Value

'In an attempt to arrive at the truth, I have applied everywhere for information but in scarcely an instance have I been able to obtain hospital records fit for any purpose of comparison. If they could be obtained they would enable us to answer many questions. They would show subscribers how their money was being spent, what amount of good was really being done with it or whether the money was not doing more mischief than good…

Florence Nightingale Notes on Hospitals 1863

"… the Trustees of our Charitable hospitals do not consider it their duty to see that good results are obtained in the treatment of their patients. They see to it that their financial accounts are audited but they take no inventory of the Product for which their money is expended.

Codman E.A. A study in hospital efficiency 1918

In health care, the overarching goal for providers, as well as for every other stakeholder, must be improving value for patients, where value is defined as the health outcomes achieved that matter to patients relative to the cost of achieving those outcomes. Improving value requires either improving one or more outcomes without raising costs or lowering costs without compromising outcomes, or both. Failure to improve value means, well, failure.

Porter M.E. and Lee T. 2013

If she were still with us today Florence Nightingale would be surprised, and given her activist proclivities and many other achievements, [1] probably rather vexed, to learn that a century and a half later many, probably most, hospitals around the world still do not have records fit for the purpose she had in mind, - to measure the value of their services by linking outcomes with costs. So too would be Ernest Codman a Boston surgeon who over a century ago carefully documented the outcomes of all his operations including all the complications, and who was so perturbed by the failure of his colleagues to do the same and by the failure of hospitals to make this mandatory, that he set up his own "End results" hospital. He was not shy in his criticisms and unsurprisingly this endeared him to neither his colleagues nor the hospitals of the day. His career went into a steep decline and his ideas went into hibernation for nearly a century.

In the last decade or so however, there has been some progress towards the linking of outcomes and costs as 'value-based' healthcare' (VBHC) which raises four questions:

- why has it taken so long to take up the Nightingale/Codman challenges?
- what has stimulated the recent interest and activity?
- what does all this mean for the implementation of VBHC?
- where to from here?

Contents

Why the long delay? ... 3
Ethos and ethics ... 3
Evidence based practice ... 3
Activity based funding ... 5
Rational ignorance ... 6
Why now? ... 7
Concerns about costs and quality ... 7
Unjustified variations ... 7
Low value clinical practice ... 7
Falling value ... 8
Low or no outcome/cost correlations ... 8
Widening perspectives and priorities ... 9
The consequences of rapid growth ... 9
Cumulative risks of error ... 9
Diminishing returns from increasing complexity ... 10
Increasing chances of interactions ... 10
Bounded rationality ... 11
What does all this mean? – six recommendations ... 12
Ask the right people the right questions ... 12
Rethink and reorganise the datasets ... 14
Start with the low hanging fruit ... 15
Measure processes as well as outcomes ... 16
Provide fast and relevant feedback ... 16
Follow the leaders ... 16
Where to from here? ... 18
The necessity-sufficiency gap ... 18
New opportunities ... 20
Problem definition and linkage ... 20
Analysis and display ... 21
Conclusions ... 26

Why the long delay?

There have been four constraints on the adoption of VBHC all of which have arisen, as they so often do in human affairs, from the unforeseen consequences of good intentions.

Ethos and ethics

In healthcare in general and in large hospitals in particular there is often a conflict between economic value and human values – between what *can* be done within the limits of available resources and what we feel *should* be done. These human values arise from both traditional medical ethics and from the desire in the wider community to help those afflicted by illness or injury who cannot help themselves because of poverty or some other disadvantage.

These admirable moral principles have been implemented in practice over the centuries by religious orders and charitable organisations and are embedded in the ethos of most public hospitals such as St Bartholomew's and St Thomas' in London, both of which were founded by Augustinian monks in 12th and 13th centuries. The same ethos lies at the heart of many national health systems that provide, through collective funding from taxation or insurance, services that would otherwise be beyond the financial resources of any one individual.

The resultant conflict between economic and ethical forces comes in several shapes and sizes, but in the current context, clinicians moving into management often find it difficult to reconcile their professional ethical obligation to optimise the care of the individual patient with the competing needs as managers to deliver these resources more equitably and cost-effectively across populations.

These tensions and how they are viewed differently by different hospital staff groups were well illustrated in a study of their attitudes to financial accountability and standardised processes of care. As shown in this simplified version of the figure from this paper, doctors and nurses working as clinicians or as managers and general non-clinical managers are clearly marching to very different drummers. This is not surprising given their respective roles and responsibilities, but these differences have to be accommodated in any system of VBHC as they can become acute through recourse to the 'rule of rescue' that runs deep in most humans even where continuation of care seems futile to those involved as decision makers or observers.

Evidence based practice

The randomised controlled trial (RCT) is rightly viewed as a key weapon in the armamentarium of evidence based medical practice. This is because it provides a powerful way to determine with measurable confidence whether a drug or other intervention is better than nothing or than some alternative. It does however require at least two groups of patients carefully matched in age, sex, disease severity etc to test for differences that are then only attributable, at least in theory, to the intervention in question. This is essential for

the type of frequentist statistical analysis needed to prove or disprove the null hypothesis and to assess the significance of any differences in outcomes.

The problem with this however is that such narrowly defined short term studies are not good guides for the untidy, uncontrolled and often uncontrollable long term world of routine clinical practice. This is because of their limited external validity due to the infinite variety and multiple pathologies of unselected patents.

Nor are RCT's of any use for the long-term tracking of outcomes and costs needed for routine service management including VBHC. Fortunately there are very suitable though less often used methods of collecting and analysing 'practice based 'evidence to complement and extend the reach of 'evidence based' practice, most notably Bayesian analysis and statistical process control. (SPC)

Bayes theorem is well suited to the generation of practice-based evidence as it closely mirrors the thought processes of clinicians. This is because it helps them to quantify and track the levels of probability as these change from a *prior* status derived from a combination of their previous knowledge and experience, and from the initial history and physical examination to a revised *posterior* status as new information becomes available from the results of laboratory and other investigations and from the effects of treatment.

In the past there have been lively and occasionally acrimonious arguments between Bayesians and frequentists about statistical philosophies and methodologies with Bayesians gaining some ground in recent years in both in general and in medicine. Also, although frequentism has traditionally dominated the analysis of RCT's it has been shown that Bayesian methods can also be used.

SPC has a long and well proven track record in monitoring industrial processes to maintain production within specified limits in the quality of goods and services. It does so by distinguishing 'common cause' from 'special cause' variation through control chart methodologies developed in the 1920's by the Walter Shewhart and later promoted and taught by William Deming. Over the next half century or so these methods were enthusiastically adopted in many industries, but not so much in healthcare, except in pathology laboratories where it is used for quality control. This is despite the need to address the problems of variation in medicine or which SPC is well suited, and persuasive arguments in favour of following in Shewhart's footsteps for both generic and specific reasons.

Over the last couple of decades however there has been a slow but steady increase in the use of SPC as documented in a practical overview and a detailed metanalysis. Such reviews provide useful taxonomies of the wide variety of possible applications, but sometimes closer inspection of a single publication can be more instructive and memorable.

An important and often unjustly overlooked example is a study of the use of one form of SPC to show that the increased mortality rates in two very different groups of patients in the UK were due to special cause variation. In the first case the patients were all children who had undergone cardiac surgery in the Bristol Royal Infirmary (BRI), and in the second they were all patients of a single general practitioner. In first case after a lengthy and detailed commission

of inquiry costing around £15M it was concluded that outcomes for this type of surgery were substandard and that 30 deaths may have been preventable had the children been treated elsewhere.

In a second, after another even more extensive and expensive inquiry costing around £21M, it was concluded that at least 215 patients had been murdered by their GP Harold Shipman.

The importance of the above study by Spiegelhalter et al is that it clearly demonstrated (Fig1) that the use of SPC would have shown highly significant excess mortality many years before they became even more painfully obvious by other routes. At the BRI such analyses might also have strengthened the hand of the anaesthetist Stephen Bolsin who had tried in vain to draw attention to his well justified belief that the mortality rates were excessive.

This power of SPC in tracking performance has been best recognised and exploited, probably not surprisingly, by cardiac surgeons although for some their interest predated the BRI events. Also, the benefit of using SPC to track *lead* indicators such as the acquisition of clinician technical skills rather than the *lag* indicators of adverse events as the result a lack of such skills, has been demonstrated, again not surprisingly, by Bolsin and colleagues as well as by others.

The message then is clear, - SPC deserves to be more widely used in both the routine monitoring of clinical performance and in measuring outcomes and costs for VBHC

Activity based funding

This seemed a good idea at the time. The time was the late 1970's, the place was the USA, and the problem was the remorseless rise in the costs of healthcare especially in hospitals.[2] This problem was not unique to the USA but was bigger and growing faster than elsewhere, both in absolute amounts and as a proportion of GDP. An estimate of trends in this metric by US Congressional Budget Office in 2007 predicted that by 2082 healthcare could at least in theory consume all of the GDP.

The solution seemed logical enough - just change the funding model from the traditional practice of paying hospitals what they were paid last year with a top up for growth and inflation, to one determined by the volume and complexity of the services provided. This was and still is determined by the number of patients in various diagnostic groups (DRGs) adjusted for disease severity and with fees for specific services. Other countries soon followed suit with similar casemix classification and funding models including Australia.

This change did achieve one of its presumed objectives - a shift in control from the providers to purchasers of services but didn't seem to do much to slow down the rate of cost increases. It also left in its wake the unhappy side effect of encouraging activity in *outputs* in the form of volumes of services, rather than in the *outcomes* that patients seek, - resolution of their problems. As a result of these and other concerns, and as outlined below, several countries are now moving away from activity based funding and towards VBHC, a trend accelerated by a shift in policy in the USA.

The management aphorism of 'what gets measured gets managed' is true enough as far as it goes but equally true, especially in fee for service or DRG episodes of care funded healthcare systems, is 'what gets funded gets done.'

Rational ignorance

Another cause of the slow uptake of VBHC arises where costs are completely or mostly covered by state or private insurance is that this leaves both the patient and the clinician unaware of the full costs of the care provided and with no great motivation to find out.

This is known to economists and political scientists as 'rational ignorance,' - a state of mind where the energy that must be expended in the search for information is not justified by the rewards of enlightenment. This is in stark contrast with the situation in countries where patients carry all or a much higher proportion of healthcare costs which leads to very different perceptions of the meaning of value. This was illustrated in a survey in the USA where "Out of pocket costs affordability" was ranked higher by 45% patient than "My health improves" at 32%. These system-driven differences in patient perceptions of the meanings of value need to be accommodated in the definition and measurement of VBHC.

Why now?

The inflection point for the recent resurgence in interest in VBHC seems to have been the justifiably oft-cited book published in 2006 by Michael Porter and Elizabeth Teisberg that probably also played a part in the 2016 decision to shift USA Medicare payments towards a VBHC model. If so, it is because it provides a comprehensive account of the problems of the health care system in the USA; the reasons why various attempts to resolve them have failed; and what could and should be done about them. Some of the content of this book is only relevant to the USA, but many of the problems are global. Some of the factors are well known and well managed, others less so. They are best considered in two groups – concerns about costs and quality; and the consequences of rapid growth:

Concerns about costs and quality

The two main world-wide problems driving the move to VBHC are well known - costs that are uncomfortably high and rising, and quality of care outcomes that are often uncomfortably low. The quality issues have been neatly expressed as the '60:30:10' challenge - only around 60% of patients receive the care for which there is good evidence that they would benefit; 30% of costs are attributable to waste including the use of interventions for which there is little or no evidence of benefit; and 10 % of patients that suffer some sort of adverse event. For both costs and quality there are four important metrics that need to be accommodated in the design and implementation of any VBHC initiative:

Unjustified variations

There are significant unjustified variations from place to place in both costs and outcomes that have been well known since the pioneering work of John Wennberg half a century ago.[3] These have been comprehensively documented in atlases published in the USA, Australia, UK. These variations are widespread and often substantial but have been strangely neglected as obvious targets for any improvement process including VBHC [4]. If one hospital or clinical service, or clinician is getting better outcomes and/or getting them at a lower cost than others, then it is important know how this was achieved and if this expertise could be applied elsewhere. The savings in healthcare costs that would accrue just from replication elsewhere of the results delivered by the best performing services can be of the order of billions of dollars as evidenced by studies in both the USA and Australia

Low value clinical practice

Rather more attention seems to have been paid to the overuse of some types of healthcare considered as being of low value because of an insubstantial evidence base as defined in consensus initiatives such as the 'Choosing Wisely' program. Investigations of the frequency of such cases in both the USA and Australia. Studies in this area have shown modest decreases in the use of low value care in one study and rather more substantial increases in efficiency and decreases in costs in another.

The disadvantage of this approach is that it mostly just provides a top-down macroeconomic view of a limited set of the *processes* but not of the *outcomes* of care. It also presents physicians with a number of cognitive and practical challenges. As outlined below however, some microeconomic bottom-up systems of measuring outcomes and costs are now beginning to emerge that start at the level of groups of individual patients and clinicians. These are promising not least because they seem to be more enthusiastically endorsed and used by clinicians probably because results are more rapidly available and more obviously relevant to their daily clinical practice

Falling value

The evidence for this is both ancient and modern. The ancient data set even predates Florence Nightingale's criticism of hospital governance which might have been mellowed had she been able to cross the Atlantic to visit the Massachusetts General Hospital where, ever since 1821, they have been assiduously counting the daily death rates and costs. However, no one seems to have thought it of interest to plot these two variables together on the same time axis and make them known to the world until Meyers and colleagues did so in 2012. [5] The figure in this paper gives a clear and fascinating illustration of the trends in deaths and costs over these two centuries. The erratic mortality fluctuations in the 19th century were presumably due to recurrent epidemics, and the steady decline in the 20th century due in part to advances in medical science, and in part to the prevention of infections in the community and better management of the social determinants of ill health.

Equally clear from this graph however, and of relevance to drivers of VBHC initiatives, is the steep rise in costs over the last 70 years. This is mainly due to the added expenses of the medical advances that have changed the results of the value equation as defined by the death rate per unit cost. This can be demonstrated by roughly transcribing the numbers from these two curves and replotting them as a value index, a variable that clearly shows a steady decline over this period of observation.

In more recent times increasing life expectancy has been noted in several countries between 1960 and 2009 but predictably at the cost of a higher proportion of GDP. This allows a rough calculation of value defined as the life expectancy gained from all causes per unit of GDP, which has fallen. Similar patterns have been observed in the gains in life expectancy over a similar period attributable to medical interventions for which costs per unit gain have increased (Fig 1) i.e. value has fallen, especially in older age groups

Low or no outcome/cost correlations

A striking example of a lack of *any* correlation between costs and outcomes can be seen in a graph that plots the cost of care for similar sets of patients in some USA hospitals against the hospital standardised mortality (HSMR). The correlation coefficient was shown to be zero and the variation from lowest to highest values on both axes to be about 400%. This type of analysis does not seem to have been widely replicated possibly because of doubts about the

validity of the HSMR when used as a ranking tool, but the lack of any correlation between the deaths in any well-defined set of diagnostic categories and the relevant costs of care together with such gross variation between the best and the worst performances is disconcerting. Other studies have also failed to find significant positive relationships between regional USA Medicare spending on a few specified DRGs in the quality and outcomes of care or with patient satisfaction or in similar studies at a hospital level. Another USA study found a negative correlation between Medicare spending and quality of care, and noted a beneficial effect in regions with higher general practitioner to specialist ratios.

Widening perspectives and priorities

Kenneth Arrow is reckoned to be one of the founding fathers of health economics as evidenced by his much cited 1963 paper in which he explored the distortions of normal market forces by the peculiarities of the healthcare industry and stressed the high levels of uncertainty and the asymmetry between the information available to patients in comparison with the larger stocks held by physicians.

There is however another important asymmetry that he seemed to overlook: - the information held exclusively by the patient about the size and shape of their problems and about the effectiveness of the healthcare systems in resolving them. This is now being redressed by increasing measurement and reporting by patients of both the outcomes - (PROMS) and the experiences of their interactions with the healthcare system (PREMS). Obviously, no assessment of the value of health care can be complete without these two dimensions of patient opinion, but there are other important reasons for their use. These include evidence that there is often a mismatch between the importance attributed by doctors and patients to various aspects of disability in chronic disease, and that the use of PROMS increase the chances of discovering symptoms and adverse events that would otherwise go undetected. Also, despite some early variable levels of enthusiasm , doctors often come to welcome the extra information provided by PROMS in both improving patient care communications and in saving them time in consultations.

Arrow's other healthcare market distorting force, the high level of uncertainty, is a very real part of clinical practice and is increasing, paradoxically related in part to advances in medical knowledge as discussed below.

The consequences of rapid growth

The rapid increase over the last few decades in the number of diagnostic and therapeutic innovations and of the numbers of medical specialists and other staff needed to deliver these services has had several linear and non-linear consequences which are often difficult to understand and manage. Some of these are well known, others less so:

Cumulative risks of error

As care processes become more complex and multidisciplinary they usually become lengthier and multi-staged which increases the cumulative risks of errors and adverse events. This is

well known in industry where the risk of error rises in proportion to the complexity of the process and the time on the production line. So too in critical care units where in one study the chance of an adverse event rose by about 6% every day.

The important and often overlooked consequence of this inescapable arithmetic is that if we keep increasing the complexity of healthcare to gain the advantages of advances in biomedical science, then the reduction of the 10% to zero in the '60:30:10 challenge' will continue to elude us. This is because novel errors and misadventures will constantly emerge to replace or add to the list of the old ones that we have slowly painfully learned how to avoid. But as Amy Edmondson has persuasively advocated the pathogenesis of failure should be ranked at some point along a spectrum that extends from praiseworthy experimentation to blameworthy transgression. The vital metrics therefore include not just the total error rate but more importantly the ratio of misdemeanours to mishaps, of well-known to novel problems, and of the effective to ineffective organisational responses to all types of aberration.

Diminishing returns from increasing complexity

The rise and fall of ancient civilisations may seem far removed from the forces nudging the healthcare industry towards the measurement and management of value, but not so, as shown by the archaeologist and historian Joseph Tainter in his disconcerting book, or in a more concise journal article. He points out that an increase in complexity is a common human response and one that gives solutions that may work well in the short term, but a point of diminishing returns (Fig 1 journal article) is often reached whether on the grand scale of Minoan or Mayan civilisations or in the knowledge-rich modern industries of education, research and healthcare. The graph (Fig 3 journal article) of the progressive reduction in productivity in healthcare in the USA between 1930 and 1980 is particularly striking and is reminiscent of other evidence of declining value as noted above.

Increasing chances of interactions

The number of possible interactions among any entities including humans, rises up an increasingly steep curve. If 10 staff are involved in the care of a patient, they must manage not 10 inter-relationships and communication channels but up to 45. In the real world the numbers are much larger, often in the hundreds. One study of communication channels in a hospital shows (Fig 4 this report) just how complex these webs can be. So too with the risks of potentially harmful interactions when several drugs are given at the same time.

These types of interactivity can also generate complex adaptive systems that have unexpected emergent properties that can be especially difficult to understand and manage. This large and important topic has many implications for VBHC which are beyond the scope of this account, except as discussed below, in the matter of prioritising potential applications.

Bounded rationality

In 1969 the polymath Herbert Simon was awarded the Nobel prize in economics for arguing convincingly for what seemed in hindsight, as is so often the case with bright ideas, to be both simple and obvious. At that time the dominant 'classical' theory among economists was that humans decide how to acquire and use money and other resources through a rational process of finding and using *all* available and relevant information to enable them to arrive at an *optimal* decision. Simon argued very reasonably that humans cannot possibly either acquire or use the massive amounts of often uncertain information of uncertain relevance and must therefore compromise by making *satisfactory* rather than optimal decisions.

We are also all prone to a host of environmental, psychological, and social constraints which combine with structural and situational uncertainty to generate states of mind identified by Simon as 'bounded rationality'. This has given rise to a whole new science of behavioural economics, a field in which one of the leading investigators Daniel Kahneman also received a Nobel prize, an award in which he acknowledged the foundational nature of Simon's contributions. Kahneman has more recently produced a very readable popular science account of his lifetime's work in unravelling the different ways of thinking about issues and problems and how these can be distort our conclusions about the world around us. The significance of this for VBHC is that the rapid increase in medical knowledge has generated new subdivisions of medical specialists who have no option but to bound their rationality within ever deeper but narrower domains of expertise. So too in the consequential growth in related domains of management.

What does all this mean? – six recommendations

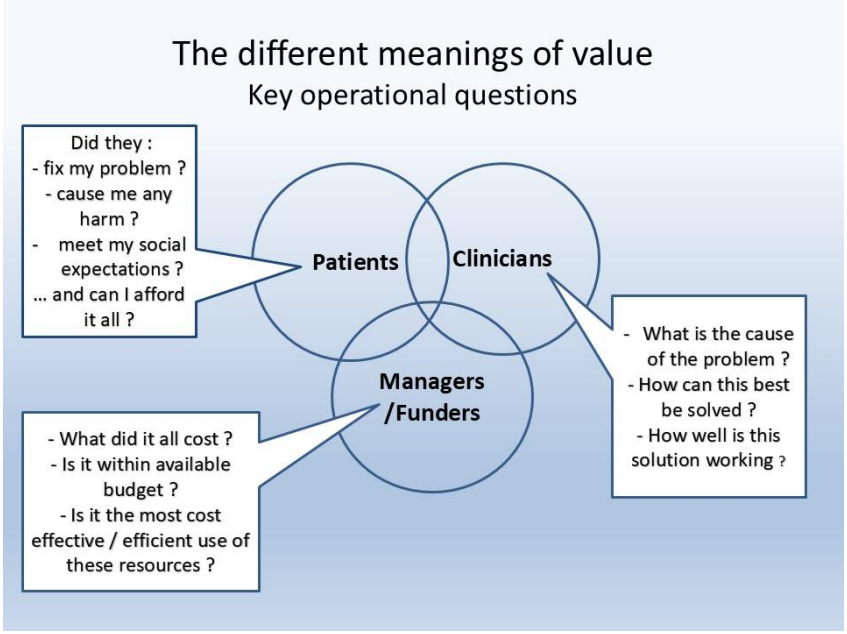
If the diverse strands of evidence and arguments presented above are accepted as relevant and important, then six recommendations emerge:

Ask the right people the right questions

The definition of value in any system of VBHC varies widely amongst all those involved:

> *'In any field, improving performance and accountability depends on having a shared goal that unites the interests and activities of all stakeholders. In health care, however, stakeholders have myriad, often conflicting goals, including access to services, profitability, high quality, cost containment, safety, convenience, patient-centeredness, and satisfaction. Lack of clarity about goals has led to divergent approaches, gaming of the system, and slow progress in performance improvement.*
>
> Porter ME 2010

These different and often conflicting interests and priorities especially those of three key stakeholders, - patients, clinicians, and managers/funders can however be crystallised in a few key operational questions:



This simple model makes it clear that the central task is to clarify and quantify the problems to be solved in a way that accommodates the different perceptions and often conflicting priorities of the three key stakeholders. In short, data collection and integration for VBHC should be problem oriented.

This is hardly a novel concept in medicine, but the best known precursor is the 'problem oriented medical record' as developed by Lawrence Weed and as encapsulated in the SOAP acronym. This was meant to systematise and better document the sequence of a physician's thought processes as they went from the patient's symptoms or Subjective information through Objective findings on examination or investigation and thence to the diagnosis or Assessment and culminating in treatment as part of a Plan. This was logical and useful as far as it went but gave no explicit recognition to the need to record outcomes.

Acronyms are useful as shorthand in glossaries, but more so if they capture conceptual as well as literal meaning as does SOAP, and best of all if they roll easily off the tongue and thus help spread an idea by word of mouth. The glossarists have given us VBHC which has little aural charm, but 'PROMs' and 'PREMs' should be joined by 'CROMs' and 'FROMs' This would focus attention on the need to integrate the measurement of the outcomes and experiences that matter most to the patient with those that must also be considered by, and that constrain the actions of, clinicians and funders.

Despite the impeccable logic of the SOAP sequence, the problem-oriented medical record as envisaged and promoted by Weed failed to gain long term traction for two reasons: - one obvious, the other less so, and both are relevant to the implementation of VBHC.

The obvious one was the requirement for data entry via complex, highly structured and time-consuming paper charts. These were probably intended to not only reinforce the concept, but to ease the path to computer data processing. Unfortunately, this idea was ahead of its time for the digital technology of the day. The technology has since advanced considerably but less so the ability of healthcare organisations to apply seemingly simple computerised solutions to complex human problems. VBHC requires substantial data processing support, but this needs to be carefully designed to answer the right questions.

The less obvious one was the implication, probably unintended, that the 'subjective' and 'objective' descriptors of the problem were intended to reflect the patient account and doctor assessments respectively. Given that the dictionary definition of subjective is: - '*influenced by or based on personal beliefs or feelings, rather than based on facts*', the consequent misinterpretation, especially by patients, was understandable.

There are indeed important differences between patient and doctor perceptions of the nature of the healthcare problem of mutual concern, but they related not to so much to subjectivity and objectivity, but to the priorities and practicalities that determine the ways in which solutions are found.

This has been well illustrated in a study of how the medical record is used not just as a journal of events but as a part of an active process by which doctors reformulate patients problems and make decisions to fit within the bounds of possibility. This task of reformulation increases in difficulty in proportion to the complexity [6] of the problems and this in turn increases the difficulty of measuring global value from its component parts:

	Low complexity	High complexity
Number of component parts	Few	Many
Levels of uncertainty about diagnosis and treatment	Low	High
Patient-doctor agreement about priorities and actions	High	Low
Responsibility for care	Single clinician /small team	Fragmented across many specialties
Risk of emergent properties	Low	High
Examples	Elective procedures Single bone fracture	Multi-system disease in the frail elderly

It is therefore advisable to start VBHC programs at the lower end of the problem complexity spectrum where the patient's and the doctor's perceptions of value are more likely to be concentric, solutions are better known, and success can be more readily distinguished from failure.

Rethink and reorganise the datasets

Wherever the problem sits on the complexity spectrum, finding answers to these questions is rarely easy as the key data sets in most hospitals are widely scattered. The challenge is therefore to reformulate and reorganise all these disconnected data sets for VBHC:

Dimension	Question	Sources	Format
Attribution	- What was the problem? - What was the diagnosis ? - Was this interpretation correct ?	P,C C,CM C	2 2,1 3
Intervention	- Investigations and treatments used ? - Were they appropriate?	C,CM C	2,1 3
Resolution	How well were the problems resolved?	P,C	2
Safety	Was any harm caused?	P,C	2
Quality	What were the patient experiences of care?	P	2
Cost	What where the relevant costs ?	F	1
Value	Did the outcomes/unit cost = best use of available funds?	N	3

Legend C:Clinician opinion, CM Casemix extracted data set P:Patient opinion, F:Financial systems, N:New integrated calculation needed 1: Routinely collected as a digitally accessible entity; 2 - Only accessible as text from chart review or from retrospective sample surveys. 3 - Not asked routinely or only collected in response to 'forensic' inquiry.

Start with the low hanging fruit

Organisation-wide ventures often advance on too broad a front with the risk that early encounters with difficult instances blunts enthusiasm and stalls progress. It is therefore best to start with the easier opportunities of which there is an ample supply of candidates identifiable by four characteristics:

Low complexity

Early targets should include elective surgery or other established interventions where both the nature of the problem and the usual solutions are well known and there is more likely to be agreement about measurable outcomes.

High variability

Large variations in outcomes and costs in different clinical services treating the same disease processes in similar patients offer the highest potential gains.

Amenability to intervention

Failure to use evidence based best practice presents another obvious starting point. This includes 'amenable mortality' which has been shown to be useful indicator of the performance of healthcare systems in international comparisons.

Selection of cases of known amenability to intervention also avoids unfair attribution of outcomes to clinicians over which they may have no control:

> *'If the person or organisation whose performance is being measured feels powerless to influence the indicator, inappropriate measurement can also lead to demotivation, dysfunction, and crisis.*
>
> Pringle et al 2002

Condition-specificity

Patient reported outcome measures (PROMS) can be either generic or condition- specific. Both are important but the condition specific variables can be more readily assessed and interpreted by both patients and doctors, especially in low complexity elective procedures. The three categories of generic indicators- physical, social and mental are admirably comprehensive and are useful for broad long term population 'wholesale '; studies, but less so for providing more immediate assessment and management of the individual problems of 'retail' clinical practice.

If the appropriate microeconomic measurements are made and appropriate actions taken, then macroeconomic benefits would be expected to follow.

Measure processes as well as outcomes

Porter and Teisberg rightly placed a high priority on measuring results as essential in VBHC, and had less enthusiasm for the processes of care but both need to be measured together because:

- It is essential to know whether evidence based best practice has been used in the assessment of outcomes and consequent management.

- If the outcome numerator is small in comparison with the denominator, as it often is, for instance, in mortality rates, then the *lead* indicators of failure to deliver evidence based best practice will be detectable long before there will be enough *lag* indicators of death rates to indicate a problem. This has been exemplified in the management of myocardial infarction.

Provide fast and relevant feedback

Summary data presented months after the event may be all that is needed by board level executives and strategic planners but is of little use to clinicians and patients to whom what is happening day by day and sometimes hour by hour is often of crucial importance. Summarisation, by definition, must always discard some information, but this may be the very information that matters most in the hurly burly of clinical practice. Information systems are therefore needed that display key datasets of costs and outcomes as simple graphics as soon as they are generated.

VBHC reporting systems need to display cost-outcome matrices disaggregated down to a patient and provider level. As discussed below this is has been well demonstrated, in systems developed in the University of Utah, and elsewhere such as in the 4 quadrant diagrams used by Stowell et al to compare costs and outcomes and radar charts have also been used as for similar purposes in assessing various interventions for prostate cancer.

Two other variables need to be included: - time trends and effect size. The first can be accommodated by statistical process control analysis for all the reasons discussed above, and the second using graphical displays developed for other purposes as discussed below.

Follow the leaders

There are many factors that help or hinder the implementation of VBHC, but examples of what has worked well for several years in routine practice are often more informative.

Four are noteworthy in this regard, either because of the development and routine use of disease-specific patient reported outcomes, or because of the development of organisation wide systems for measuring and improving various aspects of value. They all describe programs that have taken over a decade to build and refine:

- *UK NHS.* Since 2009 patients undergoing hip and knee replacement in the UK NHS have recorded the severity of their symptoms before and six weeks after surgery and this has generated a large trove of data about the effectiveness of these operations and the incidence of adverse events. This program has provided valuable data for the purpose of comparing the outcomes of different operations in different services.

- *Sloan-Kettering New York*. Patient undergoing outpatient chemotherapy at Sloan-Kettering in New York have been shown to be willing and able to collect PROM data online using an ITM system originally developed in 2003. A subsequent controlled trial of this methodology has shown improved effectiveness and efficiency of clinical practice monitored and managed with the aid of this type of technology.

- *University of Utah*. The "Value Driven Outcomes" program of the University of Utah Health Sciences Center, started in 2012, is particularly instructive because of:

    - The development of an integrated ITM platform that provides fast and relevant feedback of outcomes and costs to clinicians and other staff.

    - The demonstration of improved effectiveness and efficiency in a wide range of disorders. This work was understandably lauded in an accompanying editorial by Porter and Lee, especially the development of the 'opportunity index' as a means of selecting target areas, and the progress made towards an effective system of cost accounting.

- *NYU Lagone Health (NYULH)* This health service started an organization-wide VBHC program in 2014 and more recently reported significant downward trends in costs in several medical and surgical DRGs due to various process efficiencies. This program shares several of the hallmarks of integrated senior management, academic and clinician leadership with those in Utah. It is also notable for the development of some innovative cost accounting systems that measure both the actual variable costs and the impact on income streams and consequent clinical interventions which are significant factors in private healthcare services[7].

It might therefore have been expected that these results especially those from Utah and NYULH would have been more widely adopted or adapted. The main Utah publication has indeed attracted, as of early 2025, over 300 citations, but most (except the one from NYULH) seem to have only cited it in reviews of VBHC, or in narrower specialist applications.

In only one other instance, in Singapore, was there an attempt to replicate an health-service wide implementation. The NYULH publication has only been cited 20 times with a similar pattern of content seen in the Utah citations- review more often than replication.

This is lack of spread of these innovations is surprising, but as Sherlock Holmes famously noted in 'the case of the dog that didn't bark', something that doesn't happen can be as informative as something that does.

Where to from here?

The simple answer to this question would seem to be to just follow the six recommendations listed above, but the rarity of adoption of the Utah model of VBHC, which incorporates the first five of these suggests that they may be necessary but not sufficient. It is therefore worth exploring the causes of this gap to see how it might be reduced.

The necessity-sufficiency gap

There are a few clues about the probable origins of the Utah accomplishments in the listed contributions of the authors and authorship order in the ITM infrastructure and results papers. These suggest a productive collaboration among a few key clinical, executive, and ITM leaders which lasted for at least a decade. There are also hints in other publications by these authors: - on the methods of optimising the patient experience as a foundation stone of a value driven outcomes program; on the determinants of success in the implementation of computer decision support systems; and on the differences between patient and clinician perspectives of outcome priorities.

The main message therefore seems to be that organisation-wide VBHC needs a long-term resilient working partnership among:

- *Clinicians and patients* who can specify their individual and shared views of the problems to be solved and the outcomes desired.

- *Senior executives* who can build consensus about a shared meaning of value as outcomes per unit cost, and who can support and drive the integration and unification of service level initiatives into an organisation-wide function.

- *ITM specialists* who can translate the needs of clinicians and executives into digital systems that provide concise and timely reports that link outcome with relevant costs.

This is hardly a novel idea as such partnerships are regularly developed for many complex managerial tasks in healthcare such as the control of hospital acquired infections, or the reduction in amenable mortality.

VBHC does however have additional challenges however that make these coalitions of trust and competence both more important and more difficult to build:

- *Contested meanings*. It seems unlikely that there would be much dispute about whether a patient does or does not have an hospital acquired infection, or is alive or dead from a preventable cause, but as outlined above, the meaning of healthcare value is, like beauty, very much in eye of the beholder. This task of unification of these diverse meanings brings to mind the parable of the blind men and the elephant.

- *Executive capacity* In any organisational venture of the magnitude and impact of VBHC, it falls to senior executives, usually the CEO, to find answers for the strategic questions that arise from service level questions and operations.

  This challenging long-term task is made more difficult by the high turnover in these positions that has been noted in both Australia and the USA. These departures are apparently involuntary in up to a third of cases, with such dismissals commonly due to blame for poor cost control or to major failures in patient outcomes.

  It is therefore rather odd that a managerial resource such as VBHC that could clearly help both CEO's and their employers to make the links between costs and outcomes measurably explicit and thereby more controllable has not been more widely adopted. It may be that this due to the incompetence of up to a third of CEOs, but this seems highly improbable. A more likely explanation is the common human tendency to make fundamental attribution errors, - to mistakenly allocate blame for problems to the 'dispositional' limitations of individuals rather than to the 'situational' constraints of the environment or system in which these individuals are enmeshed.

- *The casemix factor.* One such situational constraint that is often overlooked is the distortion of casemix costing and funding. This is important because value assessment must apply across the whole sequence of clinical problem solving:

  Problems → Investigations → Diagnoses→ Treatments→ Outcome

  Unfortunately, the casemix juggernaut starts in the middle, focuses attention on the volumes and costs of diagnostic and interventional entities and keeps it there. It is also blind to the specific problems to be solved and the outcomes to be delivered which are the essential elements of VBHC – or indeed of any commercial or public service enterprise. If car mechanics were to be judged by the number of worn cylinder head gaskets or radiators that they had diagnosed as the source of the problems, rather than by how well they had stopped the oil or water leaks that were bothering their customers, it seem unlikely that they would stay in business for long. Health services need to be judged on similar grounds.

  It is therefore not surprising that if the performance of executives is judged mainly on the 'metrics of the middle', they do not fare well because the urgent then obscures the important as it so often does in healthcare. If for instance there are a few cost over-runs or if bad clinical outcomes become a public scandal, heads soon roll, and the search starts yet again for a heroic / transformational leader to chart a new course. Visionary leaders are indeed needed, but for VBHC, advanced transactional skills are also required to build systems and coalitions of the willing and able. And time.

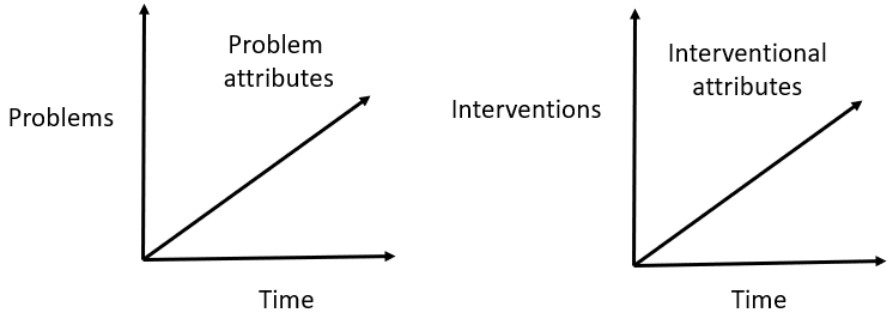
Attention therefore needs to be redirected from the deviance to the exploratory testing end of Edmondsons' spectrum of failure pathogenesis. A prime candidate for this exploration is the means of attributing costs to outcomes, a problem for which the Utah group seems to have found solutions, but which has eluded the architects of 9 out of ten of recent European VBHC programs [8], all of which have concentrated on measuring and improving outcomes. This is both essential and admirable, but it is only half of the challenge of implementing VBHC.

New opportunities

As the well-known tale would have it, a traveller lost in a foreign land seeks advice from a local citizen only to be rather unhelpfully informed: '*If that's where you want to go, I wouldn't start from here'*. Likewise, in VBHC we have no choice but to start from where we are: - somewhere in the middle of the casemix country, a wide and rugged landscape of diagnostic and interventional categories with few signposts pointing in the direction of value measurement and management. These are needed to show routes away from the metrics of the middle – one *backwards* towards problem definition, the other *forwards* towards analysis of outcomes as judged by patients, clinicians and funders.

Problem definition and linkage

The need to define and classify patients' problems was recognised over half a century ago but this ambition doesn't seemed to have attracted much interest beyond a comparison of various classifications in general practice. This is in stark contrast with the energy devoted to the detailed classification of diseases, diagnoses, pathologies and interventions in the large and still growing taxonomies such as the International Classification of Disease and the Systematized Nomenclature of Medicine. This is not surprising given that the intellectual and reputational foundation of medicine rests in large part upon the ability to diagnose, understand the causes, investigate, and treat various diseases. This has however resulted in diagnostic taxonomies that were developed primarily for medical science and practice being adopted as accounting entities and being used, often inappropriately, as performance indicators. The need in VBHC is not for an 'international classification of problems', but for a simple way of categorising problems and linking them via diagnostic and therapeutic interventions to outcomes and costs based upon the primary dimensions of clinical practice:



'Problem attributes' in this context includes all relevant clinical and epidemiological factors such as age, sex, occupation, past medical, family and social histories, while 'Interventional attributes' includes all managerial details of the providers, investigations, treatments and costs. In short, what should be done matched with what was done and all tracked over time.

The difficulty of course is the vast number and the diversity of all such problem and interventional entities, - a catalogue that expands with every advance in medical science and where the clinical context of the patient's presenting problem is critical in assessing the value of subsequent interventions and outcomes.

Among many other equally worthy candidates, gastrointestinal endoscopy provides a good example of this requirement. The indications for medical attention in this domain can thus include the screening and surveillance of asymptomatic patients with familial or other risks of malignancy, the exclusion of unlikely diseases in the worried well, and interventional procedures to stop gastrointestinal haemorrhage, or to remove pre-malignant polyps. Only when all these very different causes of 'the problem' have been specified can the relevant outcomes and costs be meaningfully linked and quantified.

Analysis and display

Processing all this complexity might seem to be an insuperable challenge for most humans unaccustomed to thinking beyond the three familiar dimensions of the physical world around them. Fortunately, this does not inhibit mathematicians or computer scientists who have no difficulty in conceiving of multi-dimensional collections of data nor in using them as the foundation of the recent dramatic advances in machine learning. [9]

In healthcare, these advances have been most obvious in medical imaging and more recently in using natural language processing to analyse the vast resources of previously unclassified text now routinely available in the electronic health record. One notable study has shown that such automated analysis of this very large 'bag of words' can accurately predict diagnosis, mortality, length of stay and readmission rates as shown by high AUC-ROC scores – including an impressive figure of 0.90 for the unwieldy categorical variable of diagnosis. Similar levels of diagnostic prediction have been confirmed elsewhere.

The potential of these machine learning techniques in health services in general and in VBHC in particular is therefore both and obvious and substantial, but the challenge is to clarify the significance of all the connections within these new constellations of data.

Fortunately, a few conceptual and practical frameworks are available from other disciplines that can be adapted for VBHC, and that can be augmented by machine learning to assist in the analysis and display of key information in multiple dimensions:

- *2-D maps.* Strange as it may seem, a mapping technique developed to track the flow of heat in steam engines by the engineer Mathew Sankey in the late 1890's, provides a starting point for this task. The charts that carry his name are still in use today, mostly to track energy production and consumption, but they have more recently attracted the attention of clinicians.

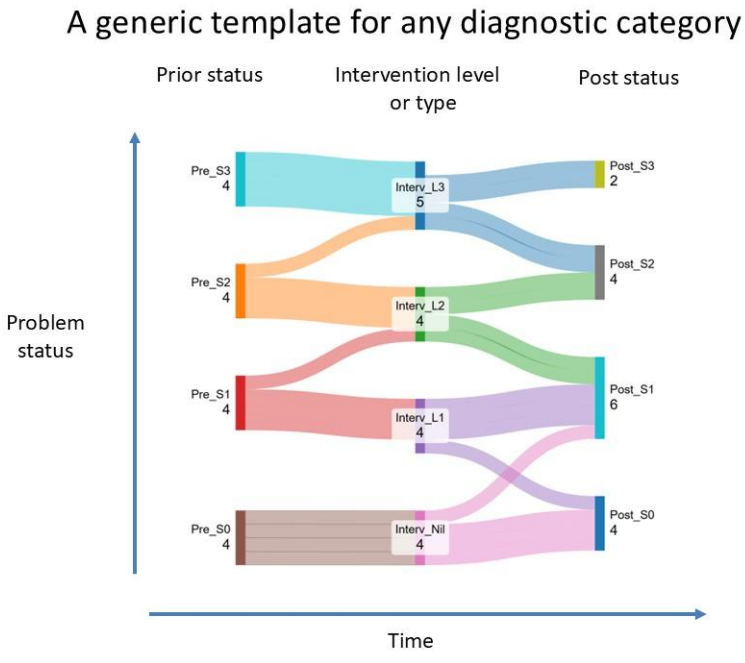
Or perhaps it's not so strange. In essence, the fundamental tasks of both engineers and clinicians are similar in that both often have to solve complex stock and flow problems of a type that is known to challenge even the brightest and best educated.[10] Sankey charts help by displaying and quantifying the levels of stocks and the rates of flow of energy and information along interconnecting channels to show the sources and destinations of both wanted and unwanted outcomes.

In steam engines the wanted outcome is the maximum conversion of the latent heat of coal into steam and thence into mechanical muscle. The unwanted outcome is any wastage of heat or steam along the way. Sankey charts show the measured sources, channels and destinations of both productive and unproductive heat.

In clinical practice the wanted outcome is the best solution for the patient's problem. Unwanted outcomes include the wastage of any unnecessary use of clinical or other resources, or the imposition of new problems in the form of adverse events. Here too Sankey charts can show the measured sources, channels and destinations of both productive and unproductive clinical activities.

A common temptation in using Sankey charts to map complex processes is to include more nodes and links than are needed. This may give an impressive display of the level of complexity to be understood and managed, but the resultant rat's nest tangle often obscures rather than clarifies the key components and relationships.

Three recent clinical applications in patients with cancer, complex abdominal surgery, and severe trauma have resisted this temptation by limiting their attention to three key elements: – *common well defined diagnostic categories; established clinical pathways; and clear and measurable outcome indicators.* These elements can also provide the foundation of a generic template:



Even in this very simple and entirely imaginary sketch map [11] several potential applications and extensions are apparent:

- Existing numerical problem severity scales such as the Oxford Hip Score or the Crohn's Disease Activity Index can be used to incorporate both patient and clinician derived indicators to assess the impact of interventions.

- Interventions can also be graded according to level of intensity and can refer to either a single event such as a hip replacement, or to a clinical pathway comprising a sequence of activities, or to 'bundles of care'.

- Although the focus in this paper is on VBHC in hospitals, the time span of the model can be extended to show the whole trajectory of any disease management process in the community before and after the hospital episode of care. This is in keeping a whole cycle of care approach as advocated by Porter and Teisberg.

- The potential of combining machine learning with such mapping techniques is clear as was shown in one of the above studies that used an early variety of this technology to provide adjustable filters for subsets of clinical problems and investigational indicators.

- The possibility also occurs of emulating the machine learning capabilities of satellite navigation systems such as Google maps to highlight which managerial routes to take or to avoid in the clinical problem-solving journey. And whether to take the roundabout scenic academic route or the duller but faster operational freeway.

- *2-D cost-benefit curves.* If VBHC systems designers can be persuaded to initially limit their ambitions to the key elements specified above (common well-defined diagnostic categories; established clinical pathways; and measurable clinical outcome indicators), then the task of measuring and attributing costs by demanding processes such as time driven activity-based costing becomes more manageable. Discrete sets of outcomes (o) and costs (c) can then be quantified more readily, and value (v) calculated either individually or cumulatively from v=o/c. This indicator can then be plotted against other important variables such as the ever-increasing complexity of health care interventions.

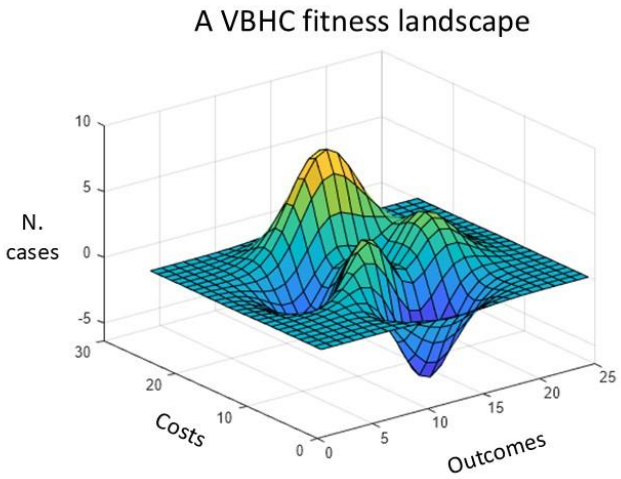
  As Tainter has pointed out, investment in complexity sooner or later always reaches a point of diminishing returns. In the current context this can be illustrated by simply changing the names of the axes in his graph (Fig 1) to reflect the similar impact of complexity on value.

  This configuration also raises the interesting possibility of the development of a 'calculus of value' in which integrals would represent the area under the curve of cumulative value and derivatives the instantaneous rate of change of its slope.

- *3-D landscapes and beyond* The simple 3-D configurations of clinical practice as defined above also needs to be related to case volumes for both operational and ethical reasons.[12]

  A useful theoretical framework for this purpose is the 'fitness landscape', a concept originally devised as a means of representing the reproductive capacity of different genotypes but more recently adapted for use in the social sciences

  In the format as originally conceived the x and y axes represent the competing genotypes and the z axis reproductive fitness. For present purposes the x and y axes can be assigned to the competing elements of cost and outcomes and the z axis to a 'reproductive fitness' dimension measured in case volumes of a given individual or service level provider.[13] As in the genotype application, fitness can be normalized to zero so that negative values can be shown below some global baseline chosen to aid comparisons among providers. An imagined scenario of this type is shown below in a configuration reformatted from a standard surface plot program example:



37

Reproductive fitness whether biological or organisational cannot be judged just by a single success, but by whether it can be sustained. In VBHC as elsewhere outcomes and costs can change in any one of nine different combinations, some clearly more desirable than others. All variables therefore need to be monitored for any changes, otherwise significant benefits or harm may go unnoticed for years. [14]

- n-D implementations in practice. These conceptual models may give some idea of how various combinations of the 8 key data elements of clinical problem solving might reconfigured in theory, but in practice they and all their subcategories need to be integrated in a single and simple system for analysis and display.

  This may seem a tall order but such a system has been produced for very large data sets of global economic, social, health, climate change and other important indicators. This was originally developed, and the need for such a system persuasively demonstrated, by Hans Rosling, with a particular emphasis upon the need to correct widespread ignorance in both general and expert groups about many known facts about these global problems. For present purposes a valuable feature of this system is that it is freely available in the public domain for interactive explorations, and comes with advice about how it can be adapted for user-supplied data sets supplied in familiar spreadsheet formats.

  In a recent very limited assessment, this system proved easy enough to populate with a minimal hypothetical data set of the cumulative costs and outcomes generated in the management of some diagnostic group by five providers whose caseloads increased at similar rates over five years. A few of the possible different performance relationships were set to a normalized starting point of 100 with this result .

  The purpose here is not to imply that this system is suitable for large scale VBHC in the real world. Far from it. It is just to show that the underlying simple and sparse data model is generic enough to be able to accommodate both the key data elements needed to measure and manage value, and the output of machine learning processing of electronic health records as subcategories of these elements.

  In short, it provides a useful 'sandpit' for clinicians and managers to explore the interactions among the key data elements of clinical problem solving, and to help build a business case for full development of VBHC information systems.

Conclusions

> *"I shall take the simple-minded view that a theory is just a model of the universe, or a restricted part of it, and a set of rules that relate quantities in the model to observations that we make. It exists only in our minds and does not have any other reality …A theory is a good theory if it satisfies two requirements: It must accurately describe a large class of observations on the basis of a model that contains only a few arbitrary elements, and it must make definite predictions about the results of future observations*
>
> *Stephen Hawking* [15]

Given that we are still not quite sure how to measure and manage health service value 160 years or so after Florence Nightingale rightly and vigorously argued the case, it is probably not surprising that finding answers to the questions listed on the first page proved to be a lengthier and more complex undertaking than estimated at the outset.

My only ambition at the outset was simply to provide an imagined readership of clinicians, managers, ITM engineers and perhaps a few patients, with a short and comprehensible guide to the vast and diverse scientific literature relevant to VBHC.

It is for others to judge whether it is comprehensible, but the aim for brevity was clearly wide of the mark. This is because it soon emerged, and took many words to make the case, that the key components and cognitive processes of clinical problem solving are no different from any other sort however unlikely this may seem at first sight.

Thus whether the problem is keeping wounded soldiers alive on the Crimean battlefield; making steam engines more efficient; monitoring trends in hospital mortality or in global health indicators; or working out why some genotypes assist reproductive success more than others, the key dimensions are always the same; the key components of clinical problem solving can readily be generalised; and the stages of clinical and generic problem solving sequences are also very similar.[16] The other common characteristic is how often well proven solutions in so many different fields have so often been overlooked or ignored.

It may seem pretentious to borrow the words of a renowned astrophysicist as an epigraph for these rather mundane and probably unsurprising conclusions**,** but the frequency of the core elements and their similar interconnections in so many diverse fields suggest that the proposed framework could indeed accommodate a large class of observations. As to predictions, it seems likely that this model, augmented by the power of machine learning systems, would be useful for problem solving in domains well beyond value-based healthcare. It might also make demonstrably effective solutions less frequently overlooked or ignored by making the evidence unignorable.

M. Ward
22/01/25

Notes

[1] As is generally well known Florence Nightingale recruited, trained, and led a small band of women in the military hospitals treating soldiers wounded in the 19th century Crimean War. This pioneering enterprise subsequently became the model of what professional nursing should look like. She achieved this through her diplomacy and lobbying skills in speaking truth to power, whether in persuading the secretary for war Sir Sydney Herbert back in London to supply the essential material resources, or in nudging sceptical or over-burdened medical officers in the Crimea towards more effective methods of infection control.

Less well known is that she substantially enhanced the power of her leadership and interpersonal skills with innovative analytical methods that made her arguments difficult if not impossible to ignore. This can clearly be seen in the 'polar' or 'coxcomb diagram' that she devised and used to summarise the causes of death of soldiers in the Crimea.

In its original format it is not easy to interpret, but by roughly converting the areas on the diagram into numbers and replotting them in a more familiar and modern style of a linear histogram, the message becomes very clear. This is simply that the death rates from 'zymotic diseases' – infections that were both dangerous and easily transmissible such as typhus, typhoid, and cholera, were many times higher than that from battle wounds and that she could dramatically reduce this mortality by insistence on simple hygienic precautions and other infection control measures.

On the strength of this and other analytical accomplishments she was to become the first female member of the Royal Statistical Society. This recognition would have had the support of the statistician William Farr, one of the founding fathers of epidemiology and a long term colleague adviser and admirer of her work. The 'lady with the lamp' thus not only illuminated and comforted her patients by night but also shed much needed scientific light by day on the outcomes of clinical practices and how to improve them.

After she returned from the Crimea, she set to work to introduce into UK civilian hospitals the same changes that had she had used to reduce the mortality on the battlefield. To this end she clearly understood the epidemiological significance and practical importance of measuring and managing unjustified variations:

> *'These methods if generally used would enable us to ascertain the relative mortality of different hospitals as well of different diseases at the same and different ages, the relative frequency of different diseases and injuries, among the classes that enter hospitals in different countries and in different districts of the same country.'*
>
> *Florence Nightingale: Notes on Hospitals 1863*

In short, she identified the key principles of the process of improvement in any organisation and implemented them in practice well before Walter Shewart and William Deming formalised them in the PDSA cycle in the next century.

It could well be argued that her contributions were as important in reshaping clinical practice by controlling infections in hospitals, as those of the more famous actions of John Snow in reshaping public health by controlling an outbreak of cholera by removing the handle of Broad Street pump.

[2] This would have come as no surprise to the economist William Baumol who introduced the concept of 'cost disease' in the 1960's to explain why the costs of employing people in some industries such healthcare, education and the performing arts increase much more rapidly than in others such as the production of cars, clothes, and computers. He attributed this to the disparity in labour productivity possibilities between these two groups as demonstrable in data showing that between 1998 and 2018 for example, costs in health care and education rose by about 200% while those in the production of consumer goods dropped by 100% .

He famously first illustrated this problem with the limited opportunities for productivity increases by a string quartet and later applied a similar analysis to the costs of medical care. One difference of course, is that to play a Mozart quartet today still needs only four players as it did in Mozart's day whereas to investigate and treat many diseases now requires a large and still growing number of specialists.

[3] Although as Wennberg points out due recognition should be given an earlier pioneer in the study of unjustified variation, Dr J. Allison Glover, a medical officer in the Ministry of Health in the UK, who in the 1930's noted that the wide variations in the chances of a child undergoing tonsillectomy, and who showed that this was not due to any clinical differences, but to the school they attended, geographic proxy indicator for the operative enthusiasm of the local surgeons.

[4] This odd reluctance in the medical profession has a long history as it was experienced by Wennberg when he first tried to get his findings published in leading medical journals in the 1970's:

'*Naturally, this conclusion did not sit well with our fellow physicians. We published in Science only after being turned down by medical journals with wide clinical readerships, such as the New England Journal of Medicine and the Journal of the American Medical Association. Editors rejected our paper on the assumption that patient demand simply had to be the explanation for our observations, and thus the findings would be of no interest to their readers. But the sheer magnitude of the variation in incidence of hospitalization and surgery among these neighbouring medical communities suggested that patient demand could not be the sole cause. And that suggested the importance of physician behavior as a major source of variation.'*

It must have been some comfort for him to see that when was finally accepted it was by the prestigious journal Science and has since been frequently cited. Nonetheless, opportunities to explore and reduce unjustified variations are still underutilised in clinical practice.

[5] It is rather ironic that it was also the Massachusetts General from which Ernest Codman resigned. Even If he knew that this information was being collected however, he would probably have been annoyed that the only 'end result' that seemed to be of interest was the final one, and not any of the precursor states such as complication rates and other outcomes that might have helped find ways to reduce mortality.

[6] The term 'complexity' is used here as it is in system dynamics or organisation sciences and is distinguished from 'complicated' by the larger number, diversity and interactivity of the component parts and by the lack of predictability of outcomes from a given starting point.

[7] '(*The value-based management*) *team used variable direct cost data from NYULH's customised activity-based costing system … for cost accounting… This customised system includes several unique features that facilitated identification of opportunities for improved value. First, it is based on actual costs (ie, acquisition costs of drugs and supplies) rather than the ratio of cost to charge, thus eliminating influence of charges and insurance contracts. Second, it incorporates both inpatient physician and hospital billing so that collective impact of clinical care could be assessed: for instance, the Task Force could assess whether certain types of consults or care by certain types of physicians affected outcomes. Third, it includes patient level quality data such as expected mortality, readmissions and length of stay (LOS); and hospital-acquired conditions, facilitating a focus on value.'* Chatfield et al 2019

[8] '*Costs remain largely a black box for payers, and a blind spot for hospital managers. As the denominator of the value ratio, cost is difficult to measure for several reasons. First, most hospital cost accounting systems are department-, not patient-based, and are designed for billing of transactions reimbursed under fee for-service contracts. In most health care organisations, there is virtually no accurate information on the cost of the full cycle of care for a patient for a particular medical condition. As a result, cost allocations are often based on charges, not actual costs. Second, most providers are reluctant to share cost information in order to ensure that their net profit margin remains confidential, particularly since this information could weaken their negotiating stance with payers*.' Implementing Value-Based Health Care in Europe 2020

[9] Any further exploration of the sciences of these multi-dimensional arrays or 'tensors' is well beyond the scope of this review and the competence of the author, but a useful glimpse of a few of the concepts involved can be found in a series of brief and simplified explanations

[10] This was well demonstrated by the systems dynamics scientist John Sterman who showed that even MIT engineering post graduates often failed to solve simple stock and flow problems or wrongly declared them insoluble. He also pointed out the serious impact of such errors in the understanding and management of climate change and other major environmental and economic problems.

[11] For those who may wish to experiment with their own data in Sankey charts there are several free online facilities including the one used for this example. Another is part of a more comprehensive set of graphic charting software.

[12] The ethical reasons for paying attention to case volumes are reflected in the different frequency distributions of health care costs and outcomes.

The distributions for costs are markedly skewed with an estimated 1% of patients in the USA accounting for 22% of funding destinations, raising questions of equity in countries where wealth distribution is similarly skewed. Those for many health outcomes on the other hand are more often of a normal distribution that raises two more considerations – one mathematical, the other epidemiological.

The mathematical one, often overlooked in service improvement strategies, is that in normal distributions 70% of any population always lies within one standard deviation either side of the mean where the same level of improvement has a much larger impact than in the 2 % of outliers where

performances are less than two standard deviations below the mean but where attention is more often paid.

The epidemiological one can be seen in any report of service delivery or outcome variations related to local social circumstances. In the Australian Atlas of Healthcare Variation for instance, the incidence of lower limb amputations for diabetic complications is much higher in remote outback areas, probably because of suboptimal treatment of diabetes in these parts of the country. This atlas also shows many examples of wide variations in the other direction where the chances of patients undergoing some elective procedures and operations can be as much as 20 times higher in urban than rural-remote areas. This is most often due to the supply-sensitive influences described by John Wennberg.

13 The similarity between genetic and organizational evolution resides in the common need to balance the costs and benefits of the alternate strategies of exploration versus exploitation. In both types of fitness landscape, the temptation to move from suboptimal to optimal peaks always carries some risk of getting trapped in an even lower valley along the way.

14 The usually cited example of an overlooked potential benefit was the delay in the use of lemon juice to prevent scurvy by the British Royal Navy for some 50 years after the demonstration of its effectiveness by one of its medical officers James Lind.

A more recent serious example of failure to track harm was the demonstration by Spiegelhalter et al that the use of a simple method of statistical process control would have shown alarming time trends in mortality rates in two very different healthcare environments more than a decade before they came to light by other routes, and only after many more preventable deaths had occurred. It is therefore both surprising and disappointing that after a further two decades this important paper has only been cited a couple of hundred times or so, and that the power of statistical process control to show trends in important variables whether positive or negative has been slow to gain wider recognition.

15 It seems unduly modest for anyone with Hawking's remarkable intellectual accomplishments to describe any of his own ideas as 'simple minded,' especially as another eminent mathematician and scientist has had similar thoughts:

> *'The progress of science consists in observing interconnections and in showing with a patient ingenuity that the events of this ever-shifting world are but examples of a few general relations, called laws. To see what is general in what is particular, and what is permanent in what is transitory, is the aim of scientific thought.'*
>
> *Alfred North Whitehead*

16 One small but important detail of these generalizations is to acknowledge that there are dimensions of the costs that are other than financial. These may be less easily quantified but we ignore them at our peril as the tragedy of human responses to climate change so clearly shows.