PIRU POLICY INNOVATION RESEARCH UNIT

Developing a Global Healthcare Innovation Index

Tiago Cravo Oliveira, Eliana Barrenho, Antoine Vernet
Erkko Autio and James Barlow

PIRU Publication
2017-20

For further details, please contact:

James Barlow
Imperial College London
South Kensington Campus
London SW7 2AZ
Email: j.barlow@imperial.ac.uk
www.piru.ac.uk

POLICY INNOVATION RESEARCH UNIT

Developing a Global Healthcare Innovation Index

Tiago Cravo Oliveira, Eliana Barrenho, Antoine Vernet
Erkko Autio and James Barlow

February 2017

Acknowledgments

This work was funded by the Policy Research Programme of the Department of Health for England, via its core support for the Policy Innovation Research Unit. This is an independent report commissioned and funded by the Department of Health. The views expressed are not necessarily those of the Department or its partners.

We are grateful to Nicholas Mays and Mark Petticrew for their helpful comments on an earlier version of this report.

Contents

1. Introduction

Our understanding of medicine is being revolutionised by the pace of science. But not all the potential innovations in life sciences and medical technology are taken up into everyday practice in healthcare, even when they are shown to be beneficial.

For the poorest people in the world, many innovations are not accessible because they are either unaffordable or unsuitable for their health systems. Tackling this gap requires the development of appropriate and affordable health technologies and novel business models.

The health innovation challenge in the UK

In the UK, the NHS Five Year Forward View (NHS, 2014) has set out the goals which need to be met to address the challenges faced by the National Health Service. Achieving them requires intelligent implementation of new organisational and financial models of care. Some of these involve 'innovation' – not only in the form of wholly novel ideas but also what has been demonstrated to work elsewhere, whether they involve organisational or process innovations, new technologies or a combination of all these. The health innovation challenge in the UK is not just about the take-up and spread of new ideas. Ensuring that the country's medical devices and life science industries remain innovative and competitive, and continue to supply the technologies that help to deliver high quality healthcare is also regarded by government as essential.

In the more advanced health systems there is a disconnection between the effort on research and development (R&D) and how much of this makes it into mainstream healthcare practice. Even the most evidence-based and affordable innovations can fail or are only taken up patchily, whether we compare across countries, or between localities or health organisations within countries. And technological innovation can be a problem for those responsible for paying for health systems. New technologies often increase costs because they allow us to treat more people for a longer part of their lives.

Yet the general view amongst politicians, managers and others involved in healthcare is that health systems across the world need new thinking. They are increasingly facing escalating demand from an ageing population and the growing incidence of chronic disease. Healthcare is consuming an ever-increasing share of gross domestic product (GDP). The search is on for ways of providing the best quality healthcare as affordably as possible.

The health technology industries – pharmaceutical and biotechnology, medical devices, information technology and the built environment (design, engineering and construction) – drive much of the innovation that takes place in healthcare. They are very big business. Collectively these companies have global revenues in the order of USD 2 trillion a year, about a quarter of overall global spending on healthcare. But they too are experiencing a changing landscape – an evolving market for their products, a changing balance of power across health systems as governments and payers seek to control costs, hence pressure on their business models.

Innovation is regarded by economists and politicians as one of the main drivers of economic growth. It helps to explain why some companies, regions and countries perform better than others in terms of higher productivity and income. For companies involved in the health technology sector, and governments in countries where they are located, there is concern to ensure that their business models are sustainable and continue to successfully deliver new products to the market.

1.1 Why do we need an index of health innovation?

"There is much anecdote and received wisdom about innovation in health but it is hard to determine exactly how countries are performing compared to each other."

Over the last decade the volume of research on innovation in the health sector – from the development and commercialisation of new technologies to challenges in their adoption and diffusion – has grown. There has also been increasing interest in how to embed organisational or service delivery innovations into healthcare organisations. We now know more about the reasons why good, evidence-based ideas fail to take off in healthcare and what needs to be done to support them.

But what is lacking is an evaluation of where different countries stand in relation to their health system's innovation capacity – their capabilities for adopting innovative solutions developed elsewhere and for originating innovations themselves. There is much anecdote and received wisdom – 'the UK is good at generating innovations but poor at adopting them', 'developing countries are a growing source of new ideas', 'the USA over-adopts health technology'. But it is hard to determine precisely how the different countries are performing when compared to each other.

There are many indices or composite indicators measuring aspects of national performance in 'innovation'. Indicators include measures of entrepreneurship, technology development and scientific research, innovation in general, and innovation in public sector organisations. However, these remain underdeveloped in health. There has also been some work to measure the 'readiness' of health organisations to adopt innovations by scoring them against various theoretically-derived factors (e.g. Weiner, 2009).

There has been little progress, however, on the creation of an international comparative health innovation index. A few studies exist, such as those by Deloitte (2012) and PricewaterhouseCoopers (2011), but they have limitations and only provide a snapshot of a few countries in a particular year (see section 4).

From a public policy perspective, it is important to design and evaluate policies that are effective and efficient in stimulating innovation, from the underlying science and technology development to the adoption and diffusion of the useful innovations. To do this we require answers to questions such as: 'how much innovation is going on', 'what type of innovation is it', 'where is it going on', 'do we have the right amount or type of innovation', 'do we need more or less of certain types'?

All these questions require an ability to measure the extent and nature of innovation. A carefully constructed global health innovation index has the potential to help governments focus attention on the relative performance of a country's health system in taking up innovations and a country's health industries in generating them – in short, its national system of health innovation. Such an index would help to pinpoint a country's strengths and weaknesses, pinpointing bottlenecks and issues for attention. Identifying deficiencies in the functioning of a national system of innovation allows those systemic dimensions that are 'failing' in terms of comparative performance to be identified (Faberger et al., 2005). This in turn would help policy makers and the health sector to identify and prioritise the factors that can have the greatest impact on innovation performance.

An index would also be useful to medical technology and life sciences companies by providing an understanding of the attractiveness of each country's health system as a market for products, as a source of innovations, and as a location for R&D. This report outlines the groundwork needed to develop a global health innovation index. Before an index can be developed, a number of conceptual, methodological and practical issues must be addressed, not least what is the index measuring, for whom and for what purpose?

2. The fundamentals of innovation

There are numerous definitions of innovation, the majority embracing three concepts – novelty, diffusion and benefit – arising from the creation of new products, services or processes. Innovation requires a combination of knowledge, capabilities, skills and resources. It is often divided into three sub-processes – invention, commercialisation and diffusion – although it is now recognised that these are all inter-related in often complex ways. Invention entails the production of new scientific and technological knowledge; commercialisation involves the translation of that knowledge into working artefacts (services as well as products); and diffusion involves responding to and influencing market or other demand by matching these artefacts to user needs.

The development of composite indicators of innovation has been influenced by the view that innovation emerges within systems, with actors embedded within a policy, institutional and geographic context which helps determine the scale, direction and relative success of innovative activities (Freeman, 1995; Lundvall, 2010). Systems of innovation scholars consider innovation as the dependent variable rather than an exogenous variable affecting economic growth or output. The central assumptions of neoclassical economics (i.e. utility- and profit-maximizing behaviour by fully informed and independent stakeholders with rational preferences) are dropped in favour of an historical and evolutionary perspective that emphasises complexity, interdependence, and bounded rationality.

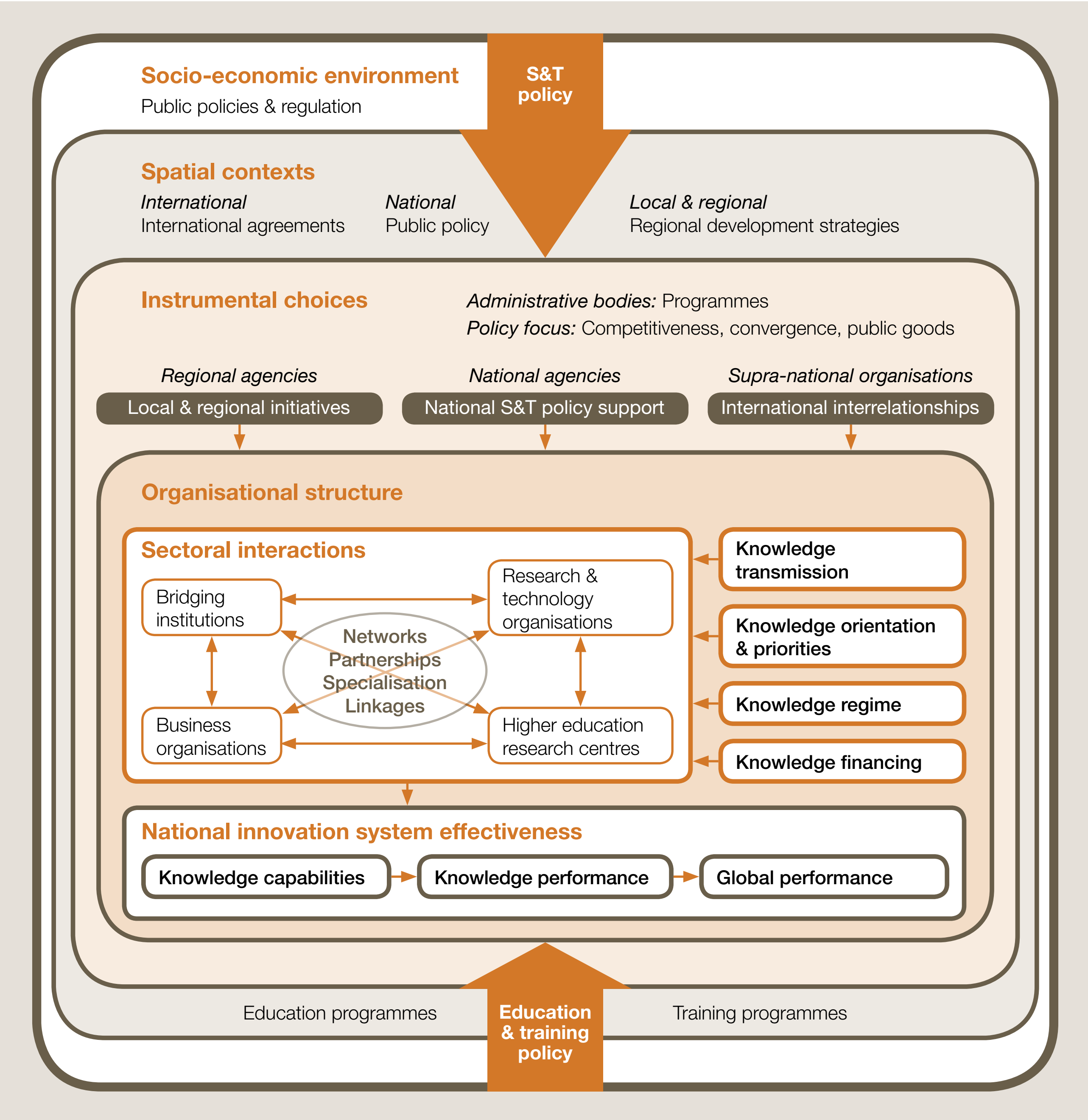
2.1 National systems of innovation

A systems of innovation perspective emphasises interdependence of actors and non-linearity of processes: organisations do not innovate in isolation and the dynamics of the innovation process are complex. It emphasises the role of institutions, defined in a variety of ways. The Organisation for Economic Co-operation and Development (OECD, 1999) uses the term ‘institution’ to describe firms, universities and research institutes; for many systems of innovation scholars (e.g. Edquist, 1997), the term describes habits, norms, routines, practices, rules and laws.

The systems of innovation view is not a formal theory – it makes no specific propositions about causal relations nor does it have well-established empirical rules. It is more of an approach or conceptual framework. It has gained traction with academic researchers, and regional and national authorities, as well as international organisations concerned with the processes of innovation, industrial transformation and links to economic growth, such as the OECD and the European Commission (Bergek et al., 2008).

A related concept is that of the technological regime. This is defined by Geels (2004) as semi-coherent sets of rules embedded in engineering practices, production process technologies, product characteristics, skills and procedures; these are themselves embedded in institutions and infrastructures. Hekkert et al. (2007) and others have emphasised the need to distinguish between national, sectoral and technological systems of innovation, although their boundaries can be far from straightforward. The European Commission (2001) has proposed a framework for a national system of innovation, illustrated in Figure 1.

Figure 1 **European Commission perspective on national innovation systems**



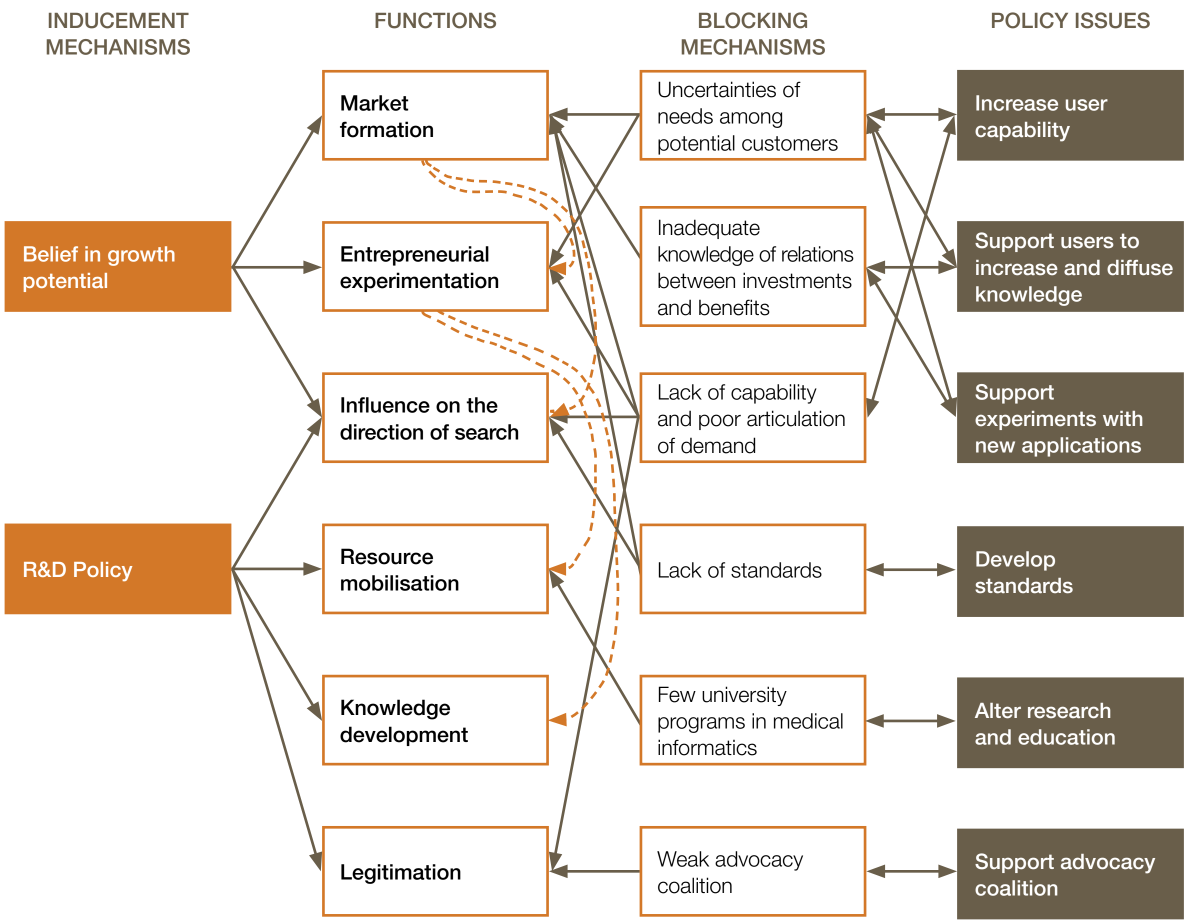
When creating composite indicators or indices of innovation, the national level has been the obvious choice because of interest in benchmarking countries against each other. However, a sectoral approach may also be used, for example, comparing the health sector against other sectors, or sectors within health. Globalisation may blur national boundaries within sectoral systems of innovation, but a likely greater challenge is to define which components and relationships should be included when defining a sector (Malerba, 2002).

What are the implications of an innovation systems perspective on policy-making? One of the consequences of dropping the neoclassical economic view of technological change and innovation (see above) in favour of an evolutionary and learning perspective is that optimality is effectively infeasible (OECD, 1999). Because economic agents are rationally bounded and innovation systems are so complex, they are unable to perceive all the opportunities for innovation and growth (Fagerberg et al., 2006). Conversely, an innovation systems perspective enables policy-makers to think about relationships within the system and ensure that its infrastructure is supported, with structural failures and bottlenecks tackled (OECD, 1999).

Focusing on the functions of a system of innovation instead of the actors within it may be useful when comparing different systems (see Box 1). This is because it reduces the risk of comparing structure rather than functionality: two systems may function equally well even though their structure may be very different (Johnson, 2001).

Box 1 Systems of innovation analytical model

A review by Bergek et al. (2008) proposes that systems of innovation generally serve seven functions. Together these make up a scheme of analysis which researchers and policy makers can use to compare different systems and identify key policy issues and goals: (1) knowledge development and diffusion, (2) influence on the direction of search, (3) entrepreneurial experimentation, (4) market formation, (5) development of positive external economies, (6) legitimation, and (7) resource mobilisation. They identify a number of issues associated with each of these functions, as well as possible indicators and metrics, and apply their analytical scheme to the study of a particular technology innovation system ('Information technology in home care'). This allows them to identify blocking mechanisms and policy goals.



2.2 Innovation in healthcare

"The general consensus is that health is 'different' from other goods and services. This has implications for how we study innovation in health."

To what extent do these concepts apply when looking at healthcare as a sector of the economy? Is not 'health innovation' simply 'innovation'? The general consensus is that health is 'different' from other goods and services (Morrisey, 2008):

"The health care product is ill-defined, the outcome of care is uncertain, large segments of the industry are dominated by non-profit providers, and payments are made by third parties such as the government and private insurers. Many of these factors are present in other industries as well, but in no other industry are they all present. It is the interaction of these factors that tends to make health care unique."

The combination of these factors has implications for how we study innovation in health. Non-profit stakeholders are unlikely to have the same incentives to innovate as profit-driven stakeholders. Extensive regulation and a culture of calculated risk mean that innovations may need to follow lengthy processes of experimentation and legitimation. The separation of payers and providers may complicate the appropriation of returns from innovation. Health systems in many high-income countries are often resistant to change because of rigidity from legacy institutions, professional bureaucracies, silos and vested interests. Those in low- and middle-income countries are often held to be highly innovative, but face significant challenges in scaling-up and replicating examples of new practice.

Modern healthcare is highly specialised and aggregates multiple sub-disciplines. It requires high-tech equipment, yet it is still very much a human interaction, a service. 'Boundaries' between organisational and technological artefacts in health innovation may be very fuzzy. The majority of health innovations involve some combination of 'soft' and 'hard' elements and the interplay between them means that many health innovations are both 'process' and 'service' innovations (Savory & Fortune, 2013). It is often the case that stakeholders have different understandings of what a specific innovation is, its defining components, and how it should be managed (Mackenzie et al., 2010). Differing interpretations of 'evidence' of an innovation's benefits may influence its adoption and diffusion (Davies & Nutley, 1999; Ferlie et al. 2005). What begins as one innovation can morph into several variants involving different adoption decisions, with certain 'core' elements of the original innovation being retained while other 'peripheral' elements are adapted to the local context for its deployment (Denis et al., 2002).

The involvement of governments in health planning, regulation and financing also complicates matters. Vested interests, power and politics are important factors, and investment in innovations which take time to deliver benefits may be shunned in favour of quick wins. Decisions to adopt and diffuse innovations in some countries may be centralised in a small number of key players (e.g. central procurement agencies, or authorities such as the National Institute for Health and Care Excellence in the UK or Canada's Agency for Drugs and Technologies in Health), or devolved to purchasers such as hospital organisations or even individual doctors. The decision-making rules and processes may be opaque, potentially disincentivising innovators.

Different parts of the health system are associated with distinct models of innovation. Innovation in the pharmaceuticals and medical devices sectors follows the more conventional for-profit product innovation model, although market mechanisms of diffusion and adoption are highly regulated, especially in the case of new drugs.

Innovation originating in healthcare providers tends to be more focused on process or service innovation and is difficult to turn into intellectual property (see Table 1). Another characteristic is that health innovation is often developed by its end-users – doctors and other health professionals – and can be hard to know about, much less capture, in formal measures (Savory & Fortune, 2013).

Table 1 Innovation in universities and public sector research establishments and in UK National Health Service organisations

	University/PSRE	NHS organisation
Context	Research and invention is primary purpose; development of published work improves professional status; experimental and risk tolerant	Operational focus with invention a by-product of practice; innovation is problem oriented; highly regulated and risk averse
Evaluation criteria for innovation	Quantity of patents and licenses; improved technical human capital; prestige and reputation; political kudos	Improved operational efficiency; improved quality of care; value of income stream from technology licenses
Technology	Hard technologies (e.g. devices, drugs)	Closely-coupled hard and soft technologies (e.g. processes, services)
Proximity of R&D effort to context of use	Distant; multidisciplinary teams drawn from specialist research staff	Close to the context of use; operationally focused individuals and teams
Mechanisms for diffusion and adoption	Market mechanisms	Diffusion of soft technology

Source: Savory (2006).

2.3 What is known about healthcare innovation

The national systems of innovation approach described above can be applied to healthcare (OECD, 2001), in the sense that different countries have different configurations of institutions, firms and individuals which combine within a system to create and deploy the knowledge, skills and artefacts that underpin healthcare. Although the interplay between these elements can be seen as a complex adaptive system (Plsek & Greenhalgh, 2001), research on innovation in the health sector tends to be separated into work that focuses on the creation of new technologies and work that explores their adoption and diffusion.

On the *supply* side, research has identified how the development of health technologies is characterized by the fact that the institutions involved in delivering health services are also fundamental components of the innovation system (Metcalfe et al., 2005). Historically, the development of many new medical devices has been closely associated with user-innovators – usually clinicians who see a need for a new product or the possibility of improving an existing one (Lettl, 2005). 'Open innovation' approaches to new product development have also emerged in the pharmaceutical and medical devices industries, partly as a response to rising costs and declining productivity of traditional approaches to R&D (Savory, 2006). Another feature of supply-side aspects of healthcare innovation is the existence of a funding gap – the 'valley of death' – for proving the feasibility of a technology in sufficiently large trials. This sits between research (where there is financial support from government bodies, charities etc.) and scalable commercial development (where banks are more likely to fund development).

Research on the ***demand*** side was rather limited until relatively recently, especially in relation to adoption and diffusion. Studies tended to focus on well-defined, bounded innovations being adopted by a single organisational unit (e.g. a new medical device adopted by a single hospital or team), rather than innovations which are less clear cut and involve complex interactions across the wider health system. The perspective was often on individuals making decisions about whether or not to take up a specific innovation. Many studies focused on a small number of causal variables, so little was known about the relative effects and interactions between them and contextual influences – much less was known about the determinants of innovations in health organisations than elsewhere (Rye & Kimberly, 2007; Fleuren et al., 2004).

There is, however, a growing number of studies on the factors influencing adoption at the organisational level. A major systematic review by Greenhalgh and colleagues (2004) found that organisations will assimilate innovations more readily if they are large (which may be a proxy for other important determinants), mature, functionally differentiated (also correlated with size), specialised, in possession of slack resources available for new projects, and they are managed through decentralised decision-making.

Robert et al. (2009) build on this work, identifying as another important factor 'organisational readiness' – organisational commitment to and efficacy in implementing change. However, the notion of organisational readiness is vague and there are major limitations to using it empirically, not least the nature of the relationship between readiness and outcomes (Weiner, 2009). Related to the notion of organisational readiness, various tools have been developed to identify the state of an organisation's culture for innovation. One example, focusing on NHS healthcare trusts in the UK (Maher & Plsek, 2009), aims to assess the current culture for innovation of any collection of individuals. The tool comprises seven dimensions: relationships, risk taking, resources, knowledge, goals, rewards, and methods (e.g. training, skills development). Individuals self-assess their performance in each dimension on a 10 point scale (from -5 to +5), but the constructs underpinning the dimensions are extremely difficult to define, much less measure (e.g. 'emotional support', 'honouring everyone's input', 'trusting, open environment').

Another stream of research has explored disinvestment (Gallego et al., 2010), namely the withdrawal of a hitherto innovative technology. This area is much less researched, partly because there are problems in defining what is meant by 'disinvestment'. Disinvestment is usually passive, driven by changes in medical practice or considerations of cost-effectiveness. There may be strong incentives to retain existing technologies, and path dependency can lead to health organisations becoming locked onto an inferior path. Occasionally, however, active disinvestment occurs, involving removal of funding for ineffective and/or unsafe technologies, usually initiated by the emergence of new evidence of ineffectiveness or harm.

"Now – at a time when the NHS is under pressure – rather than just running harder to stand still, it's time to grab with both hands these practical new treatments and technologies."

Simon Stevens, NHS England Chief Executive, NHS Confederation conference, Manchester, 2016.

Considerations of disinvestment raise the question whether health innovation is always desirable. There is a powerful rhetoric and policy belief that technological innovation is a key driver of sustained economic growth and performance improvements in sectors of the economy. The UK's Innovation, Health and Wealth report (Department of Health, 2011) states that innovation is the only way the NHS can meet the challenge of delivering more healthcare at a time of increasingly constrained resources, and that innovation must become 'core business for the NHS.' A report by the Health Foundation (2015) argues that innovative thinking in five related layers is needed if the NHS is to address the goals of the NHS Five Year Forward View (NHS 2014):

- Scientific discovery, technology and skills
- Focus on population health
- Process improvement for quality and productivity
- New ways of delivering care
- Active cost management

"It is far from clear that more innovation is necessarily a good thing, yet that is what most indices, even if unconsciously, assume."

According to the Health Foundation report, the relative importance of these will vary over the next 15 years. Innovation to improve population health and new ways of delivering care will deliver greater benefits over the longer term, while the immediate innovation requirements are for process improvement and active cost management. Interestingly, the role of science and technology is seen as the least important innovation area, in terms of impact on the challenges of rising demand and resource constraints faced by the NHS. The reasons are not discussed in the Health Foundation report, but must include the difficulty of embedding new innovations within a complex health system. Another problem for science and technology innovation in healthcare is the fact that although innovation may well support process or quality improvements, it is generally associated with increased rather than reduced costs (Bodenheimer, 2005). The reasons include the effect of new technology on increasing the demand for healthcare – research suggests that in the US medical technology accounted for 27-48% of US health spending growth between 1960 and 2007 (Barlow, 2017).

It is far from clear that more innovation is inevitably a good thing, especially where many healthcare innovations are not rigorously evaluated, yet that is what most indices assume. The quality of innovation – measured by its impact on health outcomes, the health system itself and wider societal measures such as share of GDP – matters as much, if not more, than the quantity of innovation. Unfortunately, measuring quality is far from easy. Indicators such as R&D intensity do not necessarily mean that R&D inputs are efficiently used and reflect the quality or efficiency of a country's innovation performance (Freudenberg, 2003).

3. Developing a composite indicator

"Indices have a number of benefits, but they are not without fault; many of their limitations originate in their simplified and aggregated nature, and some have questioned whether the development of composite innovation indices is possible at all."

The use of composite indices has grown in popularity since the first science and technology indicators appeared in the mid-1960s (Godin, 2003). By the late 2000s there were at least 178 composite country indices producing economic, political, social or environmental rankings (Bandura, 2008).

Composite indicators of innovation received a boost in 1992, when the guidelines for the Community Innovation Surveys (CIS) were first set out in the 'Oslo Manual' by the OECD and Eurostat (Mairesse & Mohnen, 2008). The Oslo Manual defines innovation as the implementation of a new or significantly improved product (a good or service), a new process, a new marketing method, or a new organisational method in business practices, workplace organisation, or external relations.

Indices have a number of benefits, but they are not without fault; many of their limitations originate in their simplified and aggregated nature (OECD & JRC, 2008) – see Table 2. Some have questioned whether the development of composite indicators of innovation is possible at all. As Manoochehri (2010) argues, the 'essence of innovation is novelty, so it stands to reason that some innovation will elude any pre-set measuring scheme'.

Nevertheless, the importance of indices for benchmarking ensures that their use will remain part of the portfolio of policy tools, despite their limitations. The European Union (EU) sees this approach as a way for policy makers to check whether there is an innovation gap between member states and other parts of the world, a convergence between old and new member states, and general trends over time. Comparisons between countries against individual indicators such as level of R&D or success in product innovation may flag potentially problematic areas for attention by policy makers (Mairesse & Mohnen, 2008). The creation of league tables derived from index scores makes composite indices especially attractive (Grupp & Mogee, 2004; 2010).

Table 2 Summary of benefits and limitations of indices

Benefits	Limitations
They summarise complex multi-dimensional phenomena	They can mislead policy makers through the way they are constructed and interpreted
They are easier to interpret than multiple single indicators	They may encourage simplistic conclusions and camouflage the true complexity of the issues at hand
They can be used to assess performance over time	Their results are heavily dependent on methodological choices (e.g. choice of indicators, metrics, weights), which are not always transparent, discussed and justified
They aggregate information without losing underlying data	They may fail to account for relevant issues if dimensions for which data are not available are simply ignored
They facilitate communication with general public	They may invite policy makers to develop policies that increase indicator scores without actually resulting in any actual benefit
They simplify the comparison of complex dimensions	

Adapted from: OECD and JRC (2008).

Freudenberg (2003) describes the key requirements of composite indicators. At a minimum, they should be as transparent as possible and provide detailed information on methodology and data sources, including their components, construction, weaknesses and interpretation. Decisions concerning standardisation, weighting and aggregation should be subjected to sensitivity tests. Throughout the index building exercise, both analysts and users should keep in mind that indices are simple representations and comparisons of country performance in given domains, to be used as starting points for further analysis.

The OECD and Joint Research Council have set out in detail the steps needed to develop a composite indicator elsewhere (OECD & JRC, 2008). Readers are directed to their report for an in-depth description of each step. Briefly, they are:

- **Develop a theoretical framework** that provides a rationale for selecting and combining variables into a composite indicator.
- **Identify indicators and metrics** for inclusion taking into account the quality of available data, for example in terms of measurability, coverage and relevance to the phenomenon of interest, but also in terms of the relationships between indicators.
- **Address missing values** through a well-justified and thoroughly tested imputation strategy to derive a complete dataset.
- **Test the structure** of the composite indicator using multivariate analysis to determine whether the underlying statistical structure of the dataset fits well with the proposed theoretical framework and assess differences.
- **Select suitable normalisation and standardisation strategies** to render variables in the dataset comparable.
- **Select weighting and aggregation** procedures that appropriately reflect the theoretical framework and the underlying structure of the dataset, and consider allowing for compensability among indicators.
- **Undertake sensitivity and robustness tests** to determine the impact of methodological choices on final index scores, specifically exploring different indicator variables, normalisation and standardisation schemes, weighting and aggregation strategies, and even theoretical frameworks.
- **Identify key drivers and variables** through statistical analysis of the final index scores and assess patterns across/within countries.
- **Link index results to other indicators** typically via a simple correlation to identify relationships and develop narratives to explain connections.
- **Develop visualisations of the results** to influence and enhance interpretability across different audiences.

4. Innovation indices

As previously indicated, there are very many composite indicators measuring some sort of economic, political, social or environmental phenomenon, including innovation indices. Fewer than a handful of these focus on health innovation. In this section, two recent and well-known indices on general innovation, one on the related concept of entrepreneurship, two health innovation indices and a study on global differences in health innovation diffusion are described. We then discuss the limitations of existing indices.

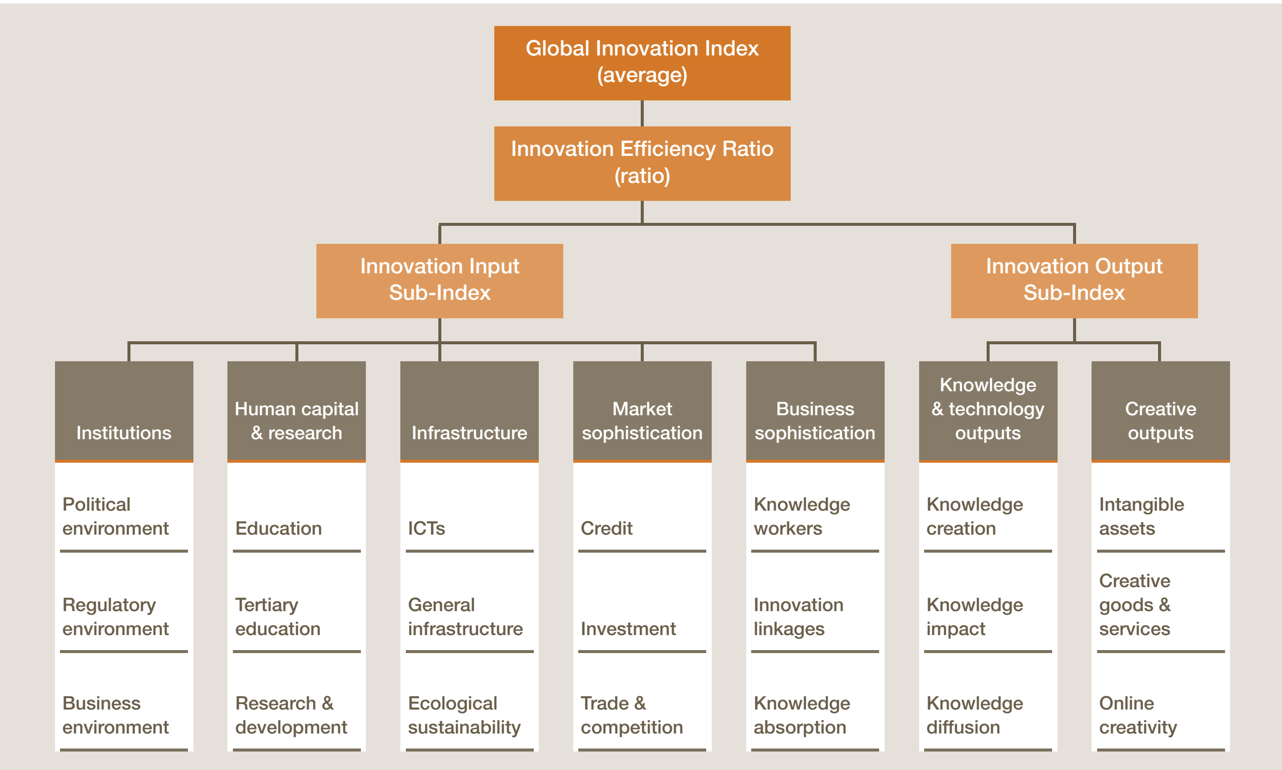
4.1 The Global Innovation Index

The Global Innovation Index (GII) is the result of a collaboration between the World Intellectual Property Organisation (WIPO, an agency of the United Nations), INSEAD and Johnson Cornell University. The 2014 GII is the 7th edition of the index. It is accompanied by a report, which in 2014 focused on human factors in innovation (WIPO, INSEAD & Johnson Cornell University, 2014). The index adopts the Oslo Manual definition of innovation (see above).

The GII is a massive enterprise comprising two sub-indices which are used to calculate an efficiency ratio, seven pillars each divided into three sub-pillars, and a total of 81 individual indicators, 20 of which are composite indicators themselves. It covers 143 economies representing 92.9% of the world population and 98.3% of the global GDP. The GII framework is illustrated in Figure 2.

The GII includes countries for which data are available for at least 51 of the 81 indicators and at least two sub-pillars in each pillar. This means there are countries in the GII which have missing values for 30 of the indicators in the framework.

Figure 2 The Global Innovation Index framework



The GII is the simple average (i.e. equal weights are used) of two composite indices: the Innovation Input and the Innovation Output sub-indices. These two sub-indices are used to calculate an Innovation Efficiency Ratio, an indication of how much return (i.e. outputs) countries are getting from their investments in innovation (i.e. inputs). The two sub-indices are calculated as simple averages of five and two pillars respectively. As such, output pillars are weighted more importantly than inputs in the final GII (see Box 2). The authors state the rationale for equal weights is to place the Innovation Input and the Innovation Output sub-indices on an equal footing, yet this favours countries with higher scores in the Innovation Output pillars (namely in 'knowledge and technology outputs', and 'creative outputs').

This does not mean the authors should add indicators/pillars to the Innovation Output sub-index – or alternatively reduce the number of indicators/pillars in the Innovation Inputs sub-index – simply to ensure that all indicators/pillars are weighted equally (recall that the choice of indicators and structure should be guided as much as possible by a robust theoretical framework, rather than methodological or analytical convenience). It does, however, mean that any interpretation of the final results (and associated rankings) must build on a recognition that different indicators/pillars have different weights on the final GII score. For example, improving the indicator 'online creativity' will have a bigger impact on the GII score than a comparable improvement in the 'business environment' indicator.

As mentioned above, more innovation is not necessarily positive. Measuring the quality of innovation should be as important as measuring the quantity of innovation. The 2013 edition of the GII therefore introduced three indicators to capture quality (Johnson Cornell University, INSEAD & WIPO, 2013):

- The average score of the top three universities in the QS World University Ranking of 2012 (included in the 'human capital and research' pillar).
- The number of patent families filed in at least three offices worldwide (included in the 'business sophistication' pillar).
- The citable documents H index (included in the 'knowledge and technology outputs' pillar).

Box 2 Weights of pillars and sub-indices in the GII

Let II represent the Innovation Input sub-index, IO the Innovation Output sub-index, II_{p1}-II_{p5} the five pillars in the Innovation Input sub-index, and IO_{p1} and IO_{p2} the two pillars in the Innovation Output sub-index. Then the GII composite score can be calculated as:

$$GII = \frac{1}{2} \times II + \frac{1}{2} \times IO =$$

$$= \frac{1}{2} \times \left(\frac{1}{5} \times \sum_{i=1}^{5} II_{pi}\right) + \frac{1}{2} \times \left(\frac{1}{2} \times (IO_{p1} + IO_{p2})\right) =$$

$$= \frac{1}{10} \times \sum_{i=1}^{5} II_{pi} + \frac{1}{4} \times (IO_{p1} + IO_{p2})$$

As a consequence, pillars in the Innovation Output sub-index are weighted 2.5 times more than the pillars in the Innovation Input sub-index. This would explain why Switzerland, which is ranked number one in the GII, is also ranked number one in the Innovation Output sub-index, yet is only 7th in the Innovation Input sub-index.

The GII is revised every year. For the 2013 edition a total of 20 indicators were modified, 10 were deleted or replaced, and 10 experienced methodological changes (e.g. change of scaling factor or change of classification). In the 2014 edition a total of 17 indicators were modified, 4 were deleted or replaced and 10 underwent methodological changes. This means that making inferences about absolute or relative performance on the basis of year-on-year differences in rankings is very difficult. Each ranking reflects the relative positioning of a country based on the conceptual framework, data coverage and the sample of countries, all of which change from one year to another. Despite admitting to these caveats, the authors of the GII do make year-on-year comparisons, highlighting, for example, a persistent innovation divide and identifying movements among the top 10 ranked countries.

The GII has been regularly audited by the European Commission JRC, focusing on the conceptual and statistical coherence of its structure, and the impact of key modelling assumptions on the GII scores and ranks. In its 2014 audit, the JRC concluded that the GII framework is statistically sound and balanced: each indicator (sub-pillar) makes a similar contribution to the variation of its respective sub-pillar (pillar). It is not clear, however, why the index should be balanced (is this because the assumption is that all indicators contribute equally to the aggregate score?). It can also be shown, using the numbers published in the report, that the actual weight of indicator 7.1.3 was more than 8 times bigger than that of indicator 2.1.5. Furthermore, the Pearson correlation between indicator 7.1.3 and the GII score is 0.76, while the same statistic for indicator 2.1.5 is 0.60, which means that indicator 7.1.3 'explains' 25% more variance in the GII than indicator 2.1.5. Setting weights to balance the contributions of individual indicators to the corresponding sub-pillars does not ensure that the contributions of individual indicators to the final GII score are balanced.

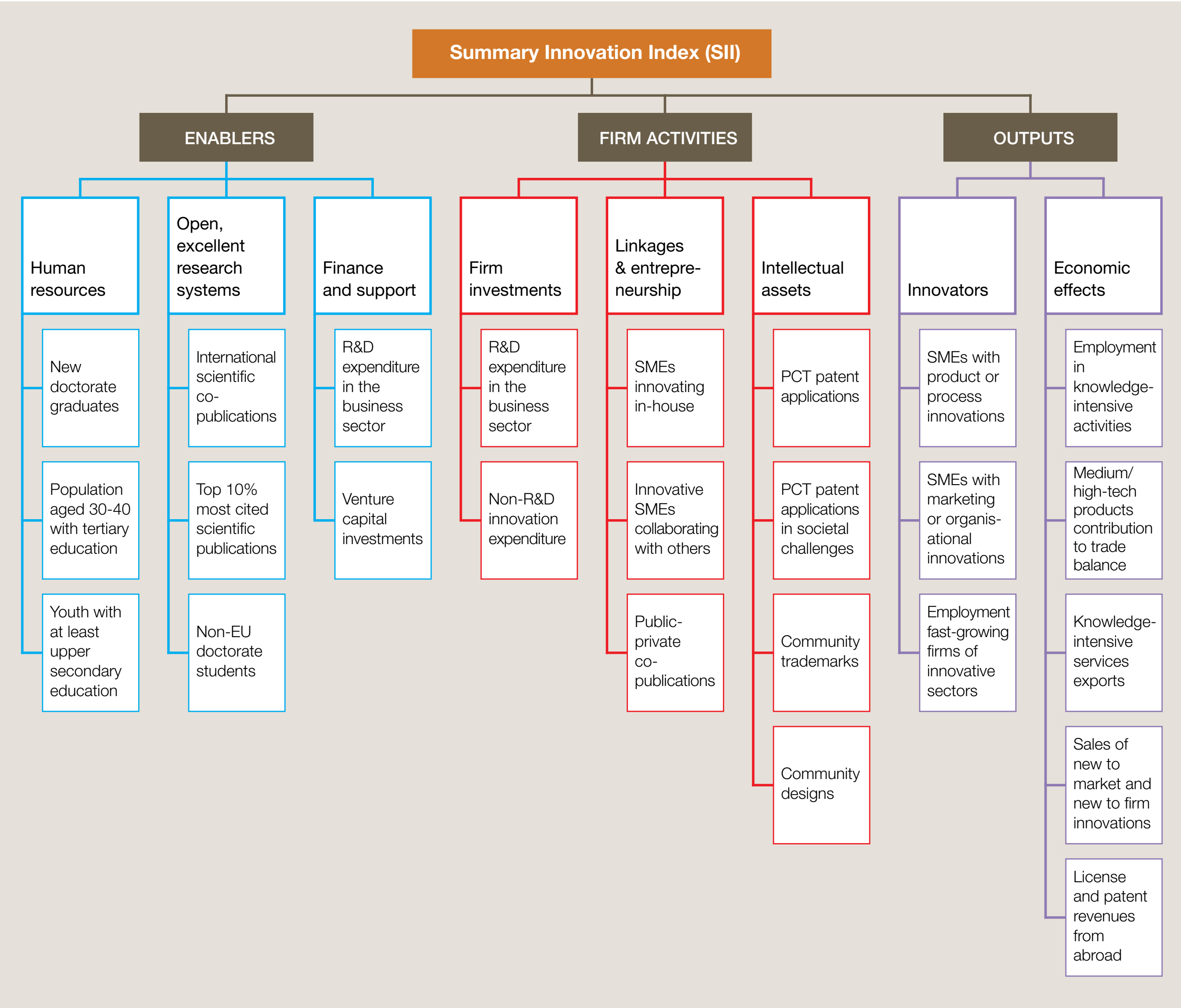
The GII team state that the use of weights as scaling coefficients during the development of the index starting in 2012 represents a significant departure from the traditional use of weights as a reflection of an indicator's importance in a weighted average (WIPO, INSEAD & Johnson Cornell University, 2014). However, as illustrated by comparing indicators 7.1.3 and 2.1.5, weights can be used as scaling coefficients within pillars and still the indicators make different contributions to the final GII score. Because categories (sub-indices and sub-pillars) have different numbers of indicators, to guarantee that weights make similar (or scaled) contributions to the GII score one would need to use different weights when calculating the scores of sub-pillars, sub-indices and the final GII score.

4.2 The Summary Innovation Index

The Innovation Union Scoreboard is an annual exercise undertaken by the European Commission since 2001, providing a comparative assessment of the research and innovation performance of EU member states as well as a selection of other European countries and 10 other global competitors[1]. The scoreboard also provides a composite indicator – the Summary Innovation Index (SII) – as a measure of innovation performance. The SII is the simple average of 25 indicators. These are grouped under three types of indicators – enablers, firm activities, and outputs – which are further categorised into 8 dimensions, as illustrated in Figure 3.

1 Iceland, FYROM, Norway, Serbia, Switzerland, Turkey, Australia, Brazil, Russia, India, China, South Africa, Canada, Japan, South Korea and the USA.

Figure 3 The Summary Innovation Index framework



The Innovation Union Scoreboard uses the most recent statistics from Eurostat and other internationally recognised sources (e.g. the OECD or the United Nations) to maximise comparability of data. The oldest data included in the SII is from 2009 (a single indicator).

Essentially, calculating the SII requires: (1) correcting outliers, (2) setting the reference year for each indicator, (3) imputing missing values, (4) determining maximum and minimum scores, (5) transforming skewed data, (6) calculating re-scaled scores, (7) calculating SIIs. Outliers are dealt with using the same approach as used in the GII. Indicators are normalised using the same Min-Max approach as the GII, but unlike the GII normalisation, the SII uses minimums and maximums chosen across all years. This makes the normalisation step – and hence the scores – sensitive to the addition of new data (all scores must be recalculated whenever data is updated).

The international benchmarking element of the SII, where EU performance is compared to the performance of global competitors, uses only 12 of the 25 indicators in the complete SII and suffers from missing values (see Annex G in European Commission (2014).

Although a full sensitivity analysis is not performed, the report does include a very useful description of how ranks changed between the previous Innovation Union Scoreboard and the 2014 edition (see figure 4-6 in European Commission (2014)). In this respect, it provides a more transparent account of why ranks changed compared to the analysis undertaken in the GII.

Compared to the GII, the Summary Innovation Index is far simpler. It has about a quarter of the indicators and fewer sub-indices. On the other hand, how SII scores are calculated and how the methodological choices affect the final scores is rather opaque. It is not clear why outliers are corrected (are they believed to be incorrect data?), nor why normalisation is performed using maximums and minimums defined over the whole study period. If a new data point is added to one year in one country, the scores of all countries over all the years in that indicator will need to be recalculated. The rationale for defining 2013 as the reference year is not discussed. More importantly, the impact of the methodological choices on the results is not tested (if a sensitivity analysis was performed, it is not reported).

4.3 The Global Entrepreneurship and Development Index

Entrepreneurs generate a broad array of economic benefits from job creation to knowledge spillovers, and hence contribute to economic growth. The institutional context not only regulates the opportunities, feasibility and desirability of entrepreneurial activity, but also its outcomes. Productive entrepreneurship is thus the outcome of an interaction between individual decision making and institutional influences (Ács et al., 2014A, 2014B; Szerb & Autio, 2014).

The Global Entrepreneurship and Development Index (GEDI) was developed to capture the essence of entrepreneurship, measuring the quality and scale of the entrepreneurial process in different countries. It was designed to profile national systems of entrepreneurship and give policy makers a tool for understanding the entrepreneurial strengths and weaknesses of their countries' economies. Currently in its fourth year, the GEDI profiles 120 countries. Guiding principles for the GEDI are listed in Box 3.

The significance of the GEDI has been recognised by a wide spectrum of business publications and has had significant practical impact (Santander, 2014). Its methodology has been validated in rigorous academic peer reviews and widely reported in media, including in The Economist, The Wall Street Journal, Financial Times and Forbes. The methodology has also been endorsed by the European Commission and used to inform the allocation of EU Structural and Cohesion Funds. Its theoretical approach has influenced entrepreneurship policy in organisations such as United Nations Conference on Trade and Development.

The GEDI can inform policy analysis, design and implementation (Ács, Autio & Szerb, 2014) by identifying factors underlying bottlenecks in a national system of entrepreneurship and compare these against relevant peers. It allows the bottleneck factors and policy measures to counter them to be examined more closely and transferable good practices to be identified. Broadly, it provides a common platform for discussion, shifting focus from individual policies and interest silos towards the system as a whole.

Central to the approach in the GEDI is the concept of a national system of entrepreneurship. This is defined in terms of a dynamic, institutionally embedded interaction between entrepreneurial attitudes, ability and aspirations, which drives the allocation of resources through the creation and operation of new ventures. The theory assumes economic growth is driven by a trial-and-error resource allocation process – entrepreneurs allocate resources towards productive uses, conditioned by contextual factors. Because of the multitude of interactions, country-level entrepreneurship is best thought of as a system, the components of which co-produce system performance.

Box 3 The guiding principles of the Global Entrepreneurship and Development Index

- Entrepreneurship is a concept of quality rather than quantity.
- Both institutional and individual (agency) factors are vital in measuring entrepreneurship.
- Measuring the pillars of entrepreneurship is based on a benchmark for each pillar. Determining benchmarks are based on annual data for the years 2006-2012.
- The averages of each pillar are equalized to provide the same marginal effect. This is particularly important from the perspective of entrepreneurship policy.
- The building blocks of entrepreneurship are integrated elements of a system. The performance of the overall system depends on the weakest pillar, and that good performance in one pillar can only partially compensate for a poorly performing one (this is operationalised in the penalty for bottleneck approach).

The GEDI theoretical basis, methodology and data sources are described in detail in its documentation. Briefly, building the GEDI involves the following steps:

1. Start with the raw data for individual-level and institutional variables.
2. Calculate pillar scores by multiplying individual and institutional variables.
3. Cap (through winsorisation, a technique which limits extreme values using a method similar to clipping or trimming above/below a certain arbitrary threshold) pillar scores at 95% (values above the 95th percentile are set to the 95th percentile) by using 2006-2012 data.
4. Normalise capped pillar values using the distance method.
5. Calculate averages for each of the normalised pillars across all countries.
6. Equalise the 15 pillar averages to have the same marginal effect.
7. Adjust each pillar score for each country using the penalty for bottleneck (an approach that penalizes the scores for all the pillars based on the lowest pillar score; it is described in greater detail below).
8. Determine each sub-index score for each country using simple averages.
9. Calculate the super-index for each country by averaging the sub-indices.
10. Determine each country's GEDI by multiplying the super-index by 100.

The GEDI introduces two novelties in the development of composite indicators in relation to weighting and aggregation: the use of institutional variables as weights for entrepreneur-level indicators, and the adjustment of resulting pillar values using a penalty for bottlenecks. The adjusted pillar scores are then simply averaged (using equal weights) to determine the sub-indices, in turn averaged to determine the super-index, or GEDI.

The concept of institutional variables as weights is inspired by the interaction variable approach in regression analysis. In the latter, two independent variables are multiplied by each other to demonstrate their combined effect on the dependent variable.

Using this approach in composite indicators has advantages because: (1) there is no need to set arbitrary weights, (2) weights are country-specific, and (3) weights operationalise the notion that different index components might produce different outcomes in different country settings.

The fifteen pillar scores are thus the result of multiplying individual and institutional variables. Before these are adjusted using the penalty for bottleneck, they are harmonised, i.e. their averages are equalised. First, averages are taken across all countries in a given pillar, resulting in fifteen averages. The authors point out that different values for these averages might suggest it is easier to obtain a high value in pillars with higher average scores than to obtain the same value in pillars with lower average scores; improving a pillar by one unit may not require the same resources as improving another pillar by the same amount.

In practice, the equalisation step involves adjusting pillar scores for all countries using a factor k so that the fifteen pillar averages all become equal to the average of the fifteen averages. This essentially means that pillars with higher average scores will be adjusted downwards closer to the average of the averages, while pillars with lower averages will be adjusted upwards. This is a significant data transformation, and it is not clear why it is performed, since the GEDI says nothing about resources needed to improve individual pillar scores.

After the equalisation of pillar averages, the pillar scores for each country are adjusted using the penalty for bottleneck technique. The penalty for bottleneck is needed to account for system dynamics produced through component interactions. It is a direct corollary of two closely related theories: the theory of the weakest link and the theory of constraints. According to the latter, improvement in system performance can only be achieved by strengthening the weakest link, while the former maintains that elements of the system are only partially substitutable (Ács, Autio & Szerb, 2014). Using a simple analogy, if a cake recipe calls for eggs and flour, substituting eggs with more flour will only get the baker so far.

According to Ács, Szerb & Autio (2014), the penalty function should reflect the magnitude of the penalty such that lower difference implies lower penalty and vice versa. It should also reflect compensation for the loss in one pillar for a gain in another pillar, although full compensability where a loss in one pillar is compensated by the same increase in another pillar is unrealistic. The penalty should reflect the law of diminishing returns, i.e., rise at an increasing rate. The average decrease in the GEDI score is 9%. The policy implications of the penalty for bottleneck are clear: poor performance on a particular feature, such as a bottleneck, should be addressed first since it will have the most negative effect on all the other features.

4.4 Innovation indicators for healthcare in emerging countries

In 2009, Deloitte published a proof-of-concept study benchmarking the ability of ten emerging market countries[2] to promote healthcare innovation relative to each other and to four developed countries[3]. The study's objectives were to: (1) facilitate debate around how to foster innovation in healthcare and life sciences, (2) understand how public and private sectors can encourage innovation, and (3) identify and prioritize levers for maximum impact. The work focused only on healthcare products and technologies, leaving innovation in delivery and services outside the scope.

2 China, India, Singapore, South Korea, Turkey, Czech Republic, Poland, Brazil, Mexico and South Africa.

3 France, Germany, UK and USA.

The framework used in the study (see Table 3) rests on the notion that innovation is dependent on factors related to four key pillars:

- Development: all activities and influential factors underlying the discovery and creation of new innovations.
- Ownership: the factors that have an impact on securing a return on investments made in innovation.
- Diffusion: all the activities involved in the distribution and adoption of innovations.
- Environment: the fundamental conditions required for doing business.

Table 3 Deloitte innovation index for emerging countries framework

Development (13)	Ownership (10)	Diffusion (15)	Environment (10)
A. Plan (R&D Focus) 1. Total GDP on R&D by government 2. # of researchers **B. People** 1. Public expenditure as % of GDP 2. Total tertiary enrollment 3. Research publications **C. Financing** 1. R&D spending by companies 2. Strength of university & industry research collaborations 3. Venture capital availability **D. Facilities** 1. # of science parks 2. Quality of scientific research institutions 3. # of clinical trials 4. # of CROs **E. R&D Output** 1. # of pharma patents filed in the WIPO by firms located in the emerging market	**A. Policy** 1. # of pharmaceutical patents in the emerging market 2. IP office staff strength **B. Intellectual Property** 1. Index of Patent Rights 2. Period of data exclusivity for new drugs in years 3. IP protection and enforcement **C. Ability to Price** 1. Price negotiations as a pre-condition of product approval a. Clearly laid out policies for Pricing b. Unbiased policies for imports and local products c. Opportunity to negotiate price with Government 2. Regulations influencing pricing a. Absence of National Medicines List/ Formulary b. Regulations against parallel imports	**A. Coverage** 1. Per capita expenditure on health 2. Accessibility of healthcare 3. Out-of-pocket expenditure on health 4. Pharmaceutical market growth 5. Pharmaceutical imports growth over previous year **B. Uptake** 1. # of hospital beds per capita 2. # of physicians per capita 3. # of nurses/mid-wives per capita 4. Total telephone subscribers (per 100) 5. Technology readiness 6. Technology usage 7. # of medical schools **C. Outcome** 1. Primary care immunization – national coverage rates 2. # of MRI & CT scanners per capita 3. # of patients organisations (per million)	**A. Political Stability and Economic Development** 1. Political stability 2. Real GDP growth 3. Macroeconomic stability **B. Business Environment** 1. Ease of Doing Business 2. Regulatory quality 3. Corruption perception index 4. General infrastructure 5. Judicial independence 6. Financial market sophistication 7. Worldwide Press Freedom Index

Each pillar is broken down into levers chosen through 'extensive research and expert group inputs'. Each lever in turn comprises several indicators – a total of 48 chosen out of 200 initially identified and researched – in a hierarchy of levels similar to those of the GII and the SII. The individual indicators for each of the levers are illustrated in Table 3. The rationale for choosing and grouping the indicators is not clear (for example, indicators of supply such as the ‘number of hospital beds per capita’ or the ‘number of physicians per capita’ are grouped under the lever ‘uptake’).

The exercise compared countries’ individual performance across each of the four pillars/levers. There is little information in the report regarding the methodological choices that were made. Pillar scores are calculated using weighted averages of normalised indicators. Indicators were standardised, if necessary, using factors such as GDP and population. Their values were then normalised to a scale of [1-7], using a ‘linear-scaling technique that accounts for any outliers’ (Deloitte, 2009), but it is unclear how this transformation of the data was performed.

The report leaves many questions unanswered. Since the theoretical framework is not discussed in detail, it is unclear how the indicators are related to innovation. It is also unclear which indicators were standardised. For example, public expenditure on education is standardised using GDP while the number of MRI and CT scanners is standardised using population. On the other hand, the number of researchers appears not to be standardised, which ceteris paribus means the index favours larger countries. The data sources of indicators are not disclosed, nor are the years to which the data refer.

The actual techniques used to normalise indicators and deal with outliers are not detailed. It is not clear how many indicators exhibit outliers, or what their meaning is, e.g. whether outliers are regarded as incorrect data? The choice of a scale of [1-7], instead of the more prevalent and interpretable [0-1] or [0-100], is not justified. The choice of weights is not discussed, nor is it clear whether scores for levers are calculated. Since one of the objectives of the report is to identify and prioritise levers for impact, it would be useful to report pillar scores (e.g. ‘ability to price’, or ‘coverage’). If a sensitivity analysis was performed, it is not reported.

With limited information on methodological choices, it is difficult to draw any significant conclusions from this study. Author statements such as ‘data needed to structure a viable healthcare innovation index are generally available for emerging markets’ are difficult to verify. It is not easy – if at all possible – to disentangle the effect of methodological choices on the results, and due to the lack of a theoretical framework, it is not clear whether the exercise is measuring innovation rather than something else. The diffusion pillar, for example, seems to be measuring basic healthcare infrastructure, capacity and expenditure. While these indicators can be useful, considering they are measures of diffusion is questionable.

4.5 The Medical Technology Innovation Scorecard

The PricewaterhouseCoopers (PwC) Medical Technology Innovation Scorecard attempts to explore the changing nature of healthcare innovation and adoptive capacity of nine countries[4]. Country performance is measured across five pillars that PwC believes have supported medical technology innovation in the US in past decades (PricewaterhouseCoopers, 2011).

Each pillar has two dimensions comprising differing numbers of indicators, with a grand total of 86 indicators (see Table 4). Out of the 18 indicators that make up the ‘leading resources for innovation’ pillar, only two are specific to healthcare and one of these is from a survey. Scores for 2005, 2010 and 2020 are calculated, but the 2005 scores are determined using only 56 indicators and the 2020 projections are based on only ten indicators, one for each dimension. The 2020 list of indicators

4 Brazil, China, France, Germany, India, Israel, Japan, UK and USA.

includes GDP growth for the 'Market incentives' dimension, physicians per capita for the 'needs and infrastructure' dimension, and non-specified interview data for the 'regulatory approval process' and 'demand and pricing factors' dimensions. The rationale for grouping these indicators under these dimensions is not provided.

Table 4 Medical Technology Innovation Scorecard framework

Powerful financial incentives	Leading resources for innovation	Supportive regulatory system	Demanding and price-insensitive patients	Supportive investment community
Market incentives	Innovation resources	Regulatory approval process	Healthcare demand	Investment environment
Healthcare incentives	Innovative output	Legal environment	Needs and infrastructure	Medical technology commercialisation

The methodological framework of PwC's innovation scorecard is the product of guidance from a steering committee composed of medical device professionals and a benchmarking analysis of the practices of eight different innovation scorecards including those produced by Deloitte, Boston Consulting Group, World Economic Forum and the Economist Intelligence Unit. Data sources included interviews with thirteen executives from the medical devices industry. It is not clear how these were selected or how the interviews were used in the scorecard.

Data were normalised using a scale of [1-9]. As with Deloitte's index, no rationale is given for choosing this scale over more conventional scales. Although not in the report, a simple calculation using the normalised data in the report's appendix shows that each of the ten dimensions is calculated as the simple average of the scores of the individual indicators. These are then themselves averaged to determine the score for each pillar, which are in turn averaged to determine the final composite score. Since the number of indicators in the dimensions is different (the 'market incentives' dimension has seven indicators while the 'innovative resources' dimension has ten), this means individual indicators are weighted differently in the final composite score (indicators in pillars with fewer indicators will be weighted more importantly than those in pillars with more indicators). A sensitivity analysis is not reported.

The results are used to make a number of statements about the evolution of countries' innovation performance. Given the impact of methodological choices on results, it is ill-advised to compare the scores of three different frameworks over three arbitrarily chosen years. For example, the authors expect scores for the US to decline in every pillar between 2010 and 2020, yet this is based on only two indicators for each pillar in 2020 compared to an average of eight in the 2010 framework.
One problem is that the study lacks a solid theoretical framework to guide discussion, with certain statements raising questions. For example:

"Historically, building of system infrastructure, such as hospitals, encouraged innovation. In the future, excess capacity could have the reverse effect. Those countries with limited infrastructure will be more driven to innovate to stretch their resources."

From the above, the relationship between capacity and innovation is not conceptually clear. What is the working definition for 'excess capacity' and how is it measured? Why is the effect of 'excess capacity' on innovation changing over time? Why has limited infrastructure not driven more innovation in the past given that it will do so in the future? Towards the end of the document, five new pillars are proposed, some in direct conflict with the pillars in the initial framework (e.g. price-insensitive customers in 2010 become price-sensitive customers in 2020).

While PwC's innovation scorecard – and the accompanying report – is an improvement over Deloitte's pilot study, both lack a solid theoretical framework to guide development. Methodological choices are not detailed nor are their impact on the results tested via sensitivity analyses. Unfortunately, these limitations greatly reduce the value of these indices for informing policy and decision making.

4.6 Global Diffusion of Healthcare Innovation

The Global Diffusion of Healthcare Innovation (GDHI) is not strictly an index. It is included here because it discusses an essential component of innovation – diffusion – and because it identifies multiple indicators that may be of interest for the development of a global health innovation index. The GDHI was commissioned by the Qatar Foundation and first presented at the World Innovation Summit for Health in Doha in 2013. It is a product of collaboration between Ipsos MORI – a market research organisation in the UK – and Imperial College London's Institute of Global Health Innovation. The GDHI is intended to provoke discussion and debate amongst healthcare leaders about how to encourage the spread of innovation and transform healthcare systems (Darzi & Parston, 2013).

The study examines the importance of a set of enablers and cultural dynamics identified as a framework for diffusion by the Institute for Global Health Innovation and illustrated in Figure 4 (Institute of Global Health Innovation, 2013). The project draws on 100 expert interviews and a survey of 1,521 healthcare professionals and 772 industry professionals across eight countries[5]. One of the findings is that diffusion can be encouraged through different mixes of enablers and cultural dynamics. Individual country dashboards include general information on the healthcare system, expert views on enablers and cultural dynamics, and survey results in the form of spider graphs.

The authors use a broad definition of innovation covering products (for example, new technology, inventions, drugs etc.), practices (ways of working, clinical protocols, workforce changes etc.) and policies (which regulate/influence the use of products and practices). With the exception of Qatar and South Africa, all questionnaires were conducted online. All quantitative data in the report are unweighted. The questionnaires were designed by Ipsos MORI and approved by the Institute of Global Health Innovation. The questionnaires included statements for each of the seven dynamics and survey respondents were asked to select on a scale of [0-10] the extent of their agreement with each statement (0 is strongly disagree and 10 strongly agree). Individual scores for each statement in each country were averaged giving a country-level ranking of the importance of the statements.

Throughout the report, and also in the summary survey results (available online at Ipsos MORI website), only means are reported. These are not especially informative since, as the authors themselves acknowledge, very different ranges of responses can have the same mean (the bigger the range, the bigger the disagreement between

5 Australia, Brazil, England, India, Qatar, South Africa, Spain and the USA.

Figure 4 Global Diffusion of Healthcare Innovation framework.

ENABLERS: facilitating factors that can be present at multiple levels and influenced in a short space of time

- Vision and strategy
- Incentives and reward
- Research funding for research, development and diffusion
- Transparency of research findings and data on demonstrable success
- Information communications technology (ICT) capability
- Specific resources to identify and promote healthcare innovation
- Communication channels across health care, with outside industries and with the public
- Development and renewal of healthcare standards and protocols

CULTURAL DYNAMICS: behaviours – organisational and personal – that are essential for rapid diffusion of innovation

- Harnessing the efforts of patients and the public as co-producers of well-being
- Addressing concerns of healthcare professionals about outcomes and sustainability
- Adapting innovations to suit the local context
- Identifying and supporting champions who embrace and promote change
- Creating the time and space for learning and new ways of working
- Delayering or eliminating old and ineffective ways of working
- Improving the next journey of system transformation

SYSTEMS CHARACTERISTICS: macro level influences on healthcare systems innovation and diffusion

respondents). In the methodology appendix, the authors explain why means should be accompanied by standard deviations so that the standard error can be calculated. With this information it is possible to determine whether the results are statistically significant given an arbitrary level of significance (frequently set at 5%) and determine the confidence intervals. While the importance of standard deviations when checking the statistical significance of mean scores is noted, unfortunately the country dashboards in the report do not include any information on statistical significance.

Although the GDHI is not an index, it exhibits a number of characteristics of composite indicators. Indeed, the only step that is missing is aggregating individual indicator scores into a single overall indicator. Given access to the underlying data, it is possible to aggregate all the information produced in the GDHI into a composite indicator. However, the authors shun the idea of composite indicators because they feel that the diverse nature of health systems and national cultures makes straightforward comparisons difficult, if at all possible. Rather, the focus is on identifying how various systems and countries – with their own unique characteristics and processes – enable and promote behaviours that help diffuse innovation.

4.7 Limitations of existing indices

"General indices of innovation have been around for a long time, but composite indicators of health innovation are in their infancy."

General indices of innovation have been around for a long time, but composite indicators of health innovation are in their infancy. This is noticeable in the methodological quality and sophistication of PwC's and Deloitte's indices compared to the GII or SII. Despite its limitations, the GII is undoubtedly the benchmark with regards to rigour, methodological sophistication and even transparency (its Joint Research Council audit is published alongside the report).

It is important to note that any composite indicator will inevitably suffer from limitations. The lack of a solid theoretical model of innovation means that indices have to make assumptions and methodological choices that are far from consensual. This makes it all the more important for developers of indices to be transparent about the conceptual and methodological choices, discuss the limitations in detail, and identify the likely impact of their choices on the findings.

Out of the general innovation indices surveyed here, the GII is unquestionably the most sophisticated methodologically, but it is unclear whether the added complexity of the GII compared to the simpler methodology of the SII is beneficial. More complex methods may be more powerful but they are also more difficult to fully comprehend and master. For example, the rationale for using Pearson correlations to adjust the weights of the individual indicators included in the GII is far from clear, especially since it does not affect the contribution of individual indicators to explaining variance in the final composite score (see Section 4.1). It is important to understand why this approach was chosen and how it affects the results. With 81 individual indicators and many sub-levels, testing the impact of this choice on the results is not easy to do, much less to communicate to readers. It should be clear to users of indices how more complex methods add value compared to simpler approaches, since there are costs in terms of transparency and interpretability.

The GEDI is also a very sophisticated index, but it too has some limitations. There are at least two issues with the penalty for bottlenecks that warrant further consideration. First, the magnitude of the penalty is arbitrarily chosen; there is no research on what it should be. Second, while policy experts tend to focus on the weakest link, an alternative to the bottleneck penalty could be a 'reward for achievement'; this would reward the whole system for the beneficial effects of having an outstanding score in one or more of the pillars. Research is scant on both the magnitude and the direction (i.e. penalty or reward) of the effect. Although there is a section on 'bottleneck sensitivity analysis' which explores the impact on the final GEDI score of different improvements to specific pillars, sensitivity, robustness and uncertainty analyses are not performed, or if performed not reported. This is unfortunate as the GEDI involves a number of significant data transformations. Without a sensitivity analysis, it is impossible to have an idea of how these methodological choices affect the scores and rankings.

The method used for calculating the individual and institutional scores in the GEDI is unclear from the report. The scores are calculated from equalised pillar values for the 15 individual level pillars, but the equalisation process, as described in the methods section, states that the adjustment is done on the 15 interaction pillars (i.e. individual indicators multiplied by the corresponding institutional variable). It is not clear whether the scores are simply averaged and what aggregation formula is used.

Health innovation indices are especially limited. While the GDHI uses a framework produced in a separate report, neither Deloitte nor PwC describe a workable framework to guide the development of their indices. As a consequence, there is some confusion over the relationship between indicators and the outcome of interest. For example, in PwC's report the relationship between the composite score and a given individual indicator is treated as linear in the calculations, yet it is described as non-linear in the text. Both in PwC's and Deloitte's indices, certain indicators of basic healthcare infrastructure are presented as measures of health innovation rather than healthcare inputs. In general, very limited information on the methods used is provided, including in the SII. It is effectively impossible to reproduce the results because there is simply no information on normalisation, correction of outliers, weighting or aggregation techniques.

Indices which do not report sensitivity, robustness and uncertainty analyses cannot realistically inform policy and decision making. Moreover, while developing an index which accounts for the quality of health innovation may be infeasible, there should at least be a discussion of why more health innovation is considered a positive outcome.

Regarding performance over time, it is important to note that many individual indicators in the indices discussed in this chapter are ratios (e.g. per capita, percentage of GDP, etc.). This means that improvements in such indicators over time may be due to changes in the numerator, the denominator, or simultaneously both the numerator and denominator. For example, if the expenditure on R&D is steady from one year to the next but the GDP decreases, the ubiquitous indicator 'expenditure on R&D' as a percentage of GDP will show improvement, even though expenditure on R&D in absolute terms has not changed. Both developers and users of composite indicators should make room for all possible interpretations of results when drawing conclusions. Both the underlying data and the ways in which it is treated can affect scores.

5. Towards a Global Healthcare Innovation Index

"While an index needs to be sufficiently complex to capture the multidimensional nature of health innovation, it should be as simple as possible so that it can be easily communicated and interpreted."

Building a composite indicator is an exercise in compromise. While an index needs to be sufficiently complex to capture the multidimensional nature of health innovation, it should be as simple as possible so that it can be easily communicated and interpreted. Realistically, it is likely that the list of individual indicators will be as much determined by data availability as by theory, especially if the objective is to survey a broad cohort of countries. Transparency – of theoretical perspectives, methods and data – must guide the exercise (OECD & JRC, 2008).

In this section, we summarise the lessons for each of the steps involved in building an index; we draw on the literature on innovation in general – and in health specifically – and from the state of the art in index development.

5.1 Theoretical framework

The theoretical framework needs to establish: (1) what exactly is being measured, (2) what are the elements that need to be included in the index and what are the relationships between indicators and the outcome, as well as interactions among indicators, (3) the scope and boundary conditions, and (4) the stakeholders involved in healthcare innovation and the potential users of the index.

The framework of choice for composite indices of innovation is that of the 'national systems of innovation'. In the context of health, the concept of a national health science and innovation system has been proposed by the OECD (2001):

"Each country may be said to have a National Health Science and Innovation System (NHSIS), i.e. the set of institutions, teams and individuals who, jointly and individually, create, store and transfer the knowledge, skills and artefacts which define new and improved health products and interventions and more efficient ways of delivering them. This system operates at two levels, the distinct institutions and major programmes and below them a complex network of topics, teams and projects which are the actual 'elements in a collective system of creation, transfer, and use' of health related knowledge."

This is a good starting point, as it incorporates several important elements:

- the stakeholders (institutions, teams and individuals),
- the processes (create, store and transfer the knowledge, skills and artefacts),
- the outputs (new and improved health products and interventions and more efficient ways of delivering them),
- the units of analysis at different levels (the distinct institutions and major programmes and below them a complex network of topics, teams and projects), and
- the three ubiquitous pillars or encompassing concepts of innovation (creation, transfer, and use).

As explained by Liu & White (2001), it is important to develop a framework that can be applied to different countries, and this is easier when the focus is on functions, inputs, processes and outputs rather than categories such as 'research institutes', 'firms' or 'public services' which are open to interpretation. This way the analyst is able to assess how research institutes in other countries influence the innovation process, and see if this type of influence is present or absent in another national system or if it operates through another type of actor.

5.1.1 Quality versus quantity

A recurring challenge is how to incorporate measures of the quality of innovation in composite indicators. For example, the number of patents filed in a given country is invariably included in innovation indices as a measure of output, yet this does not relay any information on the quality of what is being patented, whether it was used, or how much profit or benefit was derived from its use. In healthcare services and public health, it can be extremely difficult to relate changes in services to outcomes, hence incorporating measures of quality in indices is likely to be extremely difficult.

Building a composite indicator inevitably requires making value judgements. By calculating scores, and ranking countries, an index is effectively communicating that there are ‘better’ and ‘worse’ positions. Definitions of ‘good’ may simply reflect quantity rather than actual quality of innovation; a country scoring highly in a quantity index may actually be very poor at quality innovation.

There is currently no real solution to this limitation. One way to deal with it is to include indicators believed to measure the quality of innovation (e.g. the GII included three). Another option is to calculate efficiency ratios. The question however is what to use in the denominator. The option of using actual health outcomes would not really provide a measure of the quality of innovation as health outcomes depend on other factors. The analytical problem here is how to link the (change) in outcome to the innovation.

5.2 Potential indicators and metrics: a tentative structure

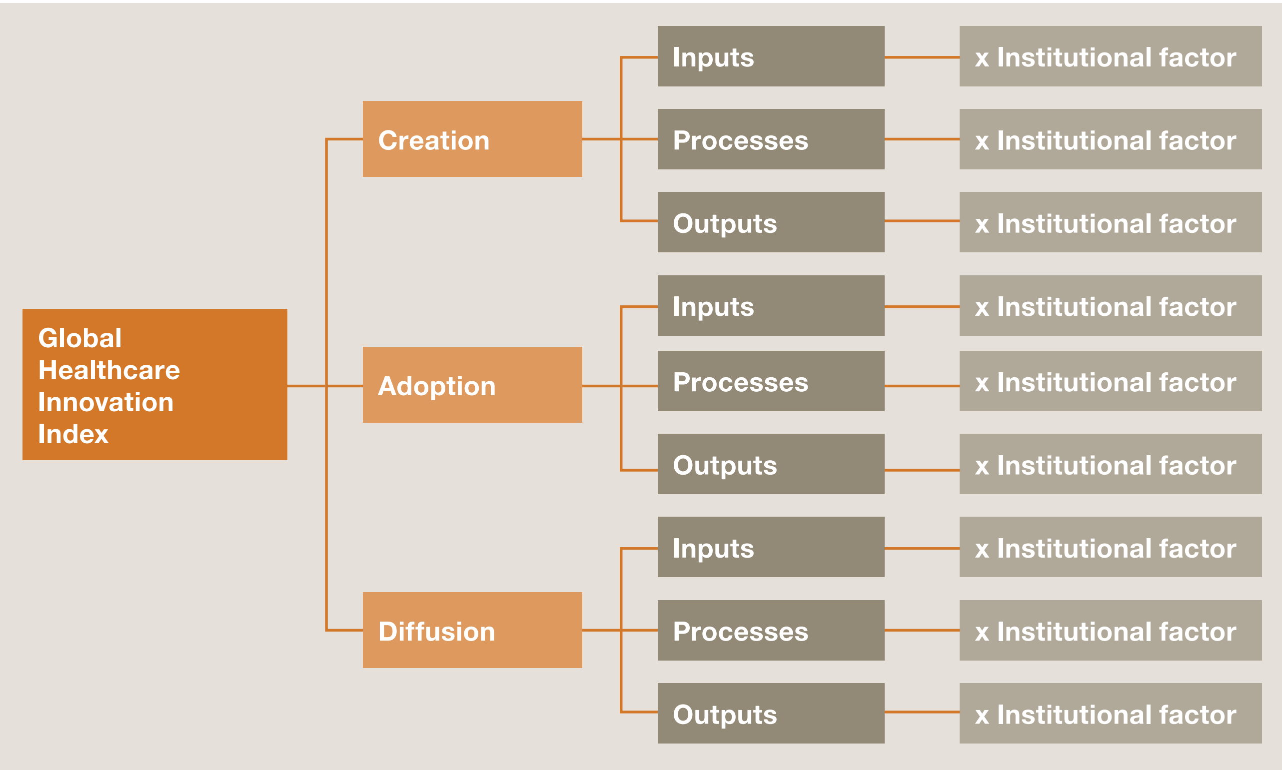
The choice of pillars, indicators and metrics should follow from the theoretical framework. As with the framework itself, the pillars and indicators themselves need to be bounded, in other words one needs to discuss what exactly is being measured by a specific indicator, its scope and boundaries.

The majority of composite indicators use three pillars – inputs/enablers, processes and outputs – to aggregate individual indicators. A fourth is usually also present to capture the institutional context and environmental indicators. A good starting point is the Australian framework (Australian Government, 2011) for measuring public sector innovation (see Table 5). The framework is clear and accurate. For example, user demand and supplier capacity are included as environmental conditions, rather than as measures of diffusion and uptake. A possible way to aggregate individual indicators is shown in Figure 5.

Table 5 Australian Government public sector innovation index framework

Inputs	Process	Outputs	Outcomes
Investment in innovation	Diffusion of innovation	Innovation (activities and implementation)	Societal and environmental impacts
Human resources and skills for innovation	Innovation collaborations	Types of innovation	Quality, efficiency and productivity
Staff attitudes and attributes to innovation	Innovation management practice	Innovation novelty	Improved employee satisfaction
Sources of innovation	Innovation culture & leadership	Innovation intensity	Benefits for users
Technological infrastructure for innovation	Innovation strategy	Intangible outputs (e.g. trademarks, copyrights)	Other intangible effects (e.g. trust and legitimacy)
Environmental conditions			
User demand and supplier capacity	Wider public sector leadership and culture	Political and legislative factors	Other enablers/barriers for innovation (e.g. research basis, skill shortage)

Figure 5 An approach to aggregating individual indicators



The main pillars or sub-indices in the tentative framework are creation, adoption and diffusion, which are three concepts that can easily be grasped, and are of interest to policy makers and users of the index. Each pillar is composed of inputs, processes and outputs; for each of these sub-categories one must find at least one indicator and couple it with one institutional variable, in a manner similar to the GEDI. In its simplest form, the index would include 9 individual indicators and 9 institutional variables, making it relatively easy to communicate and interpret.

A shopping list of individual and institutional indicators is indicated in Box 4.

5.3 Data collection

Coming up with a shopping list of indicators and important topics is relatively easy, finding data on those indicators and topics much less so. For more intangible variables, collecting primary data might be the only solution, as done by the GEDI, GDHI and PwC. This would require significant resources and raise considerable challenges in assuring commensurability, i.e. to be able to show that at least at some level the entities being measured are qualitatively similar (Fagerberg et al., 2006).

Potentially useful sources of data include the OECD's Main Science and Technology Indicators (www.oecd.org/sti/msti.htm), the World Economic Forum, the World Bank, the European Commission and Eurostat, the United Nations and the World Health Organisation. The OECD's manual on measuring health-related R&D (OECD, 2001) may also be useful. As always, a good starting point is to look at sources of data in existing indices.

Box 4 Potential individual and institutional indicators

- Expenditure in health-related R&D
- Health-related IP and patents (e.g. pharma, medical devices)
- Characterisation of clinical trials
- Bibliometrics for health-related academic output
- University and industry collaborations in health
- Supply-side measures (e.g. staff per capita, beds per capita)
- Demand-side measures (e.g. burden of disease, mortality, morbidity)
- Market size measures (e.g. expenditure as % of GDP, out-of-pocket payments)
- General environment (e.g. corruption, political stability)
- Innovation culture/leadership/attitudes (e.g. organisational readiness)
- General IT infrastructure (e.g. computer literacy)
- Health-specific IT infrastructure (e.g. internet usage for health)
- Health tech infrastructure (e.g. MRI scanners per capita)
- Human capital and education (e.g. tertiary education)
- Gender equality, female economic participation
- Technology transfer and economic dimension of globalisation
- Financing (e.g. venture capital, innovation prizes)
- Sources of innovation including user-driven and open innovation
- The Global Diffusion of Healthcare Innovation (GDHI) survey
- Regulatory process (pharma and medical devices mostly, e.g. FDA)
- Health-specific entrepreneurial activity (e.g. GEM survey for health)
- Healthcare industry providers (i.e. medtech companies)
- Procurement processes (e.g. presence agencies supporting or evaluating innovations, such as NICE in the UK)

5.4 Dealing with missing values

From this point onwards, every methodological choice will have an impact on the final index (and sub-index) scores and rankings. Transparency is the guiding principle here. Every choice must be justified clearly and the impact of that choice on the results investigated and reported.

Approaches to dealing with missing values include:

- Data deletion – omit entire records when there is substantial missing data.
- Mean substitution – use mean values computed from available cases.
- Regression – use regressions to estimate missing values.
- Multiple imputation – use a large number of sequential regressions.
- Nearest neighbour – according to some definition of similarity (e.g. geography).
- Ignoring them – take average index of remaining indicators.

The more common ones in existing indices are the last two.

5.5 Testing the structure

Tests of structure help determine whether the conceptual framework is corroborated at the empirical level, but can also be used to group countries that exhibit similar patterns (i.e. scores on individual indicators or pillars) with regards to health innovation. The main three tests are:

- Principal component or factor analysis
- Cronbach coefficient Alpha
- Cluster analysis

Tests can also signal situations where double counting might be a problem. A promising type of cluster analysis is distribution-based clustering, used by the World Bank in the Worldwide Governance Indicators project. This method is discussed below because it can be used to scale, weight and aggregate indicators in a single step.

5.6 Standardisation and normalisation

Standardisation essentially involves transforming individual indicators into ratios using appropriate denominators, frequently GDP or population. This is important when absolute values are believed to provide limited information. A good example is the number of physicians per capita. China has over 1.3 billion people and 1.9 physicians per 1,000 people, while the UK has around 66 million people but 2.8 physicians per 1,000. In absolute numbers, China dwarfs the UK, simply because more people live in China, but the ratios tell a different story. Clearly, standardised indicators should be used rather than absolute numbers.

A different process is normalisation, where for example the number of physicians per 1,000 (which is dimensionless and usually lower than 10) and the total expenditure in health per capita (which is usually measured in PPP USD per person and frequently reaches the thousands) are transformed to comparable scales. Methods for normalising data are numerous: ranking, z-scores, Min-Max, distance, distance to reference country, categorical scales, above or below mean, cyclical indicators, and percentage of annual differences over consecutive years. They have different strengths and weaknesses, and developers of composite indices should select the method that best fits the purpose of the index. Rankings can be useful for simple benchmarking, while distance to a reference country might make sense if the focus is on how one country performs compared to others.

5.6.1 Dealing with outliers

While all indices reviewed in this report correct or eliminate outliers (the method of choice for dealing with outliers is winsorisation or clipping), none of them actually explain why outliers are a problem to start with. If a country is significantly ahead of the pack in a certain indicator, should not the normalised data reflect that advantage? If the distance between the best ranked countries and the rest is big, should we artificially reduce it by correcting it? The rationale for correcting outliers is that they could polarise results and unduly bias the rankings. However, the use of a normalisation technique guarantees that any polarisation will remain within the indicator, and if the data are accurate, it can be argued that this is unlikely to lead to 'undue bias'. However, this provisional judgement needs to be explored in greater detail – namely by surveying the appropriate literature in the context of indices – before deciding to correct or eliminate outliers.

5.7 Weighting and aggregation

This is perhaps the most controversial step in building a composite indicator. Weights are a mathematical inevitability when aggregating data on multiple indicators. When indices suggest they are not using weights, they are in fact using equal weighting. The term 'equal' here is misleading because it gives the impression that all indicators matter equally in calculating the composite score. However, when there are multiple averaging steps (e.g. when calculating sub-indices and pillars) the final weights may be quite different. Simple averages (i.e. equal weighting) should only be used if all group sizes (i.e. number of individual indicators in sub-indices and pillars) are identical.

It is therefore important to re-emphasise that index development should be guided by prior theory to provide justification for assumptions made about weighting within an index. Weights can serve at least two purposes: reflect the importance of specific indicators (e.g. expenditure on R&D might be considered more important than the number of science doctorates and as such be given a bigger weight) and correct for issues with data availability and quality. Weights are also related to compensability or substitutability, i.e. the notion that different indicators can somehow substitute for each other. The GEDI methodology contributes two novel approaches, one for each of these two purposes. Firstly, institutional variables are used as weights or alternatively interaction/mediator variables. Secondly, the scores of all indicators are adjusted to reflect the worst indicator (i.e. the bottleneck). These are important steps forward in the way composite indicators are developed.

5.7.1 The unobserved components model

A different approach to aggregation – and outliers, standardisation and weighting – is provided by the World Bank's Worldwide Governance Indicators project in the form of the unobserved components model (UCM). This is a type of cluster analysis (distribution-based clustering). While it is not a simple method, it affords a number of advantages namely: (1) it is less sensitive to extreme outliers, which are not corrected/eliminated, (2) weights reflect the relative precision of the underlying data, (3) it formalises aggregation as a signal extraction problem, thus uncertainty about estimates are possible, a feature that is lacking in many composite indicators, (4) it provides better motivation for aggregating different data sources, and (5) it is conceptually more defensible than linear/geometric aggregation (Kaufmann et al., 2011).

Determining how a country fares in terms of health innovation using a plethora of individual indicators can be seen as a 'signal extraction problem'. In the context of a governance index, Kaufmann et al. (2011) explain that each of the individual data sources provides an imperfect signal of some deeper underlying notion of governance that is difficult to observe directly. This makes it hard to isolate an informative signal about the unobserved governance component common to each individual data source, and raises questions about how to optimally combine the many data sources to get the best possible signal about the nature of governance in a country based on all the available data. The UCM provides a possible solution to this signal extraction problem.

5.8 Sensitivity and robustness

This is an absolutely essential step in the development of an index, yet it generally receives little or no attention. To conduct a sensitivity analysis, it is not necessary to test all combinations of different indicators, metrics, normalisation, standardisation, weighting and aggregation techniques that are available. It is, however, necessary to at least test the best alternative methods to the ones used.

It is recommended that for every step (i.e. normalisation, weighting, etc.), at least two methods should be used and their impact on final scores compared. For the sake of simplicity and parsimony, it should not be necessary to test different combinations of methods, unless there are reasons to believe that different combinations can have significantly different impacts.

6. Conclusions

Comparative composite indices have long been of interest to policy makers, academics, industry and journalists. Over the last two decades many indices have been developed across a range of socio-economic phenomena and geographies. Our understanding of the methods of index construction has improved and more data have become available, although many composite indices are simplistic and tell us little. We draw three conclusions from this review of the use of comparative composite indices of innovation.

First, one should proceed with caution when deriving policy, research or other implications from composite indicators. Before we can be confident of their implications for the comparative performance of countries or regions, for example, and establish benchmarks to underpin policy or other decisions, it is essential to understand how innovation indicators impact on innovation processes. This in turn requires indices to be underpinned by a clear and strong theoretical framework.

Composite indicators can lead to counter-productive behaviours. One such unintended and undesirable consequence is gaming. In the case of government innovation policy an example would be influencing the score of individual indicators in ways that improve the index score without affecting innovation processes themselves – the policy aim should not simply be to increase the value of an indicator, but to address the more challenging problem of improving the conditions that the indicators are expected to capture (Archibugi et al., 2009). An important consequence of this is that the theoretical framework and hypotheses that underpin an index may well need updating to take into account the impact of past policy interventions around particular innovation factors. In other words, once an indicator is included in an index, it may become a target for improvement, so that what was known before about the observed relationship between that indicator and innovation may in turn change (Freeman & Soete, 2009). Innovation indices must therefore be used with full awareness of their limitations and the stage of evolution of the particular economies and societies that are being compared (Freeman & Soete, 2009). Composite indicators can send misleading policy messages if they are poorly constructed or misinterpreted, with ‘big picture’ results potentially inviting users to draw simplistic analytical or policy conclusions (European Commission, 2001; OECD & JRC, 2008).

Second, there are a number of issues which have only briefly been mentioned here, but which warrant greater attention in the future. One is the use of indices to explore the dynamics and evolution of health innovation across countries. To do this, one needs panel data. If the use of cross-sectional data is already fraught with difficulties, the use of panel data is much more so. This is especially true when surveys are conducted. Another issue is variation at the regional level. Countries are naturally interested not only in their national scores, but also in how well different regions are doing, something which is explored in the GEDI index.

Finally, policy makers should be careful not to draw conclusions about the relationship between composite innovation performance indicators and other indicators of policy interest, such as health outcomes. While it may be of interest to correlate innovation index scores with those of an index of health outcomes, it is essential to remember that correlation should not be mistaken for causality. We also need to be aware of double counting, which can lead to stronger correlations simply because the same variable is included in the indices being correlated. For example, in its list of recommended core indicators of health systems performance, WHO includes health expenditure and number of beds per inhabitant. Both these indicators are included in Deloitte’s emerging markets index, so correlating the index with WHO’s indicators will not be very informative about actual relationships between heath innovation and health outcomes. A global healthcare innovation index can provide a lens through which to then develop hypotheses about potential relationships between innovation performance and other healthcare phenomena, but we must remember that it is only a lens – a starting point for further research.

Appendix
Summary of key lessons

Theoretical framework

- The index should measure innovation, not health outcomes.
- Creation, adoption and diffusion are potential pillars.
- The concept of a 'national system of health innovation' is useful.
- The goodness or badness of innovation is likely to be important but difficult to conceptualise and capture.
- The framework should set the rationale for choosing indicators.
- It should describe how individual indicators relate to innovation.
- It should identify/signal potential moderating variables and interactions.
- The scope and boundaries need to be set at this stage.

Data collection

- Data availability is a necessary – but not sufficient – condition for inclusion.
- Indicators with no variation across countries are not useful.
- Involving experts and index stakeholders in this step is advisable.
- Proxy measures may be justified but the rationale must be clearly set out.
- The choice of data sources should be guided by the 6 dimensions of quality: relevance, accuracy, timeliness, accessibility, interpretability, and coherence.
- The concept of optimality is infeasible.

Missing values

- Indicators with substantial amounts of missing data should be dropped.
- If the dropped indicator is essential, then look for a proxy measure.
- Using data from 'similar' countries should be avoided.
- Testing more than one technique is useful for understanding impact on results.

Tests of structure

- Use Cronbach's Alpha to test internal consistency of pillars and sub-indices.
- Explore structure of underlying data using whatever method is appropriate.

Standardisation and normalisation

- To produce league tables, use ranking or distance, as in the GEDI, since it preserves relative differences, and compare distance with ranking when aggregating.
- Min-Max should be avoided since it does not preserve relative differences.
- Techniques which require arbitrary thresholds (e.g. distance to reference country) should be avoided, unless clearly justified.

More specifically on outliers:

- Determine whether outliers are due to incorrect data and if so, try to identify the data collection problem.
- If the problem cannot be corrected through offsetting the data, the data should not be used at all.
- If outliers reflect correct data, then explain why they are a problem before correcting/eliminating them.
- Test the impact of different correction techniques on results.

Weighting and aggregation

It is recommended that both the GEDI and the UCM approaches be used and their results compared. The UCM requires large amounts of data, so would not work well if the index has few indicators. It also does not include or a penalty for bottlenecks, although the feasibility of adding a penalty function could be explored. There are three advantages in using UCM:

- The normalisation step transforms the underlying data so that it can be compared, i.e. so that it has the same units. Unlike simple ranking or Min-Max normalisation, UCM retains some information on the size of the gaps between countries and is less sensitive to extreme outliers (indeed these are included normally rather than being corrected or eliminated).
- Weights reflect the relative precision of the underlying data.
- Viewing the development of the index as a signal extraction problem provides a conceptually stronger rationale for including different sources of data believed to somehow be related to the phenomenon of interest. The idea is that different sources provide noisy or imperfect signals of innovation and that by combining them a better estimate can be achieved.

Sensitivity analysis

Explore impact on results of:

- Dealing with missing values differently (e.g. imputation, ignore).
- Using different normalisation techniques (e.g. ranking, distance).
- Dealing with outliers differently (e.g. winsorisation, UCM, not at all).
- Using different aggregation and weighting techniques (e.g. linear, UCM).
- Adding/excluding countries from the data set.
- Adding/excluding years from the data set.

A more detailed sensitivity analysis could also test uncertainty with respect to the theoretical framework itself, by adding/excluding indicators or changing the way they relate to each other.

It is also important to think about the impact of including composite indicators as individual indicators in an index. This is frequently done, yet the implications are far from trivial, and usually not discussed.

Policy implications

The index should:

- Spur discussion among policy makers and stakeholders.
- Be a starting point for a deeper exploration of a country's innovation system.
- Act as a 'boundary object' (an artefact that promotes the sharing of knowledge between stakeholders from different backgrounds).

References

Ács Z, Szerb L, Autio E (2014a) *Global Entrepreneurship & Development Index 2014*.

Ács Z, Autio E, Szerb L (2014b) National Systems of Entrepreneurship: Measurement issues and policy implications. *Research Policy* 43(3), 476-494.

Archibugi D, Denni M, Filippetti A (2009) The technological capabilities of nations: The state of the art of synthetic indicators. *Technological Forecasting and Social Change* 76(7), 917-931.

Australian Government (2011) *Working towards a measurement framework for public sector innovation in Australia*. APSII Draft Discussion Paper.

Bandura R (2008) *A Survey of Composite Indices Measuring Country Performance: 2008 Update*. (February).

Barlow J (2017) *Managing Healthcare Innovation*. World Scientific.

Bergek A, Jacobsson S, Carlsson B, Lindmark S, et al. (2008) Analyzing the functional dynamics of technological innovation systems: A scheme of analysis. *Research Policy* 37(3), 407-429.

Bodenheimer T (2005) High and rising health care costs. Part 2: technologic innovation. *Annals of Internal Medicine* 142(11), 932-937.

Darzi A, Parston G (2013) *Global Diffusion of Healthcare Innovation (GDHI)*. Institute of Global Health Innovation, Imperial College London.

Davies H, Nutley S (1999) The rise and rise of evidence in health care. *Public Money & Management*, January-March, 9-16.

Deloitte (2009) *Innovation indicators for healthcare in emerging countries – understanding and promoting innovation in emerging markets*.

Denis J, Hebert Y, Langley A, Lozeau D, Trottier L (2002) Explaining diffusion patterns for complex health care innovations. *Health Care Management Review* 27: 60-73.

Department of Health (2011) Innovation Health and Wealth, Accelerating Adoption and Diffusion in the NHS. Available from: http://webarchive.nationalarchives.gov.uk/20130107105354/http:/www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_134597.pdf.

Edquist C (1997). Systems of innovation approaches. Their emergence and characteristics. In Edquist C (Ed.) Systems of innovation. T*echnologies, institutions and organizations*. London: Pinter Publishers/Cassell Academic.

European Commission (2001) Innovation and enterprise creation: Statistics and indicators.

European Commission (2014) *Innovation Union Scoreboard 2014*.

Fagerberg J, Mowery D, Nelson R (2006) *The Oxford handbook of innovation*. Oxford Handbooks.

Ferlie E, Fitzgerald L, Wood M, Hawkins C (2005) The nonspread of innovation: the mediating role of professionals. *Academy of Management Journal* 48(1): 117-134.

Fleuren M, Wiefferink K, Paulussen T (2004) Determinants of innovation within health care organizations. Literature review and Delphi study. *International Journal for Quality in Health Care* 16(2), 107-123.

Freeman C (1995) The “National System of Innovation”in historical perspective. *Cambridge Journal of Economics* 19(1), 5-24.

Freeman C, Soete L (2009) Developing science, technology and innovation indicators: What we can learn from the past. *Research Policy* 38(4), 583-589.

Freudenberg M (2003) Composite indicators of country performance: a critical assessment. *OECD Science, Technology and Industry Working Papers*. 1635.

Gallego G, Hall J, Haas M, Viney R (2010) *Reducing the use of ineffective health care interventions*. CHERE Working Paper 2010/5. University of Technology, Sydney

Geels F (2004) From sectoral systems of innovation to socio-technical systems. *Research Policy* 33(6-7), 897-920.

Godin B (2003) The emergence of S&T indicators: why did governments supplement statistics with indicators? *Research Policy* 32(4), 679-691.

Greenhalgh T, Robert G, Bate P, Kyriakidou O, Macfarlane F, Peacock R (2004b) *How to spread good ideas. A systematic review of the literature on diffusion, dissemination and sustainability of innovations in health service delivery and organisation*. National Co-ordinating Centre for NHS Service Delivery and Organisation R & D (NCCSDO).

Grupp H, Mogee M (2004) Indicators for national science and technology policy: How robust are composite indicators? *Research Policy* 33(9), 1373-1384.

Grupp H, Schubert T (2010) Review and new evidence on composite innovation indicators for evaluating national performance. *Research Policy* 39(1), 67-78.

Health Foundation (2015) *Shaping the future. A strategic framework for a successful NHS*. London: The Health Foundation.

Hekkert M, Suurs R, Negro S, Kuhlmann S, et al. (2007) Functions of innovation systems: A new approach for analysing technological change. *Technological Forecasting and Social Change* 74(4), 413-432.

Institute of Global Health Innovation (2013) *From innovation to transformation – a framework for diffusion of healthcare innovation*. Institute of Global Health Innovation, Imperial College London.

Johnson A (2001) Functions in Innovation System Approaches. *Innovation*. [Online] 1-19. Available from: www.druid.dk/conferences/nw/paper1/a_johnson.pdf.

Johnson Cornell University, INSEAD & WIPO (2013) *The Global Innovation Index 2013: The Local Dynamics of Innovation*.

Kaufmann D, Kraay A, Mastruzzi M (2011) The Worldwide Governance Indicators: Methodology and Analytical Issues. Hague Journal on the Rule of Law. [Online]. 3 pp.220-246. Available from: doi:10.1017/S1876404511200046.

Lettl C (2005) The emergence of radically new health care technologies: Inventive users as innovation networkers. *Technology & Healthcare* 13(3), 169-183.

Liu X, White S (2001) Comparing innovation systems: a framework and application to China' s transitional context. *Research Policy* 30(7), 1091-1114.

Lundvall B.-Å (2010) *National systems of innovation: Toward a theory of innovation and interactive learning*. Anthem Press.

Mackenzie M, O'Donnell C, Halliday E, Sridharan S., et al. (2010) Evaluating complex interventions: one size does not fit all. *BMJ* 340(7743), 401-403.

Maher L, Plsek P, Bevan H (2009) *Creating a culture for innovation. Guide for executives*. NHS Institute for Innovation and Improvement.

Mairesse J, Mohnen P (2002) Accounting for innovation and measuring innovativeness: An illustrative framework and an application. T*he American Economic Review* 92(2).

Mairesse J, Mohnen P (2008) *Innovation surveys and innovation policy*. Presentation at the conference 'En route vers Lisbonne', Luxembourg, 4-5 December 2008.

Malerba F (2002) Sectoral systems of innovation and production. *Research Policy* 31(2), 247-264.

Manoochehri, G. (2010) Measuring innovation: Challenges and best practices. *California Journal of Operations Management* 8(1), 67-73.

Metcalfe J, James A, Mina A (2005) Emergent innovation systems and the delivery of clinical services: The case of intra-ocular lens. *Research Policy* 34(9): 1283-1304.

Morrisey M (2008) Health care. *The Concise Encyclopedia of Economics*. Available from: http://econlib.org/library/Enc/HealthCare.html [Accessed: 15 July 2014].

NHS (2014) *NHS Five Year Forward View*. Available from: www.england.nhs.uk/ourwork/futurenhs [Accessed: 7 November 2014].

OECD (1999) *Managing National Innovation Systems*. Paris: OECD.

OECD (2001) *Measuring Expenditure on Health-related R&D*. Paris: OECD.

OECD & JRC (2008) *Handbook on Constructing Composite Indicators: Methodology and user guide*.

Plsek P, Greenhalgh T (2001) Complexity science: The challenge of complexity in health care. *BMJ* 323 (7313), 625-628.

Prahalad C (2010) *The fortune at the bottom of the pyramid: eradicating poverty through profits*. Pearson Education.

PricewaterhouseCoopers (2011) *Medical Technology Innovation Scorecard – the race for global leadership*.

Robert G, Greenhalgh T, Macfarlane F, Peacock R (2009) *Organisational factors influencing technology adoption and assimilation in the NHS: a systematic literature review*. Report for the National Institute of Health Research, SDO Project (08/1819/223).

Rye C, Kimberly J (2007) The adoption of innovations by provider organisations in healthcare. *Medical Care Research Review* 64: 235-278.

Santander (2014) Unlocking the potential of regional entrepreneurs. [Online]. 2014. Available from: http://bit.ly/1klxeBu [Accessed: 31 July 2014].

Savory C (2006) Does the UTTO model of technology transfer fit public sector healthcare services? *International Journal of Innovation and Technology Management* 3(2), 171-187.

Savory C, Fortune J (2013) *NHS Adoption of NHS-developed Technologies*. Final report. NIHR Service Delivery and Organisation programme.

Weiner B (2009) A theory of organizational readiness for change. *Implementation Science*, 467. Available from: doi:10.1186/1748-5908-4-67 [Accessed: 10 July 2014].

WIPO, INSEAD & Johnson Cornell University (2014) T*he global innovation index 2014: the human factor in innovation*.

PIRU POLICY INNOVATION RESEARCH UNIT

The Policy Innovation Research Unit (PIRU) brings together leading health and social care expertise to improve evidence-based policy-making and its implementation across the National Health Service, social care and public health.

We strengthen early policy development by exploiting the best routine data and by subjecting initiatives to speedy, thorough evaluation. We also help to optimise policy implementation across the Department of Health's responsibilities.

Our partners

PIRU is a collaboration between the London School of Hygiene & Tropical Medicine (LSHTM), the Personal Social Services Research Unit (PSSRU) at the London School of Economics and Political Science (LSE), and Imperial College London Business School plus RAND Europe, the Nuffield Trust and the Public Health Research Consortium.

The Unit is funded by the Policy Research Programme of the Department of Health.

Policy Innovation Research Unit
Department of Health Services Research & Policy
London School of Hygiene & Tropical Medicine
15–17 Tavistock Place, London WC1H 9SH

Tel: +44 (0)20 7927 2784
www.piru.ac.uk