science, but not how to be a scientist, and many left school being intimidated by numbers. Which is not to imply that science is only about more numbers – it is actually do do with good data, both the qualitative and the quantitative kind. Our task is to be able to provide an approach, a method, and some practical tools, that enable intimidated individuals to turn-on to being good scientists.

We also wholeheartedly agree with Ben that for the public sector – not just in healthcare – policy-making needs to become more evidence-based. In a poignant blog from the Health Foundation's Richard Taunt (69), he describes his recent attendance at two conferences in one day. At the first one, policymakers from 25 countries had assembled to discuss how national policy can best enhance the quality of health care. When collectively asked which policies they

would retain and repeat, their list included: *use of data, building quality improvement capability, ensuring senior management are aware of improvement approaches, and supporting and spreading innovations*. In a different part of London, UK health politicians happened also to be debating Health and Care to establish the policy areas they would focus on if they were to form part of the next government. This second discussion brought out a completely different set of areas: *the role of competition, workforce numbers, funding, and devolution of commissioning*. These two discussions were supposedly about the same topic, but a Venn diagram would have contained next to no overlap.

Clare Allcock, also from the HF, then blogged to comment that "*in England, we may think we are fairly advanced in terms of policy levers, but (unlike, for example in Scotland or the USA) we don't even have a strategy for implementing health system quality*." She points to Denmark who recently have declared they are phasing out their hospital accreditation scheme in favour of an approach with quality improvement methodology and person-centred care at its heart – in the words of their President of Danish Regions ..

> *"Quality work must be simplified and focused. The time has come to strengthen it by putting the patient at the centre, rather than focusing on compliance with a variety of standards. Accreditation has been justified and useful, but we move on. We need a few national targets to be met locally with strong commitment from the staff and with room for local solutions."*
>
> Bent Hansen (2)

The Danes are in effect taking the 3-paradigm A-I-R model and creating space for Column 2 improvement thinking. The UK needs to take a leaf out of their book, for without changing fundamentally the way the NHS (and the public sector as a whole) thinks about accountability, any attempt to make Column 2 the dominant paradigm is destined to be still born. It is worth noting that in large part the Column 2 paradigm was actually central to the 2012 White Paper's values, and with it the subsequent Outcomes Framework consultation – both of which repeatedly used the phrase "bottom-up" to refer to how the new system of accountability would need to work, but this seems to have become lost in legislative procedures that history will come to regard as having been overly ambitious. The need for a new paradigm of accountability however remains – and without it health workers and clinicians – and the managers they hire to support them – will continue to view metrics more as something intrusive than as something that can support them in delivering enhancements in sustained outcomes. Now is the moment to stand up and be counted, or (to use Simon Stevens' words) even "rattle the cage and advocate something different" ..

> *"Society, in general, and leaders and opinion formers, in particular, (including national and local media, national and local politicians of all parties, and commentators) have a crucial role to play in shaping a positive culture that, building on these strengths, can realise the full potential of the NHS.*
>
> *When people find themselves working in a culture that avoids a predisposition to blame, eschews naïve or mechanistic targets, and appreciates the pressures that can accumulate under resource constraints, they can avoid the fear, opacity, and denial that will almost inevitably lead to harm."*
>
> Berwick Report

Changing cultures means changing our habits – it starts with us. It won't be easy because people default to the familiar, to more of the same. Hospitals are easier to build than relationships; operations are easier to measure than knowledge, skills and confidence; and prescribing is easier than enabling.

If we are going to realise the renewable energy of people and communities, we need to put a lot of energy into changing the way the 'system' works: how policy makers prioritise what's important; enhancing the skills and competence of the workforce; investing in building community capacity and individual capability. Easy to say, but is it too big to bite off? It cannot be if we're serious about sustaining the NHS.

From the "Realising the Value" programme, led by the **Health Foundation** and **Nesta**
http://www.nesta.org.uk/blog/realising-value#sthash.rlEP7Guv.dpuf

We do not of course possess a monopoly on all possible solutions, but our experience tells us that now is the time for: evidence-rich storytelling by frontline teams, by pharmaceutical development teams, by patients and carers conversing jointly with their physicians – all of these require new thinking and rapidly evolving practise.

We also know that measurement is not a magic bullet, but what frightens us is that the majority of people seem content to avoid it altogether. As Oliver Moody recently noted in The Times ..

Call it innumeracy, magical thinking or intrinsic mental laziness, but even intelligent members of the public struggle, through no fault of their own, to deal with statistics and probability. This is a problem. People put inordinate amounts of trust in politicians, chief executives, football managers and pundits whose judgment is often little better than that of a psychic octopus. Short of making all schoolchildren study applied mathematics to A level, the only thing scientists can do about this is stick to their results and tell more persuasive stories about them. (36)

Too often, Disraeli's infamous words: "***Lies, damned lies, and statistics***" are used as the refuge of busy professionals looking for an excuse to avoid numbers.

As a first step we recommend enabling teams to access good data in (or near to) real time – data that indicates the impact that one's intervention is having – this alone often prompts a dramatic shift in the type of conversations that people working in and on the system may have. Often this can be achieved simply by converting existing KPI data into System Behaviour Chart form which, using for example the BaseLine® tool (4) takes only a few mouse clicks.

We offer 3 examples of Improvement Science in action – collectively illustrating how data may be used to evidence *both* sustained systemic enhancement, *and* to generate engagement by the people most directly connected to what is systemically happening in real time – one from acute care, one from primary care, and one by a patient with a long term condition providing their own care.

The landscape of NHS improvement and innovation support is fragmented, cluttered, and currently pretty confusing. Since May 2013 Academic Health Science Networks (AHSNs) funded by NHS England (NHSE) have been created with the aim of bringing together health services, and academic and industry members. Their stated purpose is to improve patient outcomes and generate economic benefits for the UK by promoting and encouraging the adoption of innovation in healthcare. They have a 5 year remit and have spent the first 2 years establishing their structures and recruiting, it is not yet clear how effective they are likely to be?

In the wake of the Berwick Report, Patient Safety Collaboratives – linked with AHSN areas – have been established to improve the safety of patients and ensure continual patient safety learning. The programme is coordinated by NHSE and NHS Improving Quality (NHSIQ) and will provide safety improvements across in a range of healthcare settings by tackling the leading causes of avoidable harm to patients. The intention is to empower local patients and healthcare staff to work together to identify safety priorities and develop solutions – implemented and tested within local

healthcare organisations, then shared nationally. Hopefully AHSNs will be able to coordinate all these activities with a single-minded attention to scientific method?

For our part, we two intend to contribute to these initiatives in any way we can. It is our hope that Improvement Science enables the cultural transformation we have envisioned in our papers and with our case studies. This is what we feel most equipped to help with. When in your sixties it easy to feel that time is short, but maybe people of every age should feel this way?

> **"Begin doing what you want to do now. We are not living in eternity. We have only this moment, sparkling like a star in our hand-and melting like a snowflake..."**
>
> **Francis Bacon**

Notes & Bibliography

1 Ackoff R. – From data to wisdom – 1989

2 Allcock C. – Health Foundation blog "From compliance to commitment should the NHS look to the Danish on how to accelerate change and improvement?" - 29th April 2015

3 Banks G. – Chair of the Australian Productivity Commission – "Evidence-based policy making. What is it? How do we get it? 2009

4 Baseline® http://www.saasoft.com/baseline/baseline.php is a statistical tool designed with (self) empowerment in mind which is why it is available at under £50 so that individuals who feel unsupported by their boss or their organisation can afford to buy for it themselves. For the same reason it also comes complete with free training materials.

5 Bate P. – Context is Everything -- www.health.org.uk/media_manager/public/75/publications_pdfs/Context%20is%20everything%20-%20Bate.pdf

6 Berwick D. – Lecture http://livestream.com/IFQSH/ForumLIVE - 24th April 2015

7 Berwick D. - The Science of Improvement – 2008 : http://www.allhealth.org/BriefingMaterials/JAMA-Berwick-1151.pdf

8 Berwick Report - Aug 6th 2013 https://www.gov.uk/government/publications/berwick-review-into-patient-safety

9 Bohannon J. --- I Fooled Millions Into Thinking Chocolate Helps Weight Loss - June 2015

10 Briggs Myers Isabel – Introduction to Type – 1998, 2000

11 Briggs Myers Isabel – Introduction to Type – revised by Mary H. McCaulley – 1989

12 Buckingham M. – " Most HR Data Is Bad Data" – HBR – February 9th 2015

13 Charon R. (MD) - The turn toward narrative knowledge – JAMA 2001

14 Charon R. MD - Narrative Medicine: A Model for Empathy, Reflection, Profession, and Trust – JAMA 2001

15 Collingridge J. – "Nissan is stepping on the gas at Sunderland" – Sunday Times – 5 April 2015

16 Deming W. Edwards – Out of the Crisis – 1982

17 Deming W. Edwards – The New Economics – for industry, government and education – 1993

18 Dodds S -- Case Study of Reducing Surgical Site Infection (SSI) using a Bundle of Care Model and the Science of Improvement Method. Journal of Improvement Science 2014:18; 1-24. http://saasoft.com/jois/jois_view_abstract.php

19 Dodds S. R. BLOG – Feb 10th 2013 http://www.saasoft.com/blog/?p=2753

20 Donaldson Liam – Chief Medical Officer – Dep't of Health – 2008

21 Durkin M. – at the 'design day' event initiating a nationwide network to improve patient safety - 21 January 2014

22 Dweck C. - Mindset: How You Can Fulfil Your Potential - 2012

23 Gallop Survey - http://www.gallup.com/poll/181202/britons-purpose-room-improve.aspx - January 2015

24 Gawande A. – 2014 Reith Lectures – Lecture 2 of 4 "the century of the system"

25 Glasziou P. – Six proposals for evidence based medicine's (EBMs) future – and for shared decision making – BMJ Clinical Evidence - 27 Mar 2015

26 Goldacre B. – Bad Science – 2008

27 Goldacre B. : Test, Learn, Adapt – developing public policy with randomised controlled trials - Cabinet Office Paper – Dr Laura Haynes, Orwain Service, Dr Ben Goldacre, Professor David Torgerson - - October 2014 www.gov.uk/government/publications/test-learn-adapt-developing-public-policy-with-randomised-controlled-trials

28 Goldacre B. interview "Meet the anti-Dr. Oz: Ben Goldacre" updated by Julia Belluz on December 27, 2014

29 Goldacre B.– Newsnight vignette on Evidence-based Policy for renewing our approach to democracy : Feb 15th 2015 : www.youtube.com/watch?v=u6mDhW0WvUE&utm_source=dlvr.it&utm_medium=twitter

30 Harari Y.N. – Sapiens – A brief history of mankind - 2014

31 Health Foundation – "Perspectives on context" -- March 2014

32 Horton R. - Editor-in-Chief The Lancet - Priorities for UK public health 2040 - Lancet Mini Essays - 29th Dec 2014

33 Hoyningen-Huene P. – The Context of discovery and the context of justification – 1987

34 Improvement Science Research Network Website: http://isrn.net/about/improvement_science.asp

35 Jung C. – Psychological Types - 1921

36 Kahneman *D. – Thinking, Fast and Slow* – 2011

37 Keirsey D. – www.keirsey.com/

38 Keirsey K. - - http://www.keirsey.com/4temps/bob_dylan.asp

39 Kuhn Thomas S. - the structure of scientific revolutions - 1962

40 Langley G.L. Nolan K.M. Nolan T.W. Norman C.L. Provost L.P - The Improvement Guide (and model for improvement): A Practical Approach to Enhancing Organizational Performance – API (Associates in Process Improvement) - 1996

41 Lewis C.I. – Mind and the world order – outline of a theory of knowledge – 1929

42 Mahiben Maruthappu M. & Keogh B. – "The future of health system leadership" – Lancet - June 2015

43 Maslow A. H. - Toward a Psychology of Being - 1962

44 Mauléon C. & Bergman B. - Exploring the epistemological origins of Shewhart's and Deming's theory of quality: Influences from C.I. Lewis' conceptualistic pragmatism – International Journal of Quality and Service Sciences – 2009 – Vol. 1 Iss: 2, pp.160 – 171

45 Meadows Dana H.– Thinking in systems – 2009

46 MES and In Health Associates - MES in collaboration with - 'Making Sense and Making Use of Patient Experience Data' – June 2015

47 Moody O. - When science gets tough, we're suckers for superstition - The Times 2nd May 2015

48 National Quality Board – annual report – Dep't of Health : January 2013

49 National Quality Board Handbook : A Quality Dashboard for the NHS in England -- Nov 2012

50 North Somerset Clinical Commissioning Group – https://www.northsomersetccg.nhs.uk/

51 Nuzzo R. -- Scientific method: Statistical errors -- P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume. - Feb 12th 2014

52 Pawson R. – Evidence-Based Policy: a Realist Perspective – 2006

53 Pawson R. & Tilley N. – Realistic Evaluation – 1997

54 Perla, R.J., Provost, L.P., and Parry, G.J. Seven Propositions of the Science of Improvement: Exploring Foundations. Q Manage Health Care 2013 Vol. 22, No. 3, pp. 170–186.

55 Schon Donald A. – The Reflective Practitioner: How Professionals Think in Action Paperback –1984

56 Schonberger R.J. - World Class Manufacturing – 1986

57 Schonberger Richard J. World Class Manufacturing – 1986

58 Senge P. et al – The 5th Discipline Fieldbook – 1994

59 **Shewhart W. A. – "Economic control of quality of manufactured product" – 1931**

60 Shojania Kaveh G. & Grimshaw Jeremy M. -- Evidence-Based Quality Improvement: The State Of The Science. Health Affairs, 24, no.1 (2005):138-150

61 Simcox J. A. - Commissioning for efficacious outcomes - Journal of Improvement Science - 2013 http://www.improvementscience.net/jois/index.php

62 Simcox J. A. – Detailed Descriptions of the Developmental Stages or Action Logics of the Leadership Development Framework – an on-going inquiry into the polarities and tensions between the processes of Managing and Leading seen through the lens of the LDF and Deming – 2005 http://harthill.co.uk/harthill-resources/articles-and-papers/ in systems -2009

63 Simcox. J – An Information System for the Self-Managing Patient *Journal of Improvement Science* 2013: 14 1:32

64 Snook J. – "Toyota's secret: the A3 Report" – MIT Sloan Management Review – July 2009

65 Solberg L.I. Mosser G. McDonald S. – "the three faces of performance management – improvement, accountability and research" - Journal of Quality Improvement – 1997

66 Stephen T. Ziliak & Deirdre N. McCloskey - 'The Cult of Statistical Significance – How the Standard Error costs us jobs, justice and Lives' - 2008

67 Stevens S. - "Stevens issues clarion call to NHS leaders to redesign care for patients" - 3 June 2015 -- www.england.nhs.uk/2015/06/03/redesign-care/

68 Stevens S. - 'Thinking like a patient, acting like a taxpayer – from NHS challenges to new solutions?' - Speech to NHS Confederation Annual Conference - June 2014

69 Taunt R. - Learning our lesson: England's place in the health improvement world – Health Foundation May 6th 2015

70 Taylor M.J. – "The widespread non-adherence to PDSA methodology guidelines in healthcare as demonstrated by systematic review" Imperial College London - presentation Sept 11th 2014 to the Royal College of Surgeons

71 The Learning Handbook : NHS Improving Quality, Leeds, LS1 4HR http://www.nhsiq.nhs.uk/media/2532607/nhs_iq_-_learning_handbook_-_pdsa.pdf

72 **The NHS Institute, and The Institute of Healthcare Improvement**

73 Tomolo A, Lawrence R, Aron D. Revised – Conceptual model of PDSA method – Postgrad med J 2009;85:530-7

74 Torbert W.R. – "the power of balance" - 1991

75 Torbert W.R. & Associates – Action Inquiry – The secret of timely and transforming leadership – 2004

76 **Tversky A. & Kahneman D. – Judgement under uncertainty - heuristics and biases – Science (New Series) – Volume 185 - Sept 1974**

77 **Vincent C. Professor -- Director at Imperial College Centre for Patient Safety and Service Quality. http://www.health.org.uk/news-and-events/newsletter/one-year-on-reflections-from-the-berwick-advisory-group/**

78 **Weight T. -- book review in Royal Statistical Society Electronic Newspaper – October 2011**

79 **Wheeler D.J. – Advanced Topics in Statistical Process Control – the power of Shewhart Charts – 1995**

80 **Wintour P. – The Guardian - 28 December 2014 "NHS hampered by poor management structure Rose report claims - Former M&S chief blames 'hydra-like' complexity of system and poor training and development for holding back staff"**

81 **Woodcock T. – The Health Foundation "Twelve tips for measuring improvement" - 30 April 2014 http://www.health.org.uk/blog/twelve-tips-for-measuring-improvement/**

82 **Ziliak Stephen T. &. McCloskey Deirdre N - 'The Cult of Statistical Significance – How the Standard Error costs us jobs, justice and Lives' - 2008**

Supplementary References

83 **King's Fund Report "Better Value " --- July 2015**

84 **Peckham S, Falconer J, Gillam S, Hann A, Kendall S, Nanchahal K, Ritchie B, Rogers R, Wallace A. --- "The organisation and delivery of health improvement in general practice and primary care: a scoping study --- Journal: Health Services and Delivery Research Volume: 3 Issue: 29 --- June 2015**

85 **MES in collaboration with InHealth Associates -- 'Making Sense and Making Use of Patient Experience Data' -- June 2015**

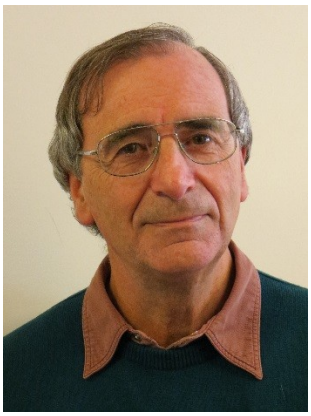
86 **SAASoft -- FISH & ISP Training --- http://saasoft.com/fish/course/**

Authors

Formerly a business coach, specializing in personal and organizational development, and system-wide performance, Julian Simcox is now semi-retired.

He is Chair of the Practice Participation Group for the GP surgery where he has been a patient since 1984. He is also the patient representative on the Quality Assurance Group for North Somerset CCG. In both roles he needs to know and understand just what leads to quality outcomes for patients.

Recently he has also become a Public Contributor for his local AHSN.



A graduate in Statistics, Terry Weight throughout a 40 year career has applied statistical methods in a wide variety of organisations and situations.

He worked as a senior manager in ICI, as a director/consultant in a Process Management consultancy, and as an improvement consultant in Shell – successfully guiding improvement in many organisations, whilst also teaching, advising and mentoring of fellow consultants.

Now retired, yet active in his local community, Terry continues to absorb and apply the methods of Improvement Science wherever he sees they can make a difference.